Insper
São Paulo

# Modelo Multi-Estado de Markov em Cartões de Crédito

Daniel Evangelista Régis

Rinaldo Artes

**Insper Working Paper**
WPE: 137/2008

# MODELO MULTI-ESTADO DE MARKOV EM CARTÕES DE CRÉDITO

Daniel Evangelista Régis

Rinaldo Artes

Ibmec São Paulo

**Resumo**

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 escore 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**: Análise de sobrevivência, *Anti-attrition coring , Credit* scoring, Modelos multi-estado de Markov, Risco de crédito.

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 estimou-se um Modelo Multi-estado de Markov (Jackson, 2006) e seu desempenho foi comparado com modelos padrões de regressão logística.

Na Seção 1 apresentamos uma breve introdução sobre tipologia de modelos de previsão de risco de crédito. Na seção 2 descrevemos o modelo multi-estado de Markov utilizado. A base de dados é apresentada na Seção 3 e os resultados obtidos a partir dos modelos, na Seção 4. Por fim, tecemos nossas considerações finais na Seção 5.

## 1. Tipologia de modelos para previsão do risco de crédito

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 obter um novo cliente pode ser 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.

### 1.1 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 identificação 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, 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 nas quais 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 ACSP. 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 escore. Essas classes definem grupos de clientes com nível semelhante de risco, permitindo à instituição o desenvolvimento de políticas específicas para cada grupo.

### 1.2 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.

## 2 - Metodologia

Neste trabalho, 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, um modelo multi-estado de Markov.

### 2.1 Conceitos de Análise de Sobrevivência

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.

Seja T a variável que indica o tempo de falha, define-se a taxa de falha instantânea no tempo t condicional à sobrevivência nesse tempo (risco de falha no instante t) como

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}, \tag{2.1}$$

O modelo semiparamétrico de Cox (ver Colosimo e Giolo, 2006, por exemplo), 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 $\lambda_0$ uma função não negativa denominada função de base. Note que $\lambda(t) = \lambda_o(t)$ quando $x_i = 0$, sendo $x_i$ um vetor de covariáveis fixas. O modelo para a função risco é chamado semiparamétrico porque apenas os efeitos das covariáveis são tratados parametricamente.

### 2.2 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.

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.

Seja $E_i(t)$, o estado assumido pelo indivíduo *i, i*= 1, ..., n, no instante *t*= 1, ..., τ. 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 $\Delta t$ é dada por

$$p_{irs}(\Delta t) = P(E_i(t+\Delta t) = s \mid E_i(t) = r). \tag{2.2}$$

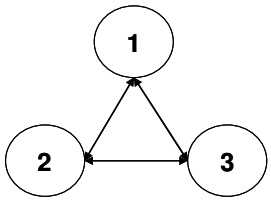
Define-se a intensidade de transição entre os estados r e s por

$$q_{irs}(t) = \lim_{\Delta t \to 0} \frac{P(E_i(t+\Delta t) = s \mid E_i(t) = r)}{\Delta t} \tag{2.3}$$

Podemos observar que a intensidade de transição, (2.3), assemelha-se a $\lambda(t)$ dada em (2.1), 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 Figura 2.1). 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 mostrar que as probabilidades

de transição (2.2) são obtidas através das componentes da matriz $P_i(\Delta t) = \exp(Q_i\ \Delta t)$ (detalhes nas referências desta seção e em Cox e Miller, 1965).




$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}$$

**Figura 2.1 – Um modelo multi-estado 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 $(t_{i1}, t_{i2}, ..., t_{i\tau_i})$ e os correspondentes estados $(E_i(t_{i1}), E_i(t_{i2}), ..., E_i(t_{i\tau_i}))$ são os dados para o indivíduo *i*. Considere um modelo multi-estado geral, com um par de sucessivos estados $E(t_j)$, $E(t_{j+1})$ nos tempos $t_j$, $t_{j+1}$. A contribuição deste par de estados para a verossimilhança é:

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

em que $L_{i,j}$ 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).

Características individuais constantes ou que variem no tempo 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

$$q_{irs}(x_i(t)) = q_{rs}^{(0)} \exp(\beta_{rs}^T x_i(t)), \qquad (2.4)$$

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 covariáveis 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.

### 2.3 Transições em Cartões de Crédito

Neste trabalho são 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 do cartão de crédito

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

Estados absorventes:

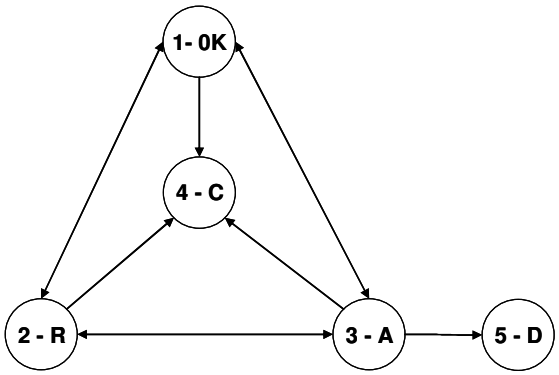
4 - Cancelamento voluntário (C): cancelamento do cartão de crédito por iniciativa do cliente.

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

Na Figura 2.3 as possíveis transições e a estrutura da matriz de intensidades são ilustradas. 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*.

As transições possíveis a partir do estado *rotativo* 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.*




$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} & 0 \\ q_{21} & q_{22} & q_{23} & q_{24} & 0 \\ q_{31} & q_{32} & q_{33} & q_{34} & q_{35} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

**Figura 2.3 – Modelo multi-estado em cartões de crédito e matriz de intensidades de transição**

## 3 – 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.

Para o desenvolvimento do modelo optou-se por utilizar uma quantidade reduzida de variáveis comportamentais que indicassem 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[1] em 6 meses, 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.

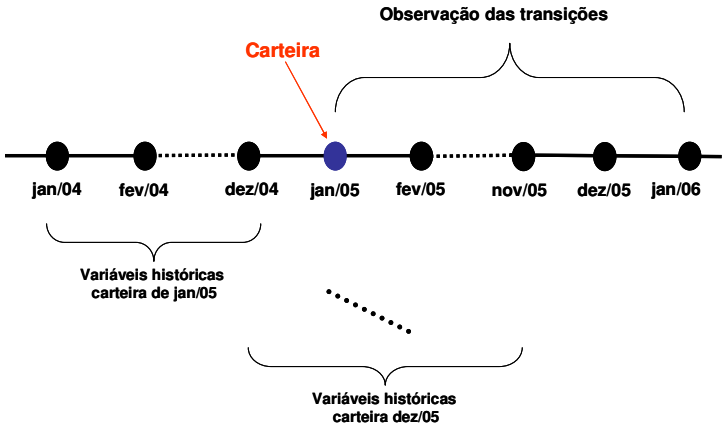



**Figura 3.1 – Estrutura da base para o modelo multi-estado de Markov**

Foram selecionadas 7 variáveis explicativas (6 comportamentais tratadas de forma contínua e 1 cadastral tratada de forma categorizada) 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). Abaixo são listadas as variáveis selecionadas:

Variável 1: variável dividida em 3 categorias em que cada categoria é associada a um determinado nível de limite de crédito, de acordo com a renda do cliente.

Variável 2: mede a utilização de crédito rotativo no histórico de 12 meses.

Variável 3: mede a inatividade do cliente no histórico de 12 meses.

Variável 4: mede o grau de problemas de atraso do cliente no histórico de 12 meses.

Variável 5: mede o grau de utilização do produto no histórico de 6 meses, atribuindo maiores pesos aos meses mais recentes.

Variável 6: mede a utilização do limite de crédito no histórico de 6 meses.

Variável 7: mede o máximo endividamento do cliente no histórico de 6 meses.

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

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 foram definidas variáveis resposta utilizando o seguinte critério:

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.

## 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. Utilizando-se o método stepwise,, 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..

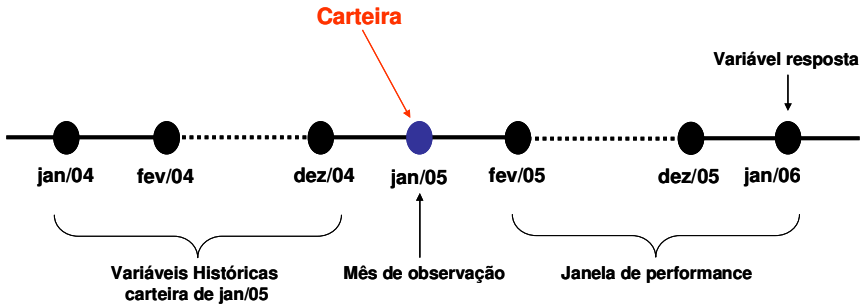



**Figura 4.1 – Estrutura da base para os modelos de regressão logística**

Para a estimação dos modelos de regressão logística foi utilizado o módulo Enterprise Miner, do software estatístico SAS. Uma observação importante é que as estimativas dos parâmetros destes modelos, foram obtidas 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 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.

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

| Variável | Estimador | Erro Padrão | P-Value |
|---|---|---|---|
| Intercepto | 5.4127 | 0.1934 | 0.0000 |
| Variável 3 | -0.0393 | 0.0159 | 0.0136 |
| Variável 4 | -0.2173 | 0.0129 | 0.0000 |
| Variável 5 | 0.0909 | 0.0219 | 0.0000 |
| Variável 6 | -0.3191 | 0.0280 | 0.0000 |
| Variável 7 | -0.1386 | 0.0295 | 0.0000 |

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

| Variável | Estimador | Erro Padrão | P-Value |
|---|---|---|---|
| Intercepto | 4.6852 | 0.1537 | 0.0000 |
| Variável 2 | -0.0230 | 0.0100 | 0.0217 |
| Variável 3 | -0.0413 | 0.0130 | 0.0000 |
| Variável 4 | -0.1884 | 0.0117 | 0.0015 |
| Variável 5 | 0.0849 | 0.0180 | 0.0000 |
| Variável 6 | -0.3034 | 0.0235 | 0.0000 |
| Variável 7 | -0.1432 | 0.0236 | 0.0000 |

De uma forma geral, o resultado dos modelos de regressão logística para prever o default do cliente apresentou resultados coerentes com a lógica de crédito nos modelos de 6 (L1) e 12 meses (L2).

### 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.

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

| Variável | Estimador | Erro Padrão | P-Value |
|---|---|---|---|
| Intercepto | 1.3625 | 0.1567 | 0.0000 |
| Variável 1 / categoria 1 | -0.5456 | 0.1452 | 0.0002 |
| Variável 1 / categoria 2 | 0.0342 | 0.1254 | 0.7851 |
| Variável 3 | -0.0452 | 0.0128 | 0.0004 |
| Variável 4 | -0.2081 | 0.0137 | 0.0000 |
| Variável 6 | 0.4766 | 0.0247 | 0.0000 |

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

| Variável | Estimador | Erro Padrão | P-Value |
|---|---|---|---|
| Intercepto | 1.2198 | 0.1484 | 0.0000 |
| Variável 1 / categoria 1 | -0.5456 | 0.1452 | 0.0002 |
| Variável 1 / categoria 2 | 0.0342 | 0.1254 | 0.7851 |
| Variável 3 | -0.0762 | 0.0112 | 0.0000 |
| Variável 4 | -0.1499 | 0.0122 | 0.0000 |
| Variável 5 | -0.0544 | 0.0215 | 0.0113 |
| Variável 6 | 0.3659 | 0.0183 | 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).

### 4.2 Modelo Multi-estado de Markov

Para a estimação do modelo multi-estado de Markov utilizou-se o pacote MSM, implementado em $R^2$. O R é um software livre destinado à computação estatística e construído de forma colaborativa com muitos desenvolvedores, detalhes sobre o software podem ser encontrados em http://cran.r-project.org/. O pacote MSM, desenvolvido por Christopher Jackson (Jackson, 2006), permite estimar modelos multi-estado de Markov em tempo contínuo.

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 (2.4), em que $q_{rs}^{(0)}$ é a função de base (*baseline*) para a transição *r-s*, e $\beta_{rs}^{T}$ 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.

Na tabela 4.6 podemos ver que, de uma forma geral, 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.

## Tabela 4.5 – Estimativas modelo multi-estado de Markov

| Transição | Variável | Estimativa | Erro Padrão | P-Value |
|---|---|---|---|---|
| 1-2 | Variável 1 | -0,03047 | 0,03062 | 0,15982 |
| 1-2 | Variável 2 | 0,14620 | 0,00503 | 0,00000 |
| 1-2 | Variável 3 | -0,02888 | 0,00718 | 0,00003 |
| 1-2 | Variável 4 | 0,04471 | 0,00948 | 0,00000 |
| 1-2 | Variável 5 | 0,01012 | 0,00903 | 0,13124 |
| 1-2 | Variável 6 | 0,11930 | 0,01240 | 0,00000 |
| 1-2 | Variável 7 | 0,10450 | 0,01091 | 0,00000 |
| 1-3 | Variável 1 | -0,01259 | 0,05174 | 0,40387 |
| 1-3 | Variável 2 | 0,06376 | 0,01025 | 0,00000 |
| 1-3 | Variável 3 | 0,00073 | 0,01048 | 0,47208 |
| 1-3 | Variável 4 | 0,09908 | 0,01470 | 0,00000 |
| 1-3 | Variável 5 | -0,00451 | 0,01599 | 0,38901 |
| 1-3 | Variável 6 | 0,04896 | 0,02139 | 0,01104 |
| 1-3 | Variável 7 | 0,04720 | 0,01794 | 0,00425 |
| 1-4 | Variável 1 | -0,00170 | 0,06546 | 0,48962 |
| 1-4 | Variável 2 | 0,03898 | 0,01451 | 0,00361 |
| 1-4 | Variável 3 | 0,03621 | 0,01198 | 0,00125 |
| 1-4 | Variável 4 | -0,06060 | 0,02668 | 0,01155 |
| 1-4 | Variável 5 | -0,08920 | 0,02327 | 0,00006 |
| 1-4 | Variável 6 | -0,00358 | 0,02753 | 0,44832 |
| 1-4 | Variável 7 | 0,01480 | 0,02247 | 0,25503 |
| 2-1 | Variável 1 | -0,00408 | 0,03258 | 0,45017 |
| 2-1 | Variável 2 | -0,08829 | 0,00503 | 0,00000 |
| 2-1 | Variável 3 | -0,02233 | 0,01279 | 0,00107 |
| 2-1 | Variável 4 | -0,03142 | 0,00746 | 0,00001 |
| 2-1 | Variável 5 | 0,02735 | 0,00913 | 0,00137 |
| 2-1 | Variável 6 | -0,03854 | 0,01475 | 0,00450 |
| 2-1 | Variável 7 | -0,01829 | 0,01380 | 0,09245 |
| 2-3 | Variável 1 | -0,24240 | 0,03054 | 0,00000 |
| 2-3 | Variável 2 | -0,02538 | 0,00496 | 0,00000 |
| 2-3 | Variável 3 | 0,00893 | 0,00660 | 0,08778 |
| 2-3 | Variável 4 | 0,04792 | 0,00595 | 0,00000 |
| 2-3 | Variável 5 | -0,02601 | 0,00960 | 0,00337 |
| 2-3 | Variável 6 | 0,02572 | 0,01401 | 0,03322 |
| 2-3 | Variável 7 | 0,03647 | 0,01794 | 0,00411 |
| 2-4 | Variável 1 | -0,08053 | 0,08983 | 0,18501 |
| 2-4 | Variável 2 | 0,03139 | 0,01425 | 0,01382 |
| 2-4 | Variável 3 | -0,06310 | 0,02185 | 0,00194 |
| 2-4 | Variável 4 | -0,06522 | 0,02306 | 0,00234 |
| 2-4 | Variável 5 | -0,01988 | 0,02675 | 0,22867 |
| 2-4 | Variável 6 | -0,02300 | 0,04060 | 0,28554 |
| 2-4 | Variável 7 | -0,01635 | 0,03943 | 0,33919 |
| 3-1 | Variável 1 | -0,23820 | 0,06056 | 0,00004 |
| 3-1 | Variável 2 | -0,13830 | 0,01253 | 0,00000 |
| 3-1 | Variável 3 | -0,01462 | 0,01279 | 0,12642 |
| 3-1 | Variável 4 | -0,03284 | 0,01048 | 0,00087 |
| 3-1 | Variável 5 | 0,04227 | 0,02206 | 0,02767 |
| 3-1 | Variável 6 | -0,00245 | 0,02447 | 0,46017 |
| 3-1 | Variável 7 | 0,02176 | 0,02427 | 0,18495 |
| 3-2 | Variável 1 | -0,01131 | 0,03679 | 0,37927 |
| 3-2 | Variável 2 | -0,00657 | 0,00686 | 0,16923 |
| 3-2 | Variável 3 | -0,03542 | 0,00840 | 0,00001 |
| 3-2 | Variável 4 | -0,04513 | 0,00670 | 0,00000 |
| 3-2 | Variável 5 | 0,04670 | 0,01315 | 0,00019 |
| 3-2 | Variável 6 | 0,02228 | 0,01575 | 0,07864 |
| 3-2 | Variável 7 | 0,02129 | 0,01615 | 0,09368 |
| 3-4 | Variável 1 | -0,04925 | 0,08283 | 0,27606 |
| 3-4 | Variável 2 | -0,10300 | 0,02056 | 0,00000 |
| 3-4 | Variável 3 | 0,05655 | 0,01793 | 0,00081 |
| 3-4 | Variável 4 | 0,04588 | 0,01401 | 0,00053 |
| 3-4 | Variável 5 | -0,04671 | 0,05006 | 0,17537 |
| 3-4 | Variável 6 | -0,10470 | 0,03603 | 0,00183 |
| 3-4 | Variável 7 | -0,09622 | 0,03432 | 0,00253 |
| 3-5 | Variável 1 | 0,03015 | 0,04335 | 0,24339 |
| 3-5 | Variável 2 | -0,01256 | 0,00823 | 0,06358 |
| 3-5 | Variável 3 | -0,03478 | 0,01003 | 0,00026 |
| 3-5 | Variável 4 | 0,00357 | 0,00793 | 0,32602 |
| 3-5 | Variável 5 | 0,04935 | 0,01578 | 0,00088 |
| 3-5 | Variável 6 | 0,04152 | 0,01899 | 0,01440 |
| 3-5 | Variável 7 | 0,03771 | 0,01959 | 0,02713 |

**Tabela 4.6 – Variáveis significativas para cada tipo de transição**

| Variável | 1-2 | 1-3 | 1-4 | 2-1 | 2-3 | 2-4 | 3-1 | 3-2 | 3-4 | 3-5 |
|---|---|---|---|---|---|---|---|---|---|---|
| Variável 1 | | | | | x | | x | | | |
| Variável 2 | x | x | x | x | x | x | x | | x | x |
| Variável 3 | x | | x | x | x | x | | x | x | x |
| Variável 4 | x | x | x | x | x | x | x | x | x | |
| Variável 5 | | | x | x | x | | x | x | | x |
| Variável 6 | x | x | | x | x | | | x | x | x |
| Variável 7 | x | x | | x | x | | | x | x | x |

## 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.

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 escore, 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 4.7 mostra as estatísticas de Kolmogorov-Smirnov e os coeficientes 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,* onde as observações são divididas em decis sendo que o primeiro decil possui os clientes de maior risco e o último decil possui os clientes de menor risco. Todos os indicadores de performance foram obtidos na amostra de validação.

**Tabela 4.7 – Indicadores de desempenho dos modelos**

| Escore | Estatística de Kolmogorov-Smirnov | Coeficiente 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 4.7. No caso de probabilidade de cancelamento os modelos obtidos através da regressão logística apresentaram resultados superiores, também mostrado na Tabela 4.7.

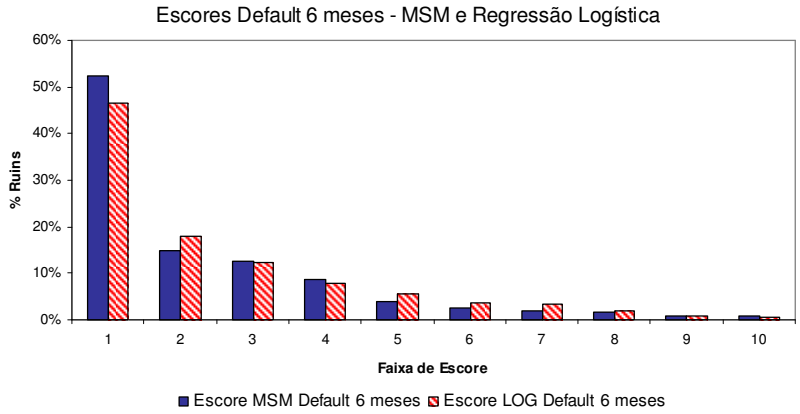



**Figura 4.2 – Gráfico de back test dos modelos de escore de default em 6 meses**

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.

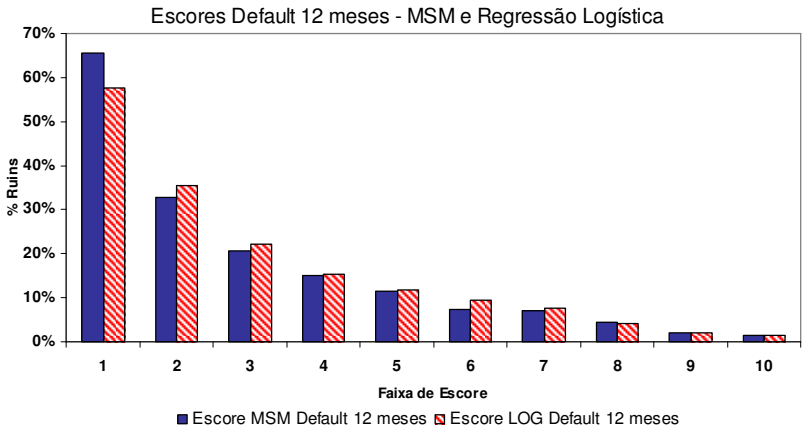



**Figura 4.3 – Gráfico de back test dos modelos de escore de default em 12 meses**

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.

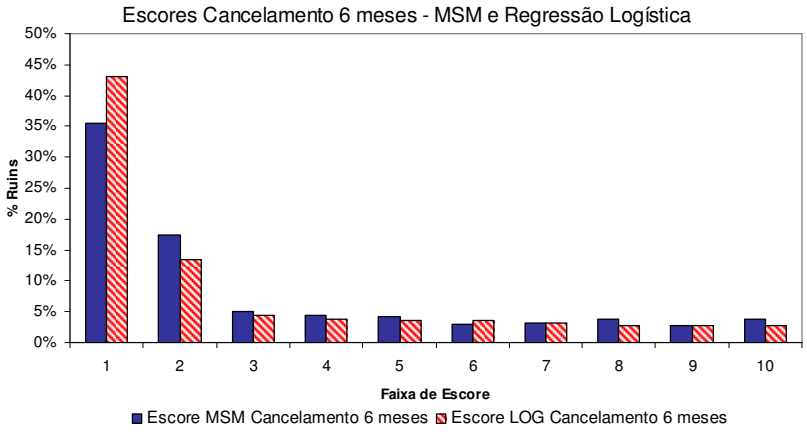



**Figura 4.4 – Gráfico de back test dos modelos de escore de cancelamento em 6 meses**

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 se melhora 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.

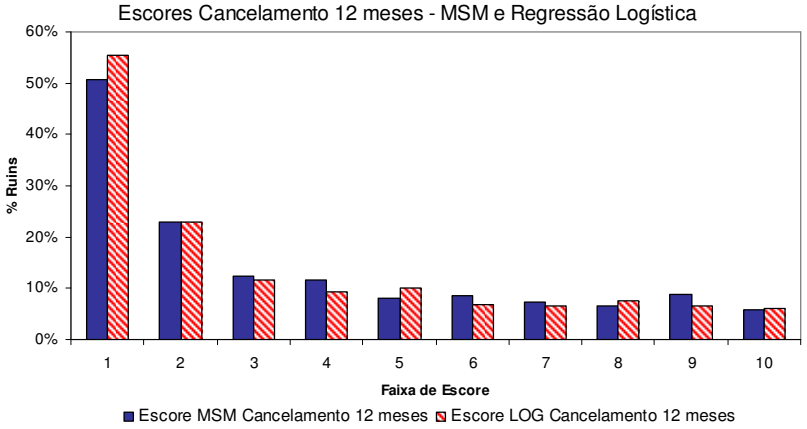



**Figura 4.5 – Gráfico de back test dos modelos de escore de cancelamento em 12 meses**

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.

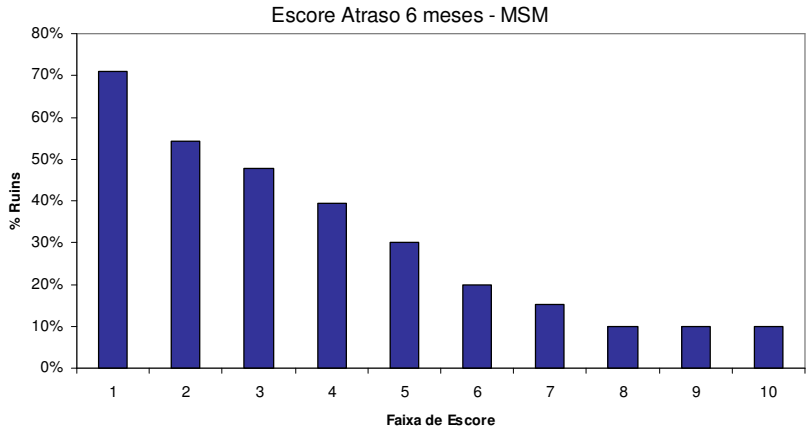



**Figura 4.6 – Gráfico de back test do modelo de escore de atraso em 6 meses**

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.

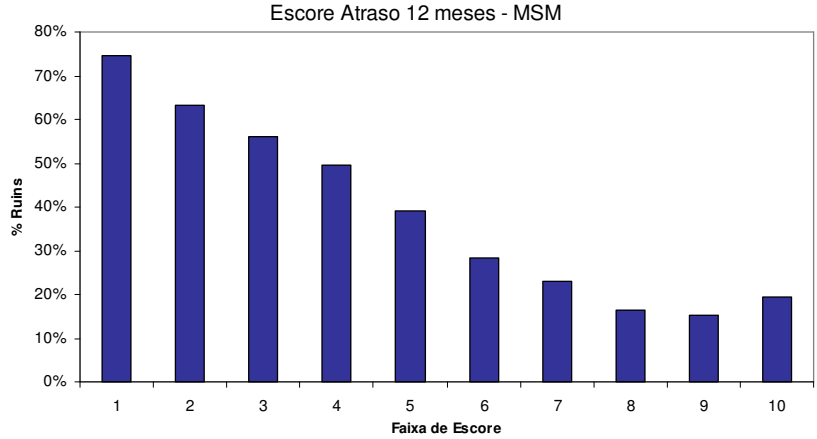



**Figura 4.7 – Gráfico de back test do modelo de escore de atraso em 12 meses**

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.

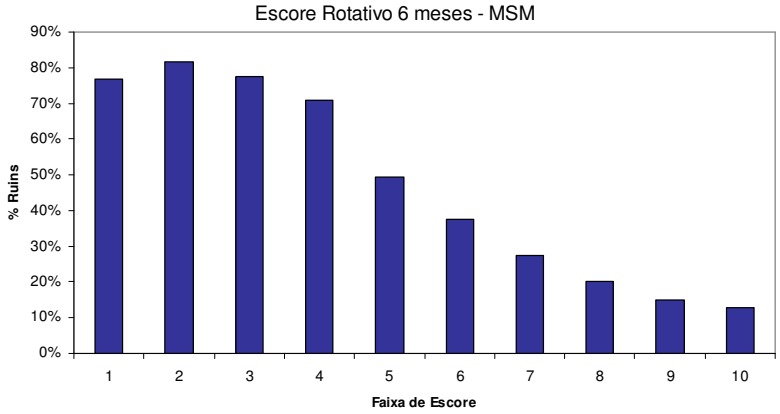



**Figura 4.8 – Gráfico de back test do modelo de escore de utilização do crédito rotativo em 6 meses**

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.

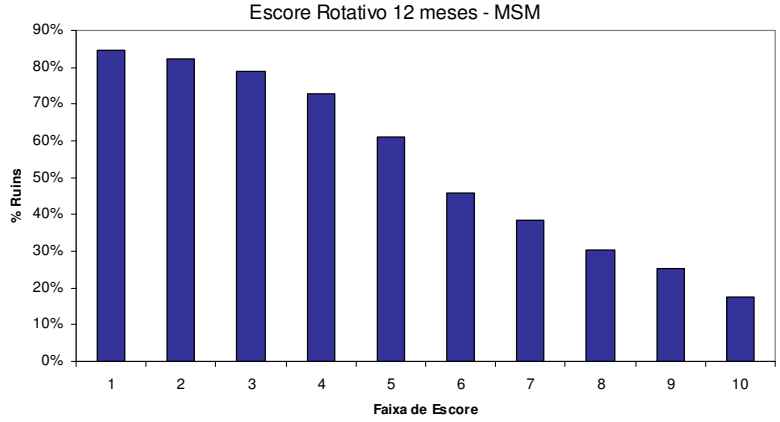



**Figura 4.9 – Gráfico de back test do modelo de escore de utilização do crédito rotativo em 12 meses**

## 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.

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

**Bibliografia**

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.

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

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

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.

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.

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

Hougaard, P. (1999). *Multi-state Models: A Review*. Multi-state Models: A Review. *Life Data Analysis,* **5**, 239-264.

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

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.

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.

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.

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.

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.

# Apêndice

## *Baselines e Transições*

### Baselines

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

**Tabela A.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 A.1 mostra a quantidade observada de cada tipo de transição 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 A.1 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 A.1 – Quantidade de transições entre os diversos estados possíveis na base de desenvolvimento do modelo multi-estado de Markov**