

**Insper Instituto de Ensino e Pesquisa  
Programa de Mestrado Profissional em Economia**

**Jorge Alfonso Marquez Sosa**

**PREVISÃO DE PERDAS FINANCEIRAS DE CONTRATOS DE  
CRÉDITO POR MEIO DE MODELOS MULTI-ESTADOS DE  
MARKOV**

**São Paulo  
2013**

Jorge Alfonso Marquez Sosa

**Previsão de Perdas Financeiras de Contratos de Crédito por meio  
de Modelos Multi-Estados de Markov**

Dissertação apresentada ao Programa de Mestrado Profissional em Economia do Insper Instituto de Ensino e Pesquisa, como parte dos requisitos para a obtenção do título de Mestre em Economia.

Área de concentração: Finanças e Macroeconomia Aplicadas  
Orientador: Prof. Dr. Rinaldo Artes – Insper

**São Paulo  
2013**

Sosa, Jorge Alfonso Marquez

Previsão de Perdas Financeiras de Contratos de Crédito por meio de Modelos Multi-Estados de Markov / Jorge Alfonso Marquez Sosa; orientador: Rinaldo Artes – São Paulo: Insper, 2013.

47 f.

Dissertação (Mestrado – Programa de Mestrado Profissional em Economia. Área de concentração: Finanças e Macroeconomia Aplicadas) – Insper Instituto de Ensino e Pesquisa.

1. Modelo Multi Estado 2. Crédito 3. Rentabilidade

# FOLHA DE APROVAÇÃO

Jorge Alfonso Marquez Sosa  
Previsão de Perdas Financeiras de Contratos de Crédito  
por meio de Modelos Multi-Estados de Markov

Dissertação apresentada ao Programa de Mestrado Profissional em Economia do Insper Instituto de Ensino e Pesquisa, como requisito parcial para obtenção do título de Mestre em Economia.

Área de concentração: Finanças e Macroeconomia Aplicadas

Aprovado em: Janeiro/2014

## Banca Examinadora

Prof. Dr. Rinaldo Artes  
Orientador

Instituição: Insper

Assinatura: \_\_\_\_\_

Prof<sup>a</sup>. Dra. Andrea Minardi  
Instituição: Insper

Assinatura: \_\_\_\_\_

Prof. Dr. Gustavo Henrique Pereira

Instituição: IME-USP

Assinatura: \_\_\_\_\_

## **DEDICATÓRIA**

Dedico este trabalho aos meus pais Luis Felipe e Ana Maria, fonte de orgulho e inspiração em minha vida. Pessoas que impulsionaram meu crescimento e que me fortaleceram nos momentos mais difíceis. Amizade sincera e amor intenso.

Dedico aos meus irmãos Luis, Viviana e Gabriela, por participarem na criação do meu senso crítico. Agradeço pelos encontros e reflexões sobre diversos pontos de nossas vidas e nossa sociedade. Exemplo de pessoas, pela honestidade, amor e respeito entre todos.

Dedico à minha namorada Karina e sua família, por entenderem e apoiarem minhas decisões. Agradeço por serem pessoas de grande carinho e de compromisso com a verdade.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, por ser meu grande amigo, entendedor de minhas alegrias e tristezas. Agradeço por sempre cuidar da saúde de meus familiares e amigos.

Agradeço a todos aqueles que de alguma forma fizeram parte deste trabalho, dando opiniões, críticas e elogios, motivando-me assim a realizar o melhor trabalho possível. Nestas condições encontra-se uma lista grande de pessoas, as quais faço questão de mencionar.

Aos meus superiores Peres, Eurico, Alessandro e Janaína por terem apoiado a ideia da realização deste Mestrado. Agradeço toda a atenção disponibilizada e o entendimento da importância deste curso para minha formação profissional.

Aos colegas de trabalho Karen, Paulo, André, Cristiane e Gláucia por darem todo o suporte técnico. Grandes conhecedores das aplicabilidades estatísticas no mundo financeiro.

Aos colegas do Mestrado, em especial ao Leandro Negrão, por compartilhar ideias, vitórias e desafios no decorrer do curso.

E por fim, ao grande mestre que aceitou prontamente a ideia desta dissertação e principal motivador das sutilezas deste trabalho, o Prof. Dr. Rinaldo Artes, pessoa que levarei com muito carinho em minhas recordações acadêmicas. Obrigado pelo apoio e carinho que tratou minha dissertação.

## RESUMO

SOSA, Jorge Alfonso Marquez. **Previsão de Perdas Financeiras de Contratos de Crédito por meio de Modelos Multi-Estados de Markov** 2013. 47 f. Dissertação (Mestrado) – Insper Instituto de Ensino e Pesquisa, São Paulo, 2013.

Com o crescimento do nível de bancarização e as projeções da continuidade da expansão do crédito, os bancos brasileiros devem buscar novas formas de manter sua vantagem competitiva, criando modelos estatísticos que melhorem a eficácia na tomada de decisão de conceder novos créditos. Neste trabalho será apresentada a modelagem por Análise de Sobrevivência, através de modelos Multi-Estados de Markov, com a finalidade de prever se um cliente trará prejuízos financeiros para a Organização. Sendo assim, uma nova conceituação de default será abordada: ao contrário de considerar apenas o default causado pela inadimplência, será levada em consideração a rentabilidade gerada pelo cliente, objetivando a maximização do lucro ao invés de apenas focar na diminuição do risco associado.

Palavras-chave: Análise de Sobrevivência; Modelo Multi-Estados de Markov; Rentabilidade

## ABSTRACT

SOSA, Jorge Alfonso Marquez. **Financial Loss Forecast of Credit Contracts by Multi-State Markov Models** 2013. 47 f.. Dissertation (Mastership) – Insper Instituto de Ensino e Pesquisa, São Paulo, 2013.

Considering the increased level of banking penetration and the forecast of continuing credit expansion, Brazilian banks must seek new ways to maintain their competitive advantage, building statistical models that improve the efficiency in decision to grant new loans. This work presents the modelling of Survival Analysis, through Multi-State Markov Models, in order to predict whether a customer will bring financial loss to the Organization. Thus, a new concept of default will be approached instead of considering only the default caused by delinquency. It will be taken into account the profitability generated by the client, in order to maximize profit rather than just focus on reducing the associated risk.

Keywords : Survival Analysis; Multi-State Markov Model; Profitability



## LISTA DE TABELAS

|  |    |
|--|----|
| Tabela 1 – Definição dos Estados dos Modelos de Rentabilidade .....    | 18 |
| Tabela 2 – Modelo de Rentabilidade Total .....                         | 20 |
| Tabela 3 – Modelo de Rentabilidade Boa .....                           | 20 |
| Tabela 4 – Índice KS das Bases Estudadas .....                         | 21 |
| Tabela 5 – Níveis de Discriminação (Índice KS) .....                   | 22 |
| Tabela 6 – Tempo Médio de Permanência – Quantidade de Produtos .....   | 24 |
| Tabela 7 – Tempo Médio de Permanência – Saldo Médio em Conta .....     | 25 |
| Tabela 8 – Tempo Médio de Permanência – Meses até Saldo Máximo .....   | 26 |
| Tabela 9 – Tempo Médio de Permanência – Comprometimento de Renda ..... | 27 |

## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 1 – Tipos de Censura à Direita .....                         | 8  |
| Figura 2 – Modelo Multi-Estados para Contratos de Crédito .....     | 12 |
| Figura 3 – Base de Modelagem .....                                  | 15 |
| Figura 4 – Estados de Transição .....                               | 17 |
| Figura 5 – Função de Sobrevivência – Quantidade de Produtos .....   | 23 |
| Figura 6 – Função de Sobrevivência – Saldo Médio em Conta .....     | 25 |
| Figura 7 – Função de Sobrevivência – Meses até Saldo Máximo .....   | 26 |
| Figura 8 – Função de Sobrevivência – Comprometimento de Renda ..... | 27 |

## SUMÁRIO

|   |    |
|---|----|
| <b>1 INTRODUÇÃO</b> .....               | 1  |
| <b>2 REVISÃO BIBLIOGRÁFICA</b> .....    | 3  |
| <b>3 METODOLOGIA</b> .....              | 6  |
| 3.1 MODELOS DE CREDIT SCORING .....     | 6  |
| 3.2 ANÁLISE DE SOBREVIVÊNCIA .....      | 7  |
| 3.3 COVARIÁVEIS .....                   | 10 |
| 3.4 MODELO DE REGRESSÃO DE COX .....    | 10 |
| 3.5 MODELO MULTI-ESTADO DE MARKOV ..... | 11 |
| 3.6 FERRAMENTAS UTILIZADAS .....        | 14 |
| <b>4 MODELAGEM</b> .....                | 14 |
| 4.1 PÚBLICO-ALVO .....                  | 14 |
| 4.2 PERÍODO DE AMOSTRAGEM .....         | 15 |
| 4.3 DEFINIÇÃO DOS ESTADOS .....         | 15 |
| 4.3.1 Rentabilidade .....               | 15 |
| 4.3.2 Estados .....                     | 17 |
| 4.4 VARIÁVEIS UTILIZADAS .....          | 18 |
| 4.4.1 Metodologia Empregada .....       | 19 |
| <b>5 RESULTADOS</b> .....               | 19 |
| 5.1 MODELOS FINAIS .....                | 19 |
| 5.2 AVALIAÇÃO DOS MODELOS .....         | 21 |
| 5.2.1 Kolmogorov-Smirnov .....          | 21 |
| 5.3 EFEITOS DAS VARIÁVEIS .....         | 23 |
| 5.3.1 Quantidade de Produtos .....      | 23 |
| 5.3.2 Saldo Médio em Conta .....        | 24 |
| 5.3.3 Meses até Saldo Máximo .....      | 26 |
| 5.3.4 Comprometimento de Renda .....    | 27 |
| 5.4 COMENTÁRIOS GERAIS .....            | 28 |
| <b>6 CONCLUSÃO</b> .....                | 29 |
| Referências .....                       | 30 |
| Apêndice .....                          | 33 |
| 1 Lista de Variáveis Estudadas .....    | 33 |
| 2 Exemplo Base de Dados .....           | 34 |
| 3 Matrizes de Migração .....            | 35 |
| 4 Código em R .....                     | 36 |

## 1 INTRODUÇÃO

Segundo dados do Banco Central do Brasil (BCB), desde 2003 o mercado de crédito no Brasil tem apresentado um crescimento acelerado. Diversos aspectos tais como alterações institucionais e fatores econômicos contribuíram para esse desenvolvimento. O principal fator econômico está relacionado à estabilização financeira que ocorreu desde a adoção do Plano Real, dando fim a altas taxas de inflação apresentada em décadas anteriores.

Associada ao aumento da renda nacional, os bancos puderam experimentar um crescimento acelerado no nível de bancarização, facilitando o acesso popular aos seus produtos e serviços. Segundo o Relatório de Bancarização e Inclusão Financeira no Brasil (Febraban, 2011), a rede de atendimento bancário cresceu aproximadamente 80% no período de 2006-2010. Com isso, as instituições financeiras gozaram de pleno desenvolvimento das linhas de crédito, tanto para recursos livres quanto para recursos direcionados. Alguns produtos contribuíram diretamente para o crescimento desse mercado, como o crédito consignado e o microcrédito. Segundo o relatório da Febraban, o crédito expandiu-se em mais de 130% no período referente a 2006-2010.

Mesmo com o expressivo crescimento ocorrido nas últimas décadas, de acordo com Andrade (2012, p. 73-74), há ainda espaço para a expansão de crédito no país. A participação do mercado de crédito tende a aumentar devido à redução das taxas de juros e à adoção do Cadastro Positivo – aprovado em 10 de junho de 2011, que tem por finalidade registrar todos compromissos financeiros e seus respectivos históricos de pagamentos, facilitando o acesso ao crédito e a oferta de melhores condições de negociação.

De acordo com o FMI (2012), os bancos brasileiros são sólidos e seguros, estando do ponto de vista operacional, preparados para um possível aumento da participação do crédito. Para isso, devem estar constantemente atualizados, mantendo suas políticas e modelos de crédito adequados a constantes mudanças na economia e mercado. Com a finalidade de manter sua vantagem competitiva, modelos estatísticos mais sofisticados devem ser desenvolvidos ou aprimorados, melhorando a eficiência na tomada de decisão.

Da mesma forma, novos conceitos e técnicas devem ser utilizados na análise de crédito de modo a torná-la mais efetiva. Utilizar a previsão de rentabilidade do cliente pode ser um grande indicador de apoio à decisão de concessão de crédito, uma vez que a principal preocupação de uma instituição financeira é quanto de retorno o cliente irá trazer para a organização.

Considerar como principal objetivo a maximização do lucro além da diminuição do risco de inadimplência pode trazer às instituições financeiras uma nova forma de conceituar o critério de default nos modelos de crédito. O conceito mais difundido refere-se especialmente ao atraso em operações financeiras. Desse modo, quando o atraso é superior a um determinado patamar de tempo, em um horizonte de tempo específico, o cliente é considerado default. Esse é um critério comum na maioria dos bancos nacionais, preocupando-se especialmente com o risco do cliente tornar-se inadimplente.

O grande diferencial deste trabalho refere-se a uma nova conceituação de default: ao invés de modelar apenas o default causado pela inadimplência, será levada em consideração a rentabilidade gerada pelo cliente. Isso porque alguns clientes, mesmo em default por atraso, mostram-se rentáveis, pois as receitas adquiridas com multas e juros de atraso podem compensar as demais perdas. Adicionalmente, clientes considerados não-default, quando liquidam suas operações antecipadamente em um prazo curto, as receitas advindas dos juros de empréstimo podem não compensar os custos obtidos.

Portanto, é fundamental acompanhar a vida do contrato, especialmente quando começa a apresentar níveis de perdas elevadas. Classificá-los como rentáveis ou não-rentáveis é uma forma de tornar a decisão de crédito mais eficaz. Prever o tempo que um contrato leva até tornar-se prejuízo pode ser um indicador de performance adequado para análise de portfólio.

Os modelos de análise de sobrevivência atendem a esses requisitos pois tratam de processos estatísticos nos quais a variável de interesse é o tempo que decorre até a ocorrência de um evento. Pelo fato de seu uso ainda não ser totalmente difundido no meio financeiro, há uma série de fatores que precisam ser estudados. Encontrar variáveis e produtos que mais se adequem a este tipo de modelagem é um obstáculo a ser enfrentado.

Portanto, o grande desafio e a motivação deste trabalho será a construção de modelos que permitam utilizar a rentabilidade como apoio a tomada de decisão de crédito, objetivando a maximização do lucro ao invés de apenas focar na diminuição do risco associado à operação.

A estrutura da dissertação contempla além desta introdução, um capítulo relacionado a revisão bibliográfica, referenciando alguns trabalhos que já foram realizados no meio acadêmico, seja nas questões que tangem a rentabilidade quanto da utilização dos modelos de Análise de Sobrevivência. No capítulo seguinte, é apresentada a metodologia utilizada, conceituando o que são os modelos de credit scoring e aprofundando as questões referentes à análise de sobrevivência, como os modelos de Regressão de Cox e por fim os Modelos Multi-Estados de Markov, base para desenvolvimento deste trabalho.

No capítulo quatro, definem-se os fatores de risco utilizados nos modelos, especificando o público-alvo e o período de amostragem. No capítulo cinco são apresentados os resultados obtidos para dois modelos distintos de Análise de Sobrevivência com suas respectivas medidas de desempenho. Por fim todas as conclusões acerca dos modelos e do processo de modelagem são apresentadas, destacando-se os resultados obtidos e os comparativos com os modelos desenvolvidos.

## **2 REVISÃO BIBLIOGRÁFICA**

Um ponto fundamental da análise de concessão de crédito é a avaliação do risco associado ao descumprimento dos pagamentos. Desse modo, tanto os estudos pontuais de concessão de crédito – como ações de pré-aprovação – quanto a modelagem de crédito, levam em consideração a probabilidade do cliente tornar-se inadimplente. No entanto, a principal preocupação de uma instituição financeira é saber quanto de retorno o cliente trará para a organização. Dessa forma, atrelar as ações comerciais e/ou estratégias do negócio à rentabilidade, podem trazer resultados mais satisfatórios para a organização do que simplesmente avaliar o risco do cliente.

A mudança nas estratégias de concessão de crédito com foco na rentabilidade tem sido discutida no meio acadêmico. Silva, Ribeiro e Sheng (2011) utilizaram a rentabilidade como indicador na tomada de decisão de crédito e verificaram que conceder crédito para clientes com baixo risco nem sempre é a opção mais rentável, pois os retornos advindos desses clientes podem ser baixos se comparados com o retorno de investimento dos demais clientes. O contrário também se mostrou verdadeiro, pois por muitas vezes, clientes de alto risco ofereciam rentabilidade maior, compensando eventuais perdas que eles poderiam ocasionar. Dessa forma, concluíram que a decisão de concessão de crédito não deve apenas considerar o risco envolvido, mas também a rentabilidade que esse cliente pode oferecer.

Banasik, Crook & Thomas (2001) se preocuparam com a questão da rentabilidade, criando modelos que associam o risco de inadimplência com o uso do cartão de crédito. Assim, descobriram que a acurácia da previsão dos modelos de concessão poderia ser melhorada justamente pelo fato da rentabilidade estar atrelada ao uso desse produto.

Dey (2010) avaliou a rentabilidade de cartões de crédito verificando a influência dos diversos fatores que compõem o risco do produto, desde a probabilidade do cliente encerrar sua conta até o uso de seu limite de crédito. Assim, propôs um método que combina efeitos de gestão de limites de crédito com ações de *pricing* de modo a maximizar a rentabilidade que o cliente gera.

Juntamente com Stepanova, Thomas (2002) utilizou ferramentas de Análise de Sobrevivência para construir modelos de credit scoring que avaliam aspectos da rentabilidade tão bem quanto do default por inadimplência. O grande diferencial encontrava-se no fato de que alguns clientes, mesmo em default, eram rentáveis, pois os juros adquiridos com o atraso puderam compensar as perdas obtidas. Adicionalmente, alguns clientes considerados não-default, liquidavam suas operações em um prazo mais curto que o previsto, fazendo com que os juros do empréstimo não compensassem os custos obtidos.

Segundo Andreeva (2006), as aplicações de análise de sobrevivência em Credit Scoring se iniciaram com Narain (1992), mostrando que as estimativas do tempo de vida de um empréstimo obtido a partir de um modelo exponencial podem melhorar significativamente as decisões de crédito.

Finlay (2008) comparou alguns modelos de credit scoring utilizando duas definições distintas de bom/mau cliente – a definição tradicional baseada na inadimplência e uma definição atrelada à rentabilidade do cliente. Sua grande preocupação estava no fato de que a principal meta de uma área comercial é a geração de lucro, independente da taxa de default do produto. O autor enumera uma série de motivos para se utilizar a modelagem por rentabilidade ao invés da classificação de bom/mau pagador convencional. Em contrapartida, enumera também os motivos pelo qual a modelagem por rentabilidade não é tão popular, como por exemplo, a dificuldade de se calcular de maneira precisa a rentabilidade que um produto como o cartão de crédito gera, devido a quantidade enorme de transações realizadas, cada uma com sua taxa de juros e despesas específicas.

Régis (2007) aplicou a análise de sobrevivência para prever mudanças de estado no ciclo de vida do uso de cartão de crédito. Os estados correspondiam ao status do cartão, como: em dia, rotativo, em atraso ou cancelado. Por meio de modelos Multi-Estados de Markov – utilizados quando há a presença de mudanças sucessivas de estados – gerou escores com a finalidade de confrontar os resultados com modelos de regressão logística, já que estes são os mais utilizados no ambiente financeiro. Assim, criou comparativos para modelos de escore de default, escores de cancelamento, escores de atraso e escores de utilização do crédito rotativo.

Bellotti & Crook (2009) também aplicaram modelos de análise de sobrevivência em contas de cartão de crédito. Eles verificaram a possibilidade de incluir variáveis macroeconômicas, como taxa de juros de mercado e taxa de desemprego, com a finalidade de melhorar o desempenho de seus modelos.

Andreeva (2006) constatou que em suas aplicações, a Análise de Sobrevivência e a Regressão Logística apresentaram resultados similares, no entanto a escolha do modelo de Análise de Sobrevivência pôde trazer alguns benefícios adicionais, tornando-o mais atrativo do que o Modelo de Regressão Logística. Determinar o tempo até a ocorrência de um evento de default é um desses atrativos, pois através deste parâmetro é possível construir escores de rentabilidade.

Conforme constatado, diversos autores realizaram comparações entre os modelos de Regressão Logística e modelos de Análise de Sobrevivência. Os Modelos de Análise de Sobrevivência por muitas vezes se mostraram superiores, por essa razão, será o foco deste trabalho.



### 3 METODOLOGIA

#### 3.1 MODELOS DE CREDIT SCORING

A aplicação de modelagem estatística é fundamental para o mercado de crédito, seja para selecionar os melhores clientes, saber precificar um produto, adequar os limites de crédito ou ainda auxiliar a área de cobrança a realizar ações mais direcionadas. Todas essas aplicações são resultado do uso de um arcabouço de técnicas, desde uma análise estatística descritiva a modelos de crédito mais complexos.

A análise de agrupamentos para seleção de clientes, árvores de decisões para análise de políticas de crédito e a utilização de modelos estatísticos a fim de discriminar clientes de diferentes perfis são algumas das técnicas empregadas. Visam assegurar sobretudo uma análise mais sensível dos clientes, reduzindo os riscos envolvidos e possibilitando aumento da rentabilidade.

Há uma série de modelos que podem ser empregados pelas instituições financeiras. Muitas delas inclusive utilizam mais de um modelo. A FDIC<sup>1</sup> (2007) listou em um de seus relatórios de Proteção ao Consumidor, alguns tipos de modelos de score utilizados no mercado, conforme traduzidos abaixo:

*“- Modelos de Application Scoring: são utilizados para avaliar clientes com potencial de crédito, classificando-os segundo seu risco de inadimplência. As informações disponíveis para o desenvolvimento desses modelos referem-se à data de solicitação de crédito pelos clientes.*

*- Modelos de Attrition Scoring: auxiliam na identificação dos clientes mais prováveis a encerrarem suas contas.*

*- Modelos de Bankruptcy Scoring: ajudam a prever os emprestadores mais prováveis a declarar falência.*

*- Modelos de Behavior Scoring: utilizam informações comportamentais do cliente para uma melhor gestão de seus produtos.*

*- Modelos de Collection Scoring: são usados para determinar estratégias de cobrança.*

---

<sup>1</sup> FDIC (Federal Deposit Insurance Corporation) é uma agência independente, criada pelo congresso americano para manter a estabilidade e a confiança pública do sistema financeiro dos Estados Unidos da América.

- ***Modelos de Fraud Detection Scoring:** identificam clientes com potencial de atividades fraudulentas.*
- ***Modelos de Payment Project Scoring:** avaliam o percentual do balanço que é mais provável a ser repago.*
- ***Modelos de Recovery Scoring:** auxiliam na recuperação de crédito após atuação da área de cobrança.*
- ***Modelos de Response Scoring:** são utilizados para gerenciar custos de aquisição, especialmente em campanhas de marketing.*
- ***Modelos de Revenue Scoring:** prediz o potencial de receita gerada por uma nova conta durante os primeiros meses de vida.”*

Dentre os modelos estatísticos, os mais utilizados para avaliação de crédito são o *Application Scoring* e o *Behavior Scoring*, diferenciando-se apenas por sua empregabilidade. Enquanto os modelos de *Application Scoring* são utilizados para a avaliação de crédito de um cliente novo, os modelos de *Behavior Scoring* são empregados para clientes que já possuem algum relacionamento com a organização.

### 3.2 ANÁLISE DE SOBREVIVÊNCIA

De acordo com Colosimo & Giolo (2006), os modelos de análise de sobrevivência são bastante utilizados na área da medicina, cuja principal preocupação, em especial de indústrias farmacêuticas, é o tempo que demora para um paciente curar ou falecer, dependendo do critério de default adotado. Esta variável resposta de tempo até a ocorrência do evento é conhecido como **tempo de falha**.

Para se conceituar tempo de falha, três elementos são essenciais: o tempo inicial, a escala de medida do tempo e a falha. Para modelos de crédito, o **tempo inicial** refere-se ao início do acompanhamento do contrato; a **escala de medida** – tempo decorrido – é realizada em dias, visto que as taxas de juros de financiamento são calculadas diariamente; e a **falha** – o acontecimento de interesse – é a inadimplência ou prejuízo do cliente. Portanto, pode-se

considerar tempo de falha o tempo que um cliente leva para se tornar inadimplente ou causar prejuízo.

Outro aspecto importante na análise de sobrevivência que a diferencia de outras metodologias é a presença de **censura**. Em muitas situações o tempo de sobrevivência de alguns indivíduos não é conhecido, ou seja, não é observado o evento de interesse durante o período analisado. Na análise de sobrevivência, estes indivíduos podem ser considerados como observações censuradas, aumentando potencialmente a qualidade do modelo estimado.

A censura pode ser classificada como censura à direita, censura à esquerda ou censura intervalar. A censura à direita é a mais comum em análise de sobrevivência e aparece quando ao finalizar o estudo, o evento de interesse ainda não ocorreu. Já a censura à esquerda acontece quando o evento ocorreu antes do período de observação, mas não se sabe exatamente quando. A censura à direita pode ser classificada em vários tipos (MILLER, 1981), conforme figura 1.

- tipo I: ocorre quando existe um período pré-determinado de tempo ( $t_c$ ) a partir do qual todos os dados estão censurados;
- tipo II: ocorre quando o estudo é finalizado ao se atingir um número pré-determinado de eventos de interesse;
- tipo III ou censura aleatória: quando a censura não está relacionada com o tempo de sobrevivência e ocorre de forma aleatória.

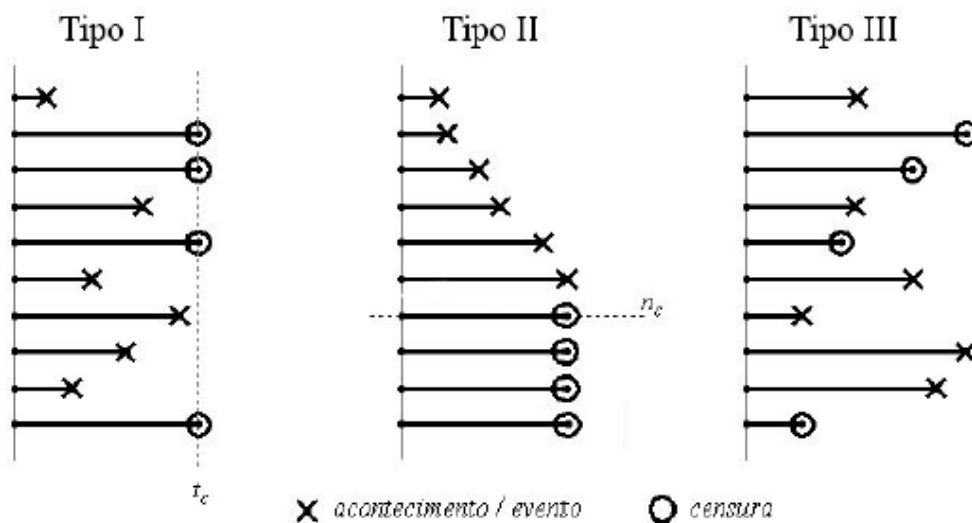


Figura 1 – Tipos de Censura à Direita (Fonte: CORREIA, 2007)

Na prática bancária da área de Crédito, não há mais o acompanhamento do desempenho do contrato quando o cliente liquida ou renegocia sua operação, ou ainda, quando seu contrato é vendido para outra instituição financeira, caracterizando censura à direita. Do mesmo modo, dependendo do produto e da quantidade de meses observados, não se consegue estabelecer o instante exato que ocorreu o evento de interesse. Dessa forma, uma base de dados com essas características, pode ser caracterizada como censura do tipo intervalar. No entanto, mesmo censurados, utiliza-se a informação de que, até aquele momento, o cliente era adimplente, pois mesmo incompletas, as informações censuradas oferecem elementos relevantes para a análise. Além do mais, a omissão das censuras pode acarretar vícios nas estimativas dos parâmetros dos modelos.

A probabilidade de um indivíduo sobreviver (não se ter observado o evento de interesse) até determinado tempo  $t$  é caracterizada pela Função de Sobrevivência  $S(t)$ , conforme estabelecida abaixo (mais detalhes em COLOSIMO & GIOLO, 2006):

$$S(t) = P[T \geq t] = 1 - F(t) = \int_t^{\infty} f(t)dt, \quad (3.1)$$

Em que  $T$  é a variável aleatória que indica o tempo até a ocorrência do evento de interesse (falha),  $f(t)$  é sua função densidade de probabilidade (*fdp*):

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (3.2)$$

$F(t)$  é sua função de distribuição acumulada e  $S(t)$  é uma função monótona decrescente com  $S(0) = 1$  e  $\lim_{t \rightarrow \infty} S(t) = 0$ , denominada Função de Sobrevivência.

Conforme estabelecido, falha é quando o evento de interesse – como inadimplência ou prejuízo do cliente – ocorre. A função  $h(t)$ , definida em (3.3) é denominada Função de Taxa de Falha ou Risco, sendo interpretada como o risco instantâneo de falha para um indivíduo que tenha sobrevivido até o instante  $t$ :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T \leq t + \Delta t | T \geq t]}{\Delta t} = \frac{f(t)}{S(t)}. \quad (3.3)$$

Há diferentes formas de modelar a função de risco ou a função de sobrevivência: modelos não-paramétricos, paramétricos e semiparamétricos. Os modelos não-paramétricos não assumem que a distribuição de T seja conhecida; o estimador de Kaplan-Meier (ver em CORREIA, 2007, p. 25-29) para a função de sobrevivência pode ser utilizado nesta abordagem. Nos modelos paramétricos, a distribuição de probabilidades de T é conhecida e seus parâmetros são modelados (ver em FERNANDES, 2010, p. 11-14); as distribuições mais comuns são Exponencial, Weibull e Log-normal. Por fim, os modelos semiparamétricos ou Modelos de Cox são flexíveis pois, apesar de não pressupor uma distribuição de probabilidades de T, possui um componente paramétrico (COX, 1972), conforme será apresentado na seção 3.4.

### **3.3 COVARIÁVEIS**

Em geral, na área de Crédito, ao se aplicar Análise de Sobrevivência busca-se a construção de modelos que permitam a incorporação do efeito de uma série de fatores na previsão do tempo de falha de um cliente. Dentre esses fatores destacam-se características cadastrais (idade, sexo, etc.), comportamentais (BANASIK, CROOK e THOMAS, 2001 – associaram o risco de inadimplência com o uso do cartão de crédito, por exemplo) ou até mesmo outras variáveis exógenas (BELLOTTI & CROOK, 2009 – verificaram a possibilidade de incluir em seus modelos variáveis macroeconômicas, como taxa de juros de mercado e taxa de desemprego). Essas variáveis explicativas, covariáveis, podem ser dependentes do tempo ou constantes.

### **3.4 MODELO DE REGRESSÃO DE COX**

O Modelo de Regressão de Cox, ou simplesmente Modelo de Cox, incorpora covariáveis na modelagem dos tempos até a ocorrência de um evento de interesse – prejuízo financeiro, por exemplo – por meio da função de taxa de falha. Neste caso, a função de taxa de falha de um indivíduo é dada por:

$$h(t|x_k) = h_0(t) \cdot \exp\{\beta^T x_k\} \quad (3.4)$$

sendo  $h_0(t)$  uma função não-negativa do tempo (função de base – baseline),  $\beta = (\beta_1, \dots, \beta_p)^T$  o vetor de coeficientes de regressão e  $x_k$  o vetor de covariáveis associadas ao indivíduo  $k$ . Os dois componentes multiplicativos são de natureza distinta:  $h_0(t)$  é um componente não-paramétrico, enquanto o outro é um componente paramétrico. Por esse motivo o modelo é denominado semi-paramétrico.

Para estimação dos parâmetros, Cox (1975) propôs o Método da Máxima Verossimilhança Parcial, que sob condições de regularidade bastante fracas, leva a estimadores com as propriedades assintóticas usuais de inferência baseada na verossimilhança. Assim, obtêm-se a estimação dos parâmetros através da maximização da Função de Verossimilhança Parcial, estabelecida a seguir:

$$L(\beta) = \prod_{i=1}^n \left( \frac{\exp(x_i^T \beta)}{\sum_{j \in R(t_i)} \exp(x_j^T \beta)} \right)^{\delta_i}, \quad (3.5)$$

na qual  $R(t_i)$  é o conjunto dos índices das observações sob risco no tempo  $i$  e  $\delta_i$  é o indicador de falha para o indivíduo  $i$ .

Os estimadores obtidos por esse método são, sob condições de regularidade, consistentes e assintoticamente normais (COX, 1975).

### 3.5 MODELO MULTI-ESTADO DE MARKOV

Os Modelos Multi-Estados de Markov são considerados uma generalização dos modelos de Análise de Sobrevivência (FERNANDES, 2010). São mais comumente utilizados quando há presença de mudanças sucessivas nas condições do indivíduo (estados) ao longo do tempo. Por exemplo: cliente em dia, cliente em atraso e cliente em default (este último é denominado

absorvente, já que ao atingi-lo não é facultado ao cliente o retorno a qualquer um dos estados anteriores).

A complexidade desse modelo depende da quantidade de estados e das possíveis transições que podem ocorrer entre eles. Avaliando o comportamento de um contrato de crédito, pode-se verificar a alternância de seu estado, podendo o contrato estar em dia, adiantado, em atraso ou até mesmo lançado a perda, conforme pode ser visto na figura a seguir (as setas indicam as possíveis transições de estados):

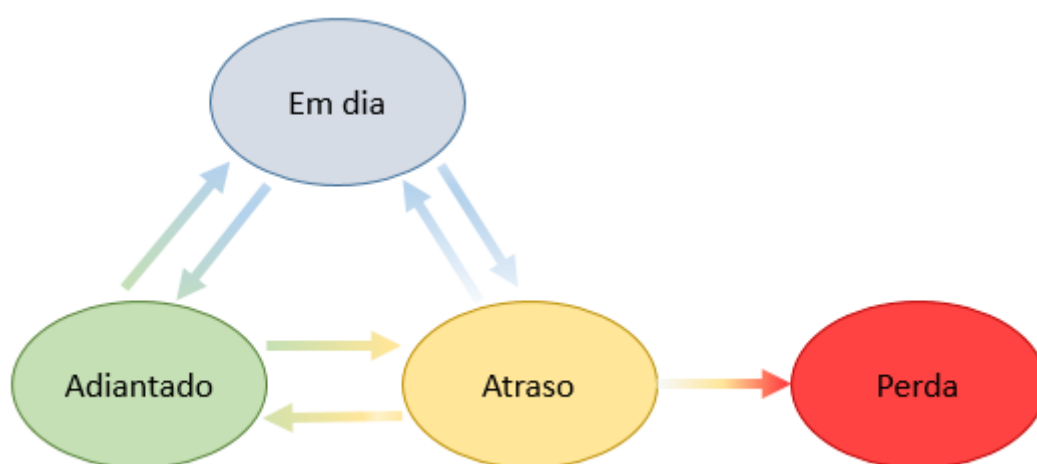


Figura 2 – Modelo Multi-Estado para Contratos de Crédito

Os processos multi-estados são caracterizados por intensidades de transição entre os estados, seguindo a formulação (3.6) que coincide com (3.3) – Em (3.6),  $T_{khj}$  representa o tempo que o indivíduo  $k$  leva para passar do estado  $h$  ao estado  $j$  e  $\lambda_{khj}$  é a função de risco de mudança do estado  $h$  ao estado  $j$ , para o indivíduo  $k$ , representando o risco instantâneo do indivíduo migrar do estado  $h$  para o estado  $j$ .

$$\lambda_{hj}(t, z(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (3.6)$$

Suponha que um processo seja caracterizado por possuir  $H$  estados. Dessa forma, o processo multi-estado pode ser caracterizado pela Matriz de Intensidade de Transição – que contempla

todas as possíveis transições do indivíduo – conforme estabelecido a seguir (omitindo-se o índice para indivíduos):

$$\Lambda(t) = \begin{pmatrix} \lambda_{11}(t) & \lambda_{12}(t) & \dots & \lambda_{1(H-1)}(t) & \lambda_{1H}(t) \\ \lambda_{21}(t) & \lambda_{22}(t) & & \lambda_{2(H-1)}(t) & \lambda_{2H}(t) \\ & \vdots & \ddots & \vdots & \\ \lambda_{(H-1)1}(t) & \lambda_{(H-1)2}(t) & \dots & \lambda_{(H-1)(H-1)}(t) & \lambda_{(H-1)H}(t) \\ \lambda_{H1}(t) & \lambda_{H2}(t) & & \lambda_{H(H-1)}(t) & \lambda_{HH}(t) \end{pmatrix}_{H \times H} \quad (3.7)$$

Por meio da Matriz Intensidade de Transição  $\Lambda(t)$  pode-se encontrar a Matriz de Probabilidades de Transição  $P(t)$  (ver detalhes em ARASHIRO, 2008), escrita em função de  $\Lambda(t)$ , conforme apresentado abaixo:

$$P(t) = \exp(\Lambda(t)) \quad (3.8)$$

Esta matriz traz a probabilidade de um indivíduo que está em um estado representado pela linha  $i$  atingir o estado representado na coluna  $j$  após um tempo  $t$ .

Marshall e Jones (1995) propõe que  $\lambda_{khj}$  seja modelada de modo semelhante ao modelo semiparamétrico de Cox. Assim,

$$\lambda_{khj}(t | x_{kit}) = \lambda_{0hj} \exp(\beta^T x_{kit}) \quad (3.9)$$

sendo  $\lambda_{0hj}$  a função baseline para a transição entre os estados  $h$  e  $j$  e  $x_{kit}$  é o vetor de covariáveis para o indivíduo  $k$ , observado no instante  $t$ . Detalhes do procedimento de estimação deste modelo podem ser encontrados em Marshall e Jones (1995) e Jackson (2011).



### **3.6 FERRAMENTAS UTILIZADAS**

Para realização deste estudo serão utilizados dois softwares distintos: R e SAS. No primeiro, será utilizado o pacote MSM (JACKSON, 2011) nas análises referentes aos Modelos Multi-Estados de Markov, calculando a Matriz de Intensidade de Transição, Matriz de Probabilidades de Transição e o tempo médio que um indivíduo demora para mudar seu estado. No apêndice 3 pode ser encontrado o código do programa em R.

A base de dados será estruturada em SAS e servirá como apoio para a finalização do estudo, através dos cálculos dos indicadores estatísticos de Modelos.

## **4. MODELAGEM**

Os dados analisados são provenientes de uma grande instituição financeira. Por questões relacionadas ao sigilo da informação, não será identificada nenhuma variável que comprometa as políticas e estratégias de crédito dessa organização.

### **4.1 PÚBLICO-ALVO**

Ao segmentar os clientes de uma instituição financeira, é notória a presença de três públicos distintos: clientes de baixa renda, que apresentam elevadas taxas de inadimplência; clientes de média renda, que conseguem pagar suas dívidas, no entanto qualquer instabilidade financeira leva à inadimplência; e clientes de alta renda, cuja inadimplência é pouco afetada por fatores externos.

Neste trabalho, o público selecionado é o de média renda, justamente por ser o público cuja inadimplência é mais difícil de prever, visto que é muito susceptível a eventualidades. Dentro deste grupo, foram selecionados apenas clientes que já possuíam relacionamento com a

organização em estudo desde janeiro de 2009, com a finalidade de utilizar variáveis comportamentais nos modelos criados. Por fim, somente clientes que solicitaram um produto específico de crédito em qualquer período entre janeiro de 2010 e junho de 2013 foram observados.

## 4.2 PERÍODO DE AMOSTRAGEM

As bases de dados foram geradas considerando o período de janeiro de 2009 a junho de 2013. O período que compreende entre janeiro de 2009 a dezembro de 2009 foi utilizado apenas para obtenção dos valores iniciais das variáveis comportamentais. Duas amostras distintas com vinte mil clientes foram utilizadas: uma para estimar o modelo (Base de Desenvolvimento) e a outra para validá-lo (Base de Validação).



Figura 3 – Base de Modelagem

## 4.3 DEFINIÇÃO DOS ESTADOS

### 4.3.1 Rentabilidade

O principal motivador de utilizar a rentabilidade ao invés do atraso da operação na definição dos estados, deve-se ao fato de existirem clientes que por algum momento de seu contrato apresentaram diversos atrasos, no entanto conseguiram se reestabelecer e voltaram a ficar em

dia. Estes clientes poderiam, eventualmente, ser classificados como default, no entanto são clientes que trazem rentabilidade para a organização, visto que além de gerar juros de financiamento, apresentam também como fonte de receita, as multas e juros de atraso.

Neste trabalho, as receitas são classificadas de duas formas distintas: receita boa e receita ruim. A receita boa é aquela advinda dos juros de financiamento e descontos. A receita ruim é a proveniente de multas, juros de atraso e recuperação. Distinguem-se essas duas receitas para verificar se há possibilidade de criar modelos mais precisos quanto da aplicação do critério de rentabilidade.

São consideradas duas fontes distintas de despesas: provisão e perdas. Os demais custos relacionados ao negócio, como custos de contratação e impostos não serão considerados no estudo. As regras de provisão e perda seguiram a Resolução 2.682 (1997) do Banco Central.

A rentabilidade total observada no instante  $T$  considera todas as receitas e despesas acumuladas desde o início do contrato até o instante referido, sobre o saldo inicial contratado. Caso o cliente possua mais de um contrato, considera-se a soma das receitas e despesas de todos os contratos sobre a soma do saldo inicial de seus contratos, conforme fórmula (4.1), no qual  $t$  é medido em meses:

$$Rentabilidade\ Total_T = \sum_{t=0}^T \frac{Receitas\ Total_t - Despesas_t}{Saldo\ Inicial\ Contratado} \quad (4.1)$$

A Rentabilidade Boa no instante  $T$  é definida em (4.2). Neste caso, somente serão consideradas as receitas proveniente dos juros de financiamento e descontos (receita boa).

$$Rentabilidade\ Boa_T = \sum_{t=0}^T \frac{Receitas\ Boa_t - Despesas_t}{Saldo\ Inicial\ Contratado} \quad (4.2)$$

#### 4.3.2 Estados

As transições são definidas considerando a estratégia do negócio como premissa. Dessa forma, quatro estados são estabelecidos tanto para o modelo de Rentabilidade Total como para o modelo de Rentabilidade Boa, conforme a figura a seguir:

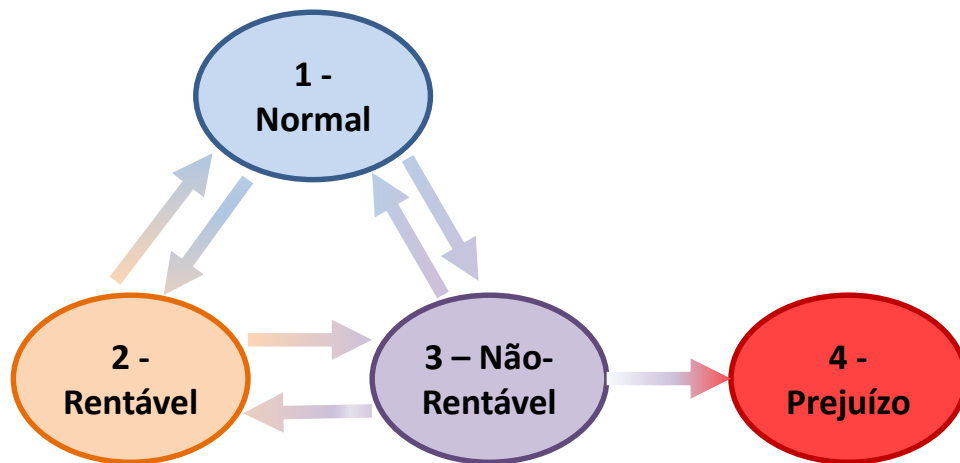


Figura 4 – Estados de Transição

São considerados os seguintes estados:

1. Normal: clientes que possuem contratos de rentabilidade aceitável;
2. Rentável: clientes que possuem contratos com rentabilidade acima do esperado;
3. Não-Rentável: clientes que possuem contratos com rentabilidade abaixo do esperado
4. Prejuízo: clientes que possuem contratos com rentabilidade inferior ao limite mínimo do produto e sem perspectiva de recuperação.

| Estado     |  | Modelo de Rentabilidade                            |  |
|------------|--|--|--|
|            |  | Rentabilidade Total                                | Rentabilidade Boa                                  |
| Recorrente | 1 - Normal (N)   | Rentabilidade Aceitável                            | Rentabilidade Aceitável                            |
|            | 2 - Rentável (R)                                       | Rentabilidade acima do esperado                    | Rentabilidade acima do esperado                    |
|            | 3 - Não-Rentável (NR)                                  | Rentabilidade abaixo do esperado                   | Rentabilidade abaixo do esperado                   |
| Absorvente | 4 – Perda Financeira Relevante (D) (Estado Absorvente) | Rentabilidade inferior ao limite mínimo do produto | Rentabilidade inferior ao limite mínimo do produto |

Tabela 1 – Definição dos Estados dos Modelos de Rentabilidade

É possível que um cliente transite entre os estados 1, 2 e 3 (vide as flechas da Figura 4). Já o estado 4 é absorvente. Neste último estado, o cliente já perdeu mais de 30% do valor emprestado. Esta faixa foi estabelecida para ser comparável aos modelos que conceituam default devido à inadimplência, que consideram um atraso superior a 90 dias. Quando o cliente atinge esse patamar de atraso, segundo a Resolução 2.682 do Banco Central, provisiona-se 30% do saldo devedor ativo. O resultado referente às matrizes de transição encontram-se no Apêndice 3.

#### 4.4 VARIÁVEIS UTILIZADAS

Uma base de 45 variáveis foi gerada com informações cadastrais e comportamentais do cliente e de seu contrato específico. Para tornar a base de mais fácil manuseio, realizou-se uma análise econômica, eliminando-se variáveis não relacionadas ao estudo. Em seguida, verificou-se a quantidade de valores ausentes, eliminando aquelas com número excessivo de observações faltantes. Por fim, realizou-se uma análise de estabilidade das variáveis, eliminando aquelas que apresentaram comportamentos inadequados, sugerindo a existência de inconsistências na

base de dados. Após a realização desses procedimentos, foram pré-selecionadas quatorze variáveis, as quais estão descritas nos Apêndices 1 e 2.

#### 4.4.1 Metodologia Empregada

As variáveis pré-selecionadas passaram por um processo de categorização, avaliando a taxa de default (percentual de clientes que migraram para o estado de perda financeira relevante em até 12 meses) dentro de cada categoria e seu respectivo WOE (*Weight Of Evidence* – ver em THOMAS, 2009, por exemplo). Assim, os modelos foram construídos utilizando-se as variáveis categorizadas. O método *Backward* foi utilizado para selecionar modelos que apresentassem menores valores de AIC (*Akaike Information Criteria*). O método *Backward*, bem como os critérios de WOE e AIC podem ser vistos em Mason, Gunst & Hess, 2003.

Foram selecionados cinco modelos das bases de desenvolvimento que apresentaram menor AIC, aos quais foram avaliados segundo o índice de KS (*Kolmogorov-Smirnov*), que avalia o quanto um modelo diferencia o “bom” do “mau” cliente (mais detalhes em THOMAS, 2009).

## 5 RESULTADOS

### 5.1 MODELOS FINAIS

Após realização do processo de seleção de modelos, os melhores modelos encontrados são apresentados a seguir:

| <b>Modelo Rentabilidade Total</b> |  |                      |  |
|-----------------------------------|--|----------------------|--|
| <b>Rótulo</b>                     | <b>Variável</b>  | <b>Nº categorias</b> | <b>Descrição</b>   |
| QTD_PRODUTOS                      | Qtde. de produtos negociados pelo cliente              | 5                    | Quanto maior a categoria, mais produtos são negociados         |
| VR_SALDO_M_CONTA                  | Saldo médio em conta corrente nos últimos 12 meses     | 5                    | Quanto maior a categoria, maior é o saldo em conta             |
| C_MESES_SALDO_MAX                 | Qtde. de meses desde o saldo devedor máximo do cliente | 5                    | Quanto maior a categoria, mais meses se afasta do saldo máximo |
| COMPR_RENDA                       | Comprometimento de renda nos últimos 12 meses          | 5                    | Quanto maior a categoria, maior é o comprometimento de renda   |

Tabela 2 – Modelo de Rentabilidade Total

| <b>Modelo Rentabilidade Boa</b> |   |                      |   |
|---------------------------------|---|----------------------|---|
| <b>Rótulo</b>                   | <b>Variável</b>   | <b>Nº categorias</b> | <b>Descrição</b>  |
| TEM_LIMITE                      | Identifica se cliente possui limites de crédito           | 2                    | Dummy se possui limites de crédito  |
| AGR_TEM_CARTAO                  | Identifica se cliente possui cartão                       | 2                    | Dummy se possui cartão de crédito   |
| VR_SALDO_M_CONTA                | Saldo médio em conta corrente nos últimos 12 meses        | 5                    | Quanto maior a categoria, maior é o saldo em conta                        |
| RECEBE_CONTA                    | Identifica se recebe em conta                             | 2                    | Dummy se recebe em conta  |
| M_COMPORT                       | Meses na base de comportamento                            | 5                    | Quanto maior a categoria, mais meses se encontra na base de comportamento |
| PC_UTLZ_CARTAO                  | % de Utilização de cartão de crédito nos últimos 12 meses | 5                    | Quanto maior a categoria, maior é o % de utilização                       |
| TEMPO_CLI                       | Tempo de relacionamento com o cliente                     | 5                    | Quanto maior a categoria, maior é o tempo de relacionamento               |

Tabela 3 – Modelo de Rentabilidade Boa

## 5.2 AVALIAÇÃO DOS MODELOS

O Índice KS (Kolmogorov-Smirnov) foi utilizado para avaliar os modelos, uma vez que é extensivamente utilizado no mercado financeiro. Para privilegiar a identificação de indivíduos com grande probabilidade de gerarem perdas financeiras relevantes, o processo de criação de escore foi aplicado apenas às probabilidades de transições para este estado. Assim, criou-se uma padronização das pontuações que variam de 0 a 1000, sendo “0” um cliente com 100% de probabilidade de migrar para o estado referente a perdas financeiras relevantes, e “1000” o cliente que apresenta 0% de probabilidade da ocorrência da mesma migração. O período de análise selecionado é de 12 meses de observação, prática comum nos modelos de crédito. Para os dois modelos serão analisados os indicadores estatísticos tanto para a base de modelagem quanto a de validação.

### 5.2.1 Kolmogorov-Smirnov

A Tabela 4 traz os índices KS dos dois modelos para as amostras de desenvolvimento e validação:

| Modelo              | Kolmogorov-Smirnov |           |
|---------------------|--------------------|-----------|
|                     | Desenvolvimento    | Validação |
| Rentabilidade Total | 40,10%             | 38,84%    |
| Rentabilidade Boa   | 29,45%             | 28,50%    |

Tabela 4 – Índice KS das Bases Estudadas

O valor de KS pode variar de 0% a 100%, sendo que o poder de discriminação aumenta à medida que se aproxima de 100%. Para efeitos comparativos, na Tabela 5 apresenta-se os níveis de discriminação por faixas de valores de KS conforme sugerido em Andrade e Oliveira (2002).



| Valores de KS | Nível de Discriminação    |
|---------------|---------------------------|
| Abaixo de 20% | Baixa discriminação       |
| De 20% a 30%  | Discriminação aceitável   |
| De 30% a 40%  | Boa discriminação         |
| De 40% a 50%  | Excelente discriminação   |
| Acima de 50%  | Discriminação excepcional |

Tabela 5 – Níveis de Discriminação (Índice KS) (Fonte: Andrade, Oliveira, 2002)

Ao verificar modelos com KS entre 30 e 40, pode-se considerar modelos satisfatórios, com boa discriminação, mesmo que o público estudado seja instável, conforme explicado na seção referente ao público alvo (Cap. 2). Caso a seleção obtida fosse advinda de um público de renda superior, possivelmente encontrar-se-ia melhores índices de KS.

Outro fator que contribui para a dimensão deste índice, refere-se à metodologia adotada. Os modelos encontrados poderiam ter apresentados valores de KS ainda melhores, pois ao realizar o processo de seleção de modelos por *Backward*, prioriza-se a minimização do AIC, o que não necessariamente equivale à maximização do KS. O principal indicador de avaliação de modelos no mercado financeiro refere-se ao KS, portanto, do ponto de vista técnico, seria mais interessante criar um processo de seleção que buscasse a maximização do KS. No entanto, pela prática, percebe-se que quanto maior o número de variáveis significativas, maior será o KS do modelo. Sendo assim, seria necessário criar um indicador que penalizasse a quantidade de variáveis de um modelo no cálculo de seu KS para assim encontrar modelos mais assertivos e parcimoniosos.

Ao comparar o índice de KS dos dois modelos, verifica-se que o Modelo de Rentabilidade Total apresenta melhores resultados do que o modelo de Rentabilidade Boa. Em um primeiro momento, esperava-se que o Modelo de Rentabilidade Boa apresentasse maior KS, visto que, ao somente considerar as receitas advindas de juros de financiamento, poderia discriminar melhor os clientes. No entanto, o Modelo de Rentabilidade Total, que além de considerar as receitas advindas dos juros de financiamento, também considera os juros de atrasos e multas,

apresenta um melhor desempenho. O mesmo comportamento e ordem de grandeza do índice KS encontrados nos modelos de desenvolvimento, encontram-se no período de validação.

### 5.3 EFEITOS DAS VARIÁVEIS

Nesta seção, aprofunda-se o estudo do modelo para Rentabilidade Total, uma vez que apresentou melhor índice KS. Para verificar os efeitos de cada variável no modelo escolhido, foram simulados cenários distintos, alterando apenas uma variável e mantendo as demais *Ceteris Paribus*. Assim, verificou-se as funções de sobrevivência para cada alteração das variáveis e o tempo médio de permanência em cada estado. As funções de sobrevivência avaliam a probabilidade de um indivíduo não ir a perdas financeiras relevantes (neste caso, possuir rentabilidade inferior ao limite mínimo do produto) em relação ao tempo decorrido.

#### 5.3.1 Quantidade de Produtos

Para analisar os efeitos da variável “Quantidade de Produtos”, foram simulados cinco cenários para diferentes valores da variável estudada, de acordo com sua categorização.

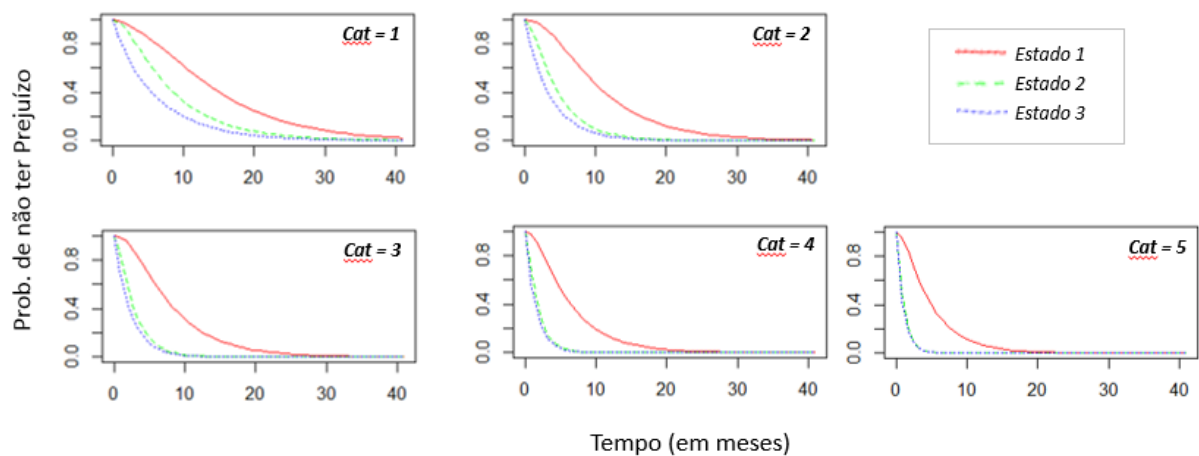


Figura 5 – Função de Sobrevivência – Quantidade de Produtos

Verifica-se que à medida em que o cliente aumenta a quantidade de produtos negociados, maior é a probabilidade de entrar em prejuízo. Do ponto de vista de mercado, torna-se muito claro que um cliente ao solicitar uma maior quantidade de produtos de crédito, possivelmente contraiu novas dívidas e está em busca de novos créditos para liquidá-las.

A Tabela 6 traz uma estimativa do tempo médio, em meses, de permanência em cada um dos estados, segundo os níveis da variável “Quantidade de Produtos”. Interessante notar que os indivíduos permanecem a maior parte do tempo no estado 1 (Rentabilidade Normal), o que pode ser considerável saudável para uma carteira de crédito. No entanto, são poucos os indivíduos que permanecem nas faixas de rentabilidade acima do esperado, e esta proporção diminui à medida que o cliente solicita uma maior quantidade de produtos. Percebe-se claramente um *tradeoff* entre concessão de novos créditos e rentabilidade do cliente, o que serve de inspiração para estudos posteriores.

| Quantidade de Produtos |       |       |       |       |       |
|------------------------|-------|-------|-------|-------|-------|
| Estado                 | Cat=1 | Cat=2 | Cat=3 | Cat=4 | Cat=5 |
| 1                      | 7,0   | 6,5   | 5,7   | 4,8   | 4,0   |
| 2                      | 2,5   | 1,2   | 0,6   | 0,3   | 0,1   |
| 3                      | 4,4   | 3,5   | 2,3   | 1,5   | 1,0   |

Tabela 6 – Tempo Médio de Permanência – Quantidade de Produtos

### 5.3.2 Saldo Médio em Conta

A Figura 6 ilustra alguns resultados interessantes ao se analisar os diferentes cenários para a variável “Saldo Médio em Conta”. À medida em que se eleva o saldo médio em conta corrente, maior é a probabilidade do indivíduo ir a perdas financeiras relevantes quando seu estado analisado já está em não-rentável (Estado 3). Ou seja, quando um cliente não é rentável, mesmo que ele possua saldo em conta elevado, tende a não ser rentável para a organização nos meses futuros.

Para os demais estados (Normal e Rentável), à medida que o cliente possui maior saldo médio em conta corrente, menor será a probabilidade de ir à perdas financeiras relevantes, o que condiz com a lógica do mercado, já que estes indivíduos têm condições de arcar com eventuais atrasos nos seus contratos de crédito.

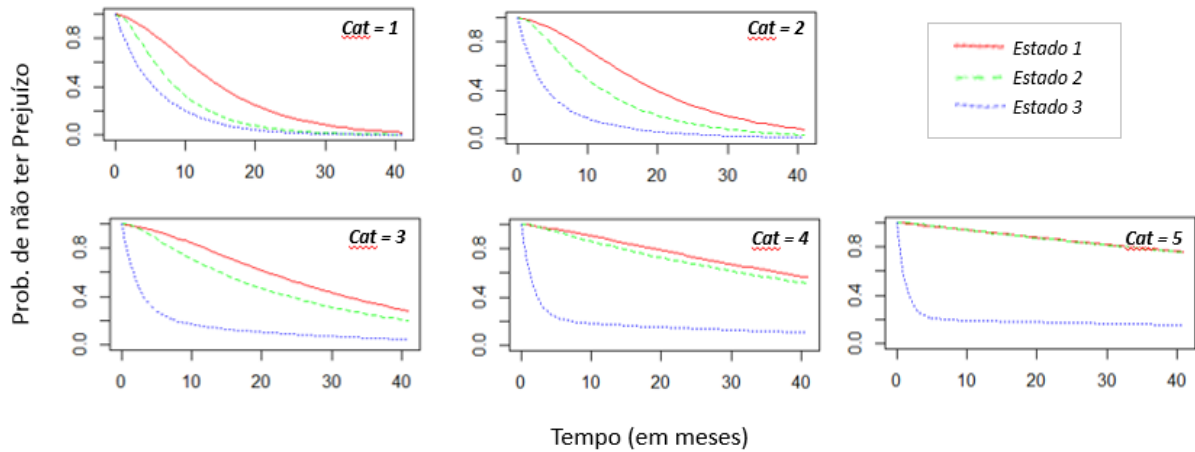


Figura 6 – Função de Sobrevivência – Saldo Médio em Conta

Chega-se a mesma conclusão quando avaliado o tempo médio de permanência em cada estado (Tabela 7). Para o cliente Não-Rentável, à medida que ele possui maior saldo médio em conta corrente, menos tempo ele permanece nesse estado, pois migra para perdas financeiras relevantes. Para os demais estados, à medida que aumenta o saldo médio, o cliente permanece no mesmo estado.

| Saldo Médio em Conta |       |       |       |       |       |
|----------------------|-------|-------|-------|-------|-------|
| Estado               | Cat=1 | Cat=2 | Cat=3 | Cat=4 | Cat=5 |
| 1                    | 7,0   | 8,2   | 9,4   | 10,9  | 12,6  |
| 2                    | 2,5   | 6,0   | 13,2  | 23,7  | 31,9  |
| 3                    | 4,4   | 3,2   | 2,3   | 1,7   | 1,2   |

Tabela 7 – Tempo Médio de Permanência – Saldo Médio em Conta

### 5.3.3 Meses até Saldo Máximo

A variável “Meses até Saldo Máximo” reflete o tempo entre o mês atual e o mês em que o máximo saldo devedor foi atingido nos últimos 12 meses. Percebe-se que quanto maior o tempo, menor é a probabilidade de ir a perdas financeiras relevantes (Figura 7). Mais uma vez a variável condiz com a realidade do mercado, pois quando o saldo devedor máximo está próximo do mês de referência, significa que mais valor em provisão deve ser alocado, o que impacta diretamente na rentabilidade do cliente.

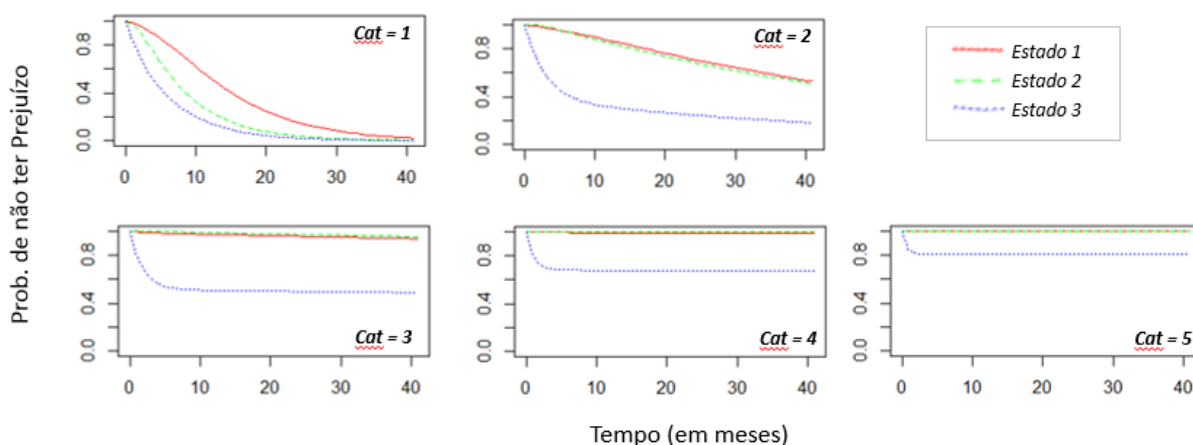


Figura 7 – Função de Sobrevivência – Meses até Saldo Máximo

Ao analisar o tempo médio de permanência em cada estado (Tabela 8), percebe-se algumas diferenças de comportamento. Tanto os clientes em estado normal e os não rentáveis, à medida que se afasta de seu saldo devedor máximo, menor é a probabilidade dele permanecer em seu estado. Para estes casos, grande parte desses clientes não migram para perdas financeiras relevantes, mas para faixas de rentabilidade mais elevadas, visto que se aloca menos provisões.

| Meses até Saldo Máximo |       |       |       |       |       |
|------------------------|-------|-------|-------|-------|-------|
| Estado                 | Cat=1 | Cat=2 | Cat=3 | Cat=4 | Cat=5 |
| 1                      | 7,0   | 5,1   | 3,4   | 2,3   | 1,5   |
| 2                      | 2,5   | 24,5  | 66,5  | 72,9  | 70,4  |
| 3                      | 4,4   | 3,0   | 1,8   | 1,0   | 0,5   |

Tabela 8 – Tempo Médio de Permanência – Meses até Saldo Máximo

### 5.3.4 Comprometimento de Renda

Uma interessante análise pode ser extraída da variável “Comprometimento de Renda”. Para os estados Normal e Rentável, à medida que um cliente compromete mais sua renda, menor é a probabilidade de se tornar uma perda financeira relevante (Figura 8). Em um primeiro momento pode parecer estranho, no entanto neste cálculo de comprometimento considera-se quanto o cliente utiliza de seu cartão de crédito e demais limites de crédito. Para os clientes Não-Rentáveis, quanto mais comprometida está sua renda, maior a probabilidade de entrar em perdas financeiras relevantes, o que condiz com a realidade do mercado. Destas análises podem partir novos estudos futuros, principalmente nas questões que se referem a atribuição de limites.

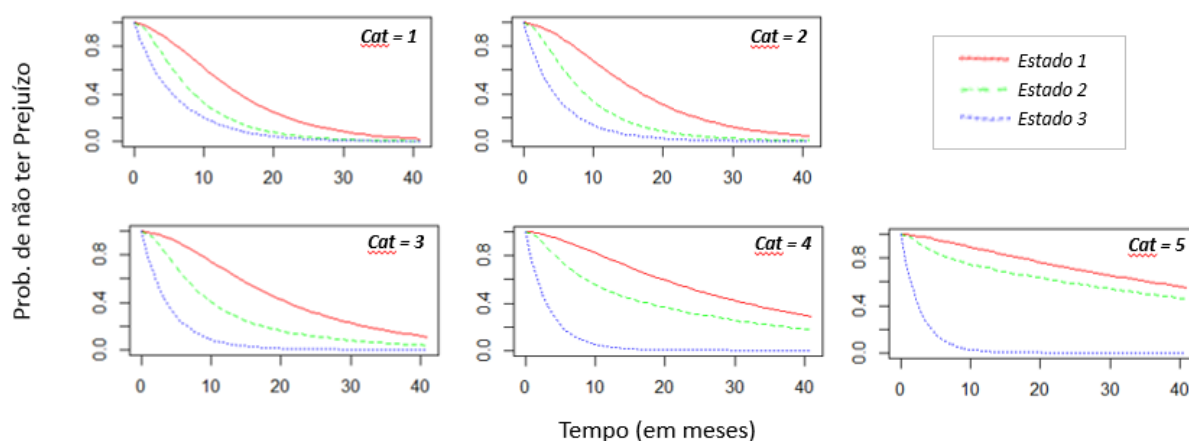


Figura 8 – Função de Sobrevivência – Comprometimento de Renda

Verifica-se, em indivíduos no estado Normal, que à medida em que se compromete mais a renda, aumenta sua permanência nesse estado (Tabela 9). Já para os estados Rentáveis e Não-Rentáveis, altera-se pouco de seu tempo médio de permanência.

| Comprometimento de Renda |       |       |       |       |       |
|--------------------------|-------|-------|-------|-------|-------|
| Estado                   | Cat=1 | Cat=2 | Cat=3 | Cat=4 | Cat=5 |
| 1                        | 7,0   | 8,7   | 10,8  | 13,3  | 16,5  |
| 2                        | 2,5   | 3,1   | 3,3   | 2,7   | 1,6   |
| 3                        | 4,4   | 4,1   | 3,7   | 3,2   | 2,7   |

Tabela 9 – Tempo Médio de Permanência – Comprometimento de Renda

## 5.4 COMENTÁRIOS GERAIS

Os modelos encontrados apresentaram performance satisfatória, no entanto esses modelos poderiam possuir melhores indicadores caso fosse criada uma metodologia específica que privilegiasse a maximização do KS, em detrimento à minimização do AIC.

Comparados com modelos de regressão logística do mercado, os índices de KS muito se assemelham. No entanto, utilizar Análise de Sobrevivência traz algumas vantagens, como por exemplo a estimação do tempo médio que um cliente demora para mudar de estado ou até mesmo o tempo médio que um cliente leva para entrar em perdas financeiras relevantes. Nos modelos de regressão logística apenas calcula-se a probabilidade de um indivíduo entrar em inadimplência em um tempo específico. Para os modelos de Análise de Sobrevivência, há flexibilidade de calcular a probabilidade de inadimplência para qualquer período.

Analisando as variáveis do modelo, verifica-se uma série de oportunidades para estudos futuros. Ao estudar a variável que corresponde à quantidade de produtos que o cliente negocia, fica claro a relação de *tradeoff* entre concessão de novos créditos e rentabilidade do cliente, pois à medida que o cliente negocia novos contratos, maior sua probabilidade de entrar em perdas financeiras relevantes, mesmo que esse cliente seja rentável no momento da análise. Por essa razão deve se avaliar com muita cautela a concessão de novos créditos para aqueles clientes que já possuem outros contratos.

Seguindo o mesmo raciocínio, à medida que um cliente negocia mais contratos, sua renda torna-se mais comprometida. Ao estudar a variável de comprometimento de renda, percebem-se comportamentos distintos dependendo do estado do cliente. Para clientes Não-Rentáveis, ao comprometer mais sua renda, maior a probabilidade de entrar em perdas financeiras relevantes. Já para os clientes em estado Normal ou Rentável, quanto mais sua renda está comprometida, menor será a probabilidade de entrar em perdas financeiras relevantes, o que contraria a análise realizada da variável “Quantidade de Produtos”. Na realidade, este aspecto não representa uma contradição de análises, e sim uma oportunidade de negócio, porque estes clientes podem ter seus limites revistos criando-se uma nova metodologia para atribuição de limites de crédito.

## CONCLUSÃO

Devido às projeções de expansão de crédito e a uma maior competitividade entre as instituições financeiras, torna-se necessário criar alternativas às análises de crédito de modo a diferenciá-las e melhorá-las comparando com o mercado. Sendo assim, novas técnicas estatísticas devem ser estudadas, de modo a observar sua aplicabilidade frente às demais técnicas presentes no mercado.

Dessa forma, analisou-se a modelagem realizada por Análise de Sobrevivência, aplicando os conceitos de Modelos Multi-Estados de Markov para prever se um cliente entrará em prejuízo financeiro. Dois modelos distintos foram desenvolvidos, o primeiro utilizando todas as receitas provenientes de um contrato financeiro, como juros advindos do financiamento, juros de atraso e multas. Já no segundo modelo, caracteriza-se como receitas apenas os juros provenientes do financiamento.

Ambos os modelos apresentaram desempenho satisfatório. Adotar escores de bureau como variáveis dos modelos, também poderiam trazer ganhos. Portanto, avalia-se positivamente todo o projeto estudado e abre-se perspectivas de novos estudos. Um grande desafio é criar uma metodologia de seleção de modelos que priorize a maximização do KS, penalizando o modelo devido sua quantidade de variáveis.

As perspectivas quanto aos estudos futuros são otimistas, já que atrelar as ações comerciais à rentabilidade podem trazer resultados mais satisfatórios, como a possibilidade de avaliar uma melhor relação entre o *tradeoff* de novas concessões de crédito x rentabilidade do cliente, e principalmente em questões estratégicas de uma instituição financeira, como uma melhor atribuição de limites.

As oportunidades de negócio advindas destes modelos podem trazer para a organização resultados valiosos, basta dar continuidade aos estudos da aplicabilidade da metodologia de Análise de Sobrevivência em estudos mais específicos.



**REFERÊNCIAS**

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control** v. 19, Issue: 6, p. 716-723, 1974.
- ANDRADE, André Pino da Silva. **Avaliação da recente expansão do crédito no Brasil: boom ou crescimento sustentável?** São Paulo, 2012. 101 p. Dissertação de Mestrado – Universidade de São Paulo, São Paulo, 2012.
- ANDRADE, F. W. M.; OLIVEIRA, J. G. C. Comparação entre Medidas de Performance de Modelos de Credit Scoring. **Revista Tecnologia de Crédito**. São Paulo. Ed. 33, 2006. Disponível em: [http://www.seradaexperian.com.br/serasaxperian/publicacoes/revista/2002/33/revista\\_0180](http://www.seradaexperian.com.br/serasaxperian/publicacoes/revista/2002/33/revista_0180). Acesso em: 06/11/2013
- ANDREEVA, G. European Scoring Models Using Survival Analysis. **The Journal of the Operational Research Society**. v. 57, n. 10, p. 1180-1187, 2006.
- ARASHIRO, Iracema Hiroko Iramina. **Modelo multi-estados markoviano não homogêneo com efeitos dinâmicos**. São Paulo, 2008. 96 p. Tese de Doutorado (Ciências. Área de concentração: Estatística) – Universidade de São Paulo, São Paulo, 2008.
- BANASIK, J.; CROOK, J.; THOMAS, L. Scoring by Usage. **The Journal of the Operational Research Society**. v. 52, n. 9, p. 997-1006, 2001.
- BAO, Yiqi. **Estimação e diagnóstico na distribuição Weibull – Binomial-Negativa em análise de sobrevivência**. São Carlos, 2012. 95 p. Dissertação em Mestrado (Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2012.
- BCB. Resolução 2.682. **Banco Central do Brasil**, p. 01-06, 1999.
- BELLOTTI, T.; CROOK, J. Credit Scoring with Macroeconomic Variables Using Survival Analysis. **The Journal of the Operational Research Society**. v. 60, n. 12, p. 1669-1707, 2009.
- CAO, Ricardo; VILAR, Juan M.; DEVIA, Andrés. **Modelling Consumer Credit Risk via Survival Analysis**. Coruña, 2008. 46 p. Universidade da Coruña, Coruña, 2008.
- COLOSIMO, Enrico Antônio; GIOLO, Suely Ruiz. **Análise de Sobrevivência Aplicada**. São Paulo: Edgard Blücher, 2006. 370 p.
- CORREIA, Alberto Filipe Neves. **Rendibilidade de Transacções de Crédito Pessoal com Recurso a Análise de Sobrevivência**. Porto, 2007. 62 p. Dissertação de Mestrado em Engenharia Matemática – Universidade do Porto, Porto, 2007.
- COX, R.. Regression Models and Life Tables. **Journal of the Royal Statistical Society**. Series B32 (2), p. 187-220, 1972.

COX, R.. Partial Likelihood. **Biometrika**. 62(2), p. 269-276, 1975

DEY, Shubhamoy. Modeling the Combined Effects of Credit Limit Management and Pricing Actions on Profitability of Credit Card Operation. **International Journal of Business and Management**. v. 5, n. 4, p. 168-177, 2010.

EXPERIAN, Serasa. Cadastro Positivo Serasa Experian. Disponível em: <<http://www.experian.com/cadastropositivo.com>>. Acesso em: 9 set 2013.

FDIC. Scoring and Modeling. **FDIC – Division of Supervision and Consumer Protection**, p. 51-65, 2007.

FEBRABAN. Bancarização e Inclusão Financeira no Brasil. **FELABAN - FEBRABAN**, No. 07/2011, 2011.

FERNANDES, Alda Marisa Ribeiro. **Análise de Dados em Modelos Multiestado**. Braga, 2010. 111 p. Dissertação de Mestrado em Estatística – Universidade do Minho, Braga, 2010.

FINLAY, S. M. Towards Profitability: A Utility Approach to the Credit Scoring Problem. **The Journal of the Operational Research Society**. v. 59, n. 7, p. 921-931, 2008.

FMI. Brazil: Financial System Stability. **International Monetary Fund Country Report**, No. 12/2006, p. 01-64, 2012.

JACKSON, C. et al. Multistate Markov models for disease progression with classification error. **The Statistician**. v. 52, Part 2, p. 193-209, 2003.

JACKSON, C. Multi-state Models for Panel Data: The msm Package for R. **Journal of Statistical Software**. v. 38, n. 8, p. 1-28, 2011.

LEE, Elisa T.; WANG, John Wenyu. **Statistical Methods for Survival Data Analysis**. New Jersey: Wiley & Sons, Inc. Hoboken, 2003. 513 p.

LOURENÇON, Adriana de Fátima. **Modelos de Análise de Sobrevida aplicados ao estudo de comportamento de retorno do doador de sangue**. Ribeirão Preto, 2007. 154 p. Dissertação de Mestrado (Faculdade de Medicina de Ribeirão Preto. Área de Concentração: Saúde da Comunidade) – Universidade de São Paulo, São Paulo, 2007.

MARSHALL, G; JONES, R. H. Multi-state Markov Models and Diabetic Retinopathy. **Statistics in Medicine**. v. 14, p. 1975-1983, 1995.

MASON, R. L., GUNST, R. F., HESS, J. L. **Statistical Design and Analysis of Experiments: With Applications to Engineering and Science**. New York: J. Wiley, 2003.

MEDINA, Rosa Puertas; SELVA, Maria Luisa Marti. Análisis del Credit Scoring. **Revista de Administração de Empresas**, São Paulo, v. 53, n. 3, p. 303-315, 2013.

MILLER, R. G. **Survival Analysis**. Wiley & Sons, 1981.

NARAIN, B. Survival Analysis and the Credit Granting Decision. **Credit Scoring and Credit Control**. p. 109-122, 1992.

POVOROZNEK, Jeferson Luis; GARCIA, Julio Alexandre Mello. **Modelos Markovianos Multi-estados**. Curitiba, 2008. 31 p. Trabalho de Conclusão de Curso (Curso de Graduação de Bacharelado em Estatística) – Universidade Federal do Paraná, Curitiba, 2008.

REGIS, Daniel Evangelista. **Aplicação do Modelo Multi-estado de Markov em Cartões de Crédito**. São Paulo, 2007. 58 p. Dissertação de Mestrado (Programa de Mestrado Profissional em Economia. Área de concentração: Finanças) – Faculdade Ibmec, São Paulo, 2007.

REGIS, Daniel Evangelista. **Aplicação do Modelo Multi-estado de Markov em Cartões de Crédito**. São Paulo, 2007. 58 p. Dissertação de Mestrado (Programa de Mestrado Profissional em Economia. Área de concentração: Finanças) – Faculdade Ibmec, São Paulo, 2007.

SCHWARTZ, R. A. An Economic Model of Trade Credit. **The Journal of Financial and Quantitative Analysis**. v. 9, n. 4, p. 643-657, 1974.

SILVA, Dany Rogers; RIBEIRO, Karem Cristina de Sousa; SHENG, Hsia Hua. Mensuração da rentabilidade do crédito comercial: aplicação em um caso atacadista-distribuidor. **Brazilian Business Review**. v. 8, n. 2, p. 22-42, 2011.

STEPANOVA, Maria; THOMAS, Lyn. Survival Analysis Methods for Personal Loan Data. **Operations Research**. v. 50, n. 2, p. 277-289, 2002.

STRAPASSON, Elizabeth. **Comparação de Modelos com censura intervalar em análise de sobrevivência**. Piracicaba, 2007. 135 p. Tese de Doutorado – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2007.

THOMAS, L. C.; EDELMAN, D.B.; CROOK, J.N. **Credit Scoring and its Applications**. Philadelphia: Monographs on Mathematical Modeling and Computation - SIAM, 2002. 250 p.

THOMAS, Lyn C. **Consumer Credit Models: Pricing, Profit and Portfolios**. New York: Oxford University Press Inc, 2009. 385 p.

## APÊNDICES

### 1 LISTA DE VARIÁVEIS ESTUDADAS

| Variável          | Descrição  |
|-------------------|--|
| C_MESES_SALDO_MAX | Qtde. de meses desde o saldo devedor máximo do cliente     |
| COMPR_RENDA       | Comprometimento de renda nos últimos 12 meses              |
| M_COMPORT         | Meses na base de comportamento                             |
| PC_UTLZ_CARTAO    | % de Utilização de cartão de crédito nos últimos 12 meses  |
| PC_UTLZ_LIM       | % de Utilização de limites de crédito nos últimos 12 meses |
| QTD_DEFAULT_12M   | Qtde. de contratos em default nos últimos 12 meses         |
| QTD_PRODUTOS      | Qtde. de produtos negociados pelo cliente                  |
| RECEBE_CONTA      | Identifica se recebe em conta                              |
| TEM_CARTAO        | Identifica se cliente possui cartão                        |
| TEM_LIMITE        | Identifica se cliente tem limites                          |
| TEMPO_CLI         | Tempo de relacionamento com o cliente                      |
| VR_SALDO_M_CONTA  | Saldo médio em conta corrente nos últimos 12 meses         |
| VR_UTLZ_CARTAO    | Valor utilizado de cartão de crédito nos últimos 12 meses  |
| VR_UTLZ_LIM       | Valor utilizado de limites de crédito nos últimos 12 meses |

## 2 EXEMPLO BASE DE DADOS

| DT_BASE | STATUS_TOTAL | CONT | CAMPO | C_MESES_SALDO_MAX | COMPR_RENDA | M_COMPOR | PC_UTILZ_CARTAO | PC_UTILZ_LIM | QTD_DEFAULT_12M | QTD_PRODUTOS | RECEBE_CONTA | TEM_CARTAO | TEM_LIMITE | TEMPO_CLI | VR_SALDO_M_CONTA | VR_UTILZ_CARTAO | VR_UTILZ_LIM |
|---------|--------------|------|-------|-------------------|-------------|----------|-----------------|--------------|-----------------|--------------|--------------|------------|------------|-----------|------------------|-----------------|--------------|
| 1       | 2            | 1    | 111   | 1                 | 3           | 3        | 2               | 1            | 3               | 1            | 2            | 2          | 1          | 4         | 2                | 3               | 4            |
| 2       | 1            | 2    | 111   | 3                 | 3           | 2        | 4               | 1            | 4               | 3            | 2            | 1          | 1          | 3         | 1                | 1               | 4            |
| 3       | 3            | 3    | 111   | 2                 | 2           | 4        | 4               | 3            | 3               | 4            | 1            | 1          | 1          | 2         | 3                | 1               | 3            |
| 4       | 4            | 4    | 111   | 1                 | 1           | 1        | 4               | 2            | 2               | 1            | 1            | 1          | 2          | 1         | 3                | 2               | 4            |
| 4       | 2            | 1    | 222   | 4                 | 4           | 3        | 1               | 3            | 4               | 3            | 2            | 1          | 2          | 1         | 1                | 4               | 1            |
| 5       | 3            | 2    | 222   | 1                 | 1           | 2        | 1               | 4            | 2               | 2            | 2            | 1          | 2          | 4         | 2                | 3               | 2            |
| 6       | 4            | 3    | 222   | 3                 | 2           | 2        | 2               | 3            | 4               | 2            | 1            | 2          | 2          | 4         | 4                | 3               | 1            |
| 3       | 1            | 1    | 333   | 3                 | 3           | 2        | 1               | 1            | 3               | 1            | 2            | 1          | 2          | 4         | 2                | 2               | 3            |
| 4       | 2            | 2    | 333   | 4                 | 4           | 1        | 1               | 2            | 4               | 3            | 1            | 2          | 1          | 4         | 2                | 4               | 1            |
| 5       | 3            | 3    | 333   | 1                 | 2           | 3        | 3               | 1            | 4               | 3            | 1            | 2          | 1          | 4         | 3                | 4               | 4            |
| 6       | 3            | 4    | 333   | 2                 | 3           | 2        | 4               | 4            | 2               | 3            | 1            | 1          | 2          | 4         | 1                | 3               | 3            |
| 7       | 3            | 5    | 333   | 1                 | 3           | 2        | 3               | 3            | 4               | 4            | 1            | 2          | 1          | 2         | 4                | 2               | 3            |
| 8       | 4            | 6    | 333   | 3                 | 4           | 4        | 2               | 1            | 2               | 3            | 1            | 1          | 1          | 2         | 2                | 2               | 1            |
| 1       | 1            | 1    | 444   | 1                 | 4           | 2        | 4               | 1            | 2               | 3            | 1            | 2          | 1          | 3         | 1                | 4               | 3            |
| 2       | 1            | 2    | 444   | 2                 | 4           | 4        | 3               | 1            | 4               | 1            | 2            | 2          | 1          | 4         | 2                | 4               | 4            |
| 3       | 1            | 3    | 444   | 4                 | 4           | 2        | 2               | 1            | 1               | 1            | 2            | 1          | 2          | 1         | 1                | 1               | 1            |
| 4       | 2            | 4    | 444   | 3                 | 4           | 2        | 4               | 3            | 2               | 2            | 2            | 2          | 2          | 4         | 2                | 3               | 2            |
| 5       | 3            | 5    | 444   | 3                 | 2           | 4        | 1               | 2            | 1               | 3            | 1            | 1          | 2          | 2         | 4                | 3               | 4            |
| 6       | 2            | 6    | 444   | 2                 | 1           | 4        | 3               | 3            | 1               | 4            | 1            | 1          | 2          | 2         | 3                | 4               | 4            |
| 7       | 2            | 7    | 444   | 2                 | 4           | 3        | 3               | 4            | 4               | 2            | 1            | 2          | 1          | 1         | 1                | 4               | 2            |

### 3 MATRIZES DE MIGRAÇÃO

As Matrizes de Migração apresentadas abaixo referem-se a transições de um horizonte de tempo de 12 meses.

| <b>Matriz de Migração</b>         |          |          |          |          |
|-----------------------------------|----------|----------|----------|----------|
| <b>Modelo Rentabilidade Total</b> |          |          |          |          |
| <b>Estados</b>                    | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> |
| <b>1</b>                          | 33%      | 62%      | 1%       | 4%       |
| <b>2</b>                          | 23%      | 72%      | 1%       | 4%       |
| <b>3</b>                          | 2%       | 6%       | 0%       | 92%      |
| <b>4</b>                          | 0%       | 0%       | 0%       | 100%     |

| <b>Matriz de Migração</b>       |          |          |          |          |
|---------------------------------|----------|----------|----------|----------|
| <b>Modelo Rentabilidade Boa</b> |          |          |          |          |
| <b>Estados</b>                  | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> |
| <b>1</b>                        | 22%      | 63%      | 7%       | 8%       |
| <b>2</b>                        | 15%      | 68%      | 11%      | 7%       |
| <b>3</b>                        | 2%       | 15%      | 34%      | 49%      |
| <b>4</b>                        | 0%       | 0%       | 0%       | 100%     |

## 4 CÓDIGO EM R

```

## Base Final para Mestrado ##
## Inputs iniciais ##
setwd("C:/Mestrado/Base Dissertação/Versão Final")
library("msm", lib.loc="C:/Program Files/R/R-3.0.2/library")
basemestrado <- read.csv("BASE_TOTAL_DESENV.csv",header=T)
basemestrado2 <- read.csv("BASE_TOTAL_VALID.csv",header=T)

## Código para gerar MSM da base Mestrado ##
statetable.msm(STATUS_TOTAL, CAMPO, data=basemestrado)
qm <- rbind(c(0,0.25,0,0.25),
            c(0.166,0,0.166,0.166),
            c(0,0.25,0,0.25),
            c(0,0,0,0))
qm0 <- crudeinits.msm(STATUS_TOTAL~CONT,CAMPO,data=basemestrado,qmatrix=qm)

## Função msm #
lista.msm<-function(variaveis)
{
  n<- msm(STATUS_TOTAL~CONT,subject = CAMPO, data = basemestrado, qmatrix =
qm0, death = 4,
        covariates=variaveis)
}

## Chama Funções ##
mestrado01.msm<-lista.msm(~AGR_QTDE_PRODUTOS+
  AGR_VR_SLD_MEDIO_CCORR_POUP+
  AGR_QT_DT_BASE_SALDO_MAXIMO+
  AGR_COMPROMETIMENTO_12M_EXP);

plot.msm(mestrado01.msm, legend.pos = c(8,1), xlab = "Tempo(anos)", ylab = "Prob de
Sobre")
plot.msm(mestrado01.msm, legend.pos = c(7,1), xlab = "Tempo(meses)", ylab = "Prob de
Prejuízo")

## Modelos Escolhidos ##

pmatrix.msm(mestrado01.msm, t=12)
sojourn.msm(mestrado01.msm,covariates=list(1,1,1,1))
plot.msm(mestrado01.msm,legend.pos=c(8,2),covariates=list(1,1,1,1),xlab =
"Tempo(meses)", ylab = "Prob de Prejuízo")

```