



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Lucas Tonegi

MODELO COM FRAÇÃO DE INADIMPLENTES PARA DADOS DE LONGA DURAÇÃO: UMA APLICAÇÃO A DADOS FINANCEIROS

Projeto de Pesquisa apresentado à disciplina Laboratório A do Curso de Graduação em Estatística da Universidade Federal do Paraná, como requisito para elaboração do Trabalho de Conclusão de Curso.

Orientadora: Profa. Dra. Suely Ruiz Giolo

**CURITIBA
2017**

Sumário

1 INTRODUÇÃO	3
2 OBJETIVOS	7
2.1 Objetivo Geral	7
2.2 Objetivos Específicos	7
3 MATERIAL E MÉTODOS	8
3.1 Material	8
3.1.1 Conjunto de Dados	8
3.1.2 Recursos Computacionais	8
3.2 Métodos	9
4 CRONOGRAMA DE ATIVIDADES	11
REFERÊNCIAS	12

1 INTRODUÇÃO

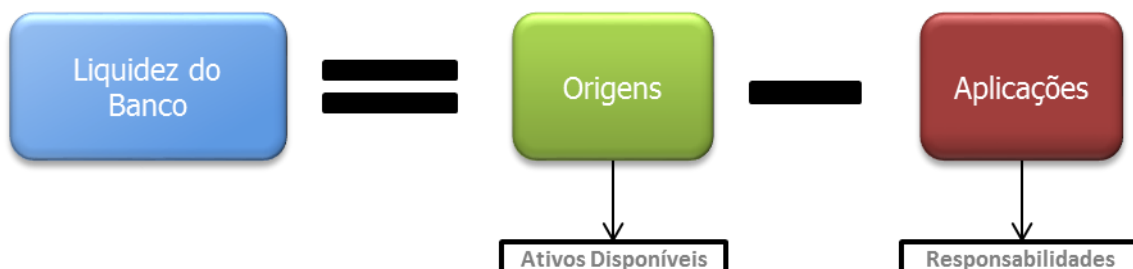
A crise econômica brasileira tem afetado cada vez mais a população no ano de 2017. São vários os fatores que geraram essa crise econômica, sendo a crise política um dos principais motivos para este momento conturbado. No primeiro trimestre de 2017, a taxa de desemprego no Brasil atingiu 13,2%, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE). E também uma alta de 1,3% na comparação com o trimestre anterior, sendo a maior taxa de desocupação no País da série histórica do indicador iniciada em 2012, segundo os dados da Pesquisa Nacional por Amostra de Domicílios (Pnad). Com problemas de estagnação da economia, alta na taxa de inflação, desvalorização da moeda, aumento na taxa de desemprego (citado anteriormente), dentre outros, é de se esperar cada vez mais o endividamento das pessoas.

O endividamento da população é algo que afeta diretamente as empresas que hoje atuam no mercado financeiro tais como: bancos, financeiras, seguradoras, operadores de crédito etc. Para entender melhor como o endividamento afeta estas empresas, é necessário entender o funcionamento de um banco e o ciclo de crédito que será explicado a seguir.

Bancos são instituições que trabalham com dinheiro, quer seja de terceiros ou seu próprio investimento, sendo que em grande parte com o dinheiro de terceiros (clientes). Devido a tal fato, os bancos oferecem vários tipos de serviços: cartão de crédito, conta corrente, conta poupança, crédito imobiliário, financiamento de automóveis, dentre outros. Os bancos têm grande importância na economia de um País, pois através dos créditos emprestados gera um aumento no capital circulante, e consequentemente o aumento da renda da população e o aumento de empregos na sociedade, além de facilitar as transações comerciais. Uma forma dos bancos terem lucro, se dá por meio dos empréstimos (créditos oferecidos), consequência da diferença entre as taxas de juros pagas e cobradas. Outra forma, é através dos ativos que se encontram disponíveis nas contas dos clientes, geralmente aplicados numa operação de curto prazo chamada de *Overnight*, que consiste em aplicar a liquidez de um banco no final de um dia (dado que a liquidez foi positiva) em títulos do governo federal (BACEN), resgatando-a na manhã do próximo dia útil. Com isso, os valores aplicados

voltam corrigidos de acordo com os juros equivalente a um dia da taxa básica Selic, como mostra a Figura 1.

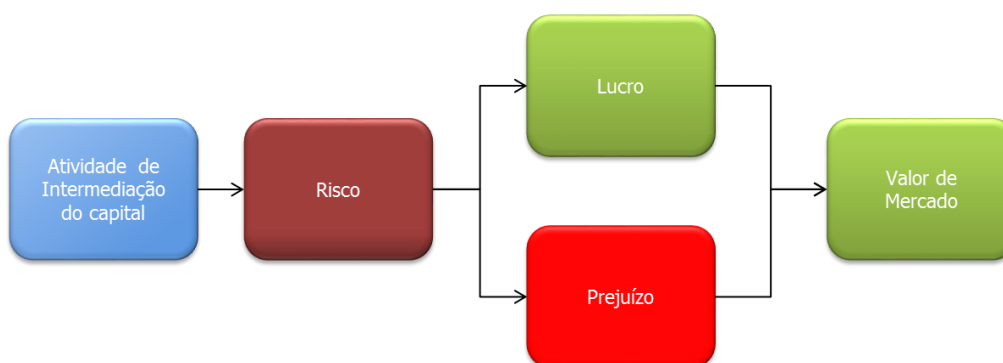
Figura 1 – Definição da liquidez de um banco.



Porém, os bancos não têm a liberdade de aplicar todo o dinheiro dos correntistas, pois são obrigados a depositar uma parte no banco central, a fim de garantir que o banco não fique sem reservas e cumpra com suas obrigações, além de garantir a Provisão para Devedores Duvidosos (PDD), que será discutida adiante.

Nota-se que toda operação tem um risco, definido como a possibilidade da perda resultante da incerteza quanto ao recebimento de valores pactuados com tomadores de empréstimos, contratantes de contratos ou emissão de títulos. E é exatamente por este risco assumido que todo empréstimo está sujeito a juros, um valor a ser pago para a instituição financeira em troca deste risco. Podendo assim, caso o cliente pague seu empréstimo em dia, gerar um lucro, e caso não pague com suas dívidas, gerar um prejuízo, conforme mostra a Figura 2.

Figura 2 – Fluxo da atividade de intermediação do capital



Inicialmente divide-se o ciclo de crédito em quatro partes: 1) Planejamento do produto, 2) Originação ou Concessão de crédito, 3) Manutenção dos produtos e 4) Cobrança (área que é afetada diretamente pela crise). O planejamento do produto é o início deste ciclo. Nesta etapa, é necessário definir como serão ofertados os produtos (já citado anteriormente). A segunda etapa deste ciclo, que consiste na concessão de crédito é a etapa em que entra o chamado *Credit Score*, que consiste em um score de obtido por meio de um modelo estatístico (usualmente o logístico) para cada cliente que chega em uma agência bancária para a solicitação de um produto (esse score seria a probabilidade de um cliente ser um bom pagador). A terceira etapa, consiste na manutenção destes produtos, em que é utilizado o chamado *Behavior Score*, um score que descreve o comportamento deste cliente dentro da instituição financeira. Caso este cliente solicite um novo produto, ele possivelmente será avaliado por este score de comportamento. E por fim, clientes que se tornam mau pagadores (inadimplentes) chegam na última etapa, a cobrança. Quando o cliente já está em cobrança, é gerado um novo score chamado *Collection Score* (também obtido, usualmente, a partir de um modelo de regressão logística) que consiste em uma estimativa da probabilidade do pagamento da dívida já em atraso.

A análise de sobrevivência tem, em geral, o interesse em estudar o tempo até a ocorrência de um determinado evento. Sendo assim, ela está sendo cada vez mais aplicada em diversas áreas de pesquisa, tal como na Medicina, em que o interesse pode ser estudar o tempo até a recidiva de uma doença após o seu tratamento, e na indústria, em que o interesse pode estar no estudo do tempo até um dispositivo eletrônico parar de funcionar. Na área financeira, há interesse dos bancos em estudar o tempo em que um cliente já devedor venha a pagar sua dívida. Esse estudo é extremamente relevante para se fazer uma previsão de devedores duvidosos (reserva que deve ser depositada para o Banco Central de todos os inadimplentes) e também, em específico, para a venda de carteiras. O processo de venda de carteiras consiste em analisar uma determinada população, já com tempo em atraso elevado (normalmente 60 meses), e definir características de possíveis pagamentos que possam ser realizados e não pagamentos. As dívidas dos maus pagadores podem ser vendidas a um preço bem inferior do que valem para uma empresa terceirizada ou outra instituição financeira. Por exemplo, uma carteira com um valor total de dez milhões de reais que podem ser recuperados, pode ter

seu valor estipulado em sete milhões de reais devido ao elevado número de dias em atraso que já se encontram.

Dado o momento de crise e instabilidade econômica atual, este projeto tem como foco o estudo do tempo até a ocorrência do pagamento de dívidas de clientes que já se encontram em atraso, buscando a identificação de possíveis fatores (covariáveis) que afetam este tempo de pagamento.

2 OBJETIVOS

2.1 Objetivo Geral

Investigar fatores que afetam o tempo até o pagamento de clientes já em inadimplência utilizando técnicas e ferramentas estatísticas para analisar perfis de possíveis maus pagadores e conseqüentemente efetuar a venda das dívidas dos mesmos.

2.2 Objetivos Específicos

1. Utilizar um método estatístico alternativo para analisar e identificar clientes de uma população que podem não estar sujeitos à ocorrência do evento e, por mais longo que seja o tempo de observação, o evento nunca ocorrerá para uma parte da população;
2. Realizar análise exploratória dos dados a fim de verificar a existência de indivíduos imunes ao evento;
3. Ajustar modelos estatísticos aos dados descritos na Seção 3.2 com a finalidade de investigar quais as covariáveis (fatores), dentre as disponíveis, influenciam o tempo até o pagamento total ou parcial de clientes inadimplentes;
4. Comparar e discutir os resultados dos modelos ajustados.

3 CLIENTES E MÉTODOS

3.1 CLIENTES

3.1.1 Conjunto de Dados

Será estudado um conjunto de dados reais fornecidos por uma instituição financeira brasileira referente a clientes que estão em inadimplência e foram observados por um período de vinte e quatro meses. Dado que o interesse neste estudo é o tempo até o pagamento, serão consideradas características referentes aos clientes e variáveis de apontamentos de negativação (*bureaus* de informação).

A variável resposta a ser considerada será o tempo (em meses) entre a entrada do cliente em atraso e o evento de interesse, que no caso desse estudo será o pagamento total ou parcial de suas dívidas. A fração de inadimplentes após o período de 24 meses de acompanhamento, constitui aproximadamente 60% do conjunto de dados a ser analisado.

3.1.2 Recursos Computacionais

O *software* R, versão 3.2.2 (R CORE TEAM, 2015) será utilizado para ajustar os modelos aos dados descritos. Alguns pacotes a serem investigados com esse propósito são: *timereg*, *survival*, *gamlss* e *smcure*. Também será utilizada uma macro do *software* SAS (Statistical Analysis System) Enterprise Guide 7.1 proposta por Corbière e Joly (2007).

3.2 Métodos

Para a análise do banco de dados descrito na Seção 3.1.1, pretende-se utilizar metodologias estatísticas no contexto de modelos de regressão logística e análise de sobrevivência.

O modelo de regressão logística é o mais usado quando a variável resposta que se deseja analisar é binária ou dicotômica. Embora existam outros modelos para analisar dados em que a resposta é binária, a regressão logística se tornou popular por ser flexível do ponto de vista matemático, de fácil utilização e por apresentar interpretação simples de seus parâmetros (GIOLO, 2012). Na área financeira, é também o modelo mais utilizado para avaliações de clientes. Contudo, outras funções de ligações estão sendo propostas para regressão binária, como é o caso das *power* e *reversal power link* que, com a adição de um parâmetro em forma de potência nas funções de ligação de diversas distribuições, fornece uma maior flexibilidade no ajuste do modelo, o que pode trazer ganhos consideráveis (LOUZADA NETO et al., 2016). Neste trabalho, a proposta é utilizar a regressão binária com função de ligação logito (modelo logístico) com o objetivo de comparar os resultados desse modelo com os do modelo de mistura com fração de inadimplentes. Essa comparação entre os modelos, visa a implantação de uma forma alternativa de modelar e analisar clientes em inadimplência.

A análise de sobrevivência é utilizada quando se deseja estimar a probabilidade de sobrevivência a um evento de interesse, denominado falha, associada a cada instante de tempo durante um período pré-estabelecido de observação. Por isso, no contexto descritivo de análise de sobrevivência, será utilizado o estimador não paramétrico de Kaplan-Meier (KAPLAN; MEIER, 1958) para a estimação da função de sobrevivência. Também pelo fato de termos na base de dados uma parte da população não susceptível ao evento de interesse (EUDES et al., 2012), será utilizado uma extensão do modelo de Cox (COX, 1972), denominado modelo de mistura de Cox com fração de cura (neste caso fração de inadimplentes), que é capaz de acomodar essa fração de indivíduos em que não ocorre o evento. Esse modelo, corresponde a uma extensão do modelo de Cox proposta por Kuk e Chen (1992). Também, no contexto de análise de sobrevivência, pretende-se considerar modelos de mistura no contexto paramétrico, assumindo para os tempos distribuições como a Weibull e a log-logística, que podem se ajustar bem a

dados financeiros de longa duração (GRAZOTTO et al., 2008). Ainda, no contexto paramétrico, poderão ser considerados os modelos com tempo de promoção, apresentados por Yakovlev et al., (1993) e, posteriormente, estudados por Yakovlev e Tsodikov (1996) e aprofundado por Chen et al. (1999). Para a formulação desses modelos é assumido que o evento de interesse pode ocorrer por um número não observável de causas. Sendo assim, a variável aleatória que representa o número de causas que podem levar à ocorrência de um evento, segue uma distribuição Poisson. Nesses modelos, a distribuição associada aos tempos T pode ser especificada por modelos paramétricos tais como: o Weibull, o log-logístico, o exponencial, o log-normal, a gamma etc. Castro et al. (2010) propuseram o modelo de tempo de promoção em que as causas, ao invés de seguir uma distribuição de Poisson, seguem uma distribuição binomial negativa. Esse modelo, poderá, possivelmente, ser considerado para a análise dos dados do estudo descrito.

Os modelos de sobrevivência semi-paramétricos e paramétricos a serem considerados nesse trabalho, levam em conta a fração de cura, ou fração de inadimplentes, no caso dos dados a serem analisados. Esta fração deve ser explicada de forma que não cause um confundimento com as censuras. Portanto, é importante entender o que é fração de cura e o que é censura. Por exemplo, em um estudo em que os indivíduos foram acompanhados por certo período de tempo (período suficientemente grande, por isso longa-duração) e o evento, após esse período, ainda não ocorreu para uma parte deles, então o evento provavelmente não mais acontecerá para a maioria deles. A essa fração (ou proporção) de indivíduos na qual o evento não ocorrerá, mesmo se observados por mais tempo, é denominada fração de cura. Por outro lado, a parcela de indivíduos, usualmente muito pequena, na qual o evento não foi observado após um período longo de acompanhamento, mas que possivelmente ocorreria, caso fossem acompanhados por mais tempo, caracterizam as censuras. Os modelos de mistura para dados de longa duração foram propostos para acomodar essas situações. Desse modo, esses modelos serão considerados nesse trabalho, tendo em vista a fração elevada de inadimplentes existente no banco de dados a ser analisado.

Portanto, pretende-se, nesse trabalho, ajustar os modelos mencionados e, após isso, será realizada a comparação desses modelos na tentativa de inovação e melhoria na identificação de possíveis clientes bons e possíveis clientes que nunca irão realizar o pagamento das suas dívidas em atraso.

REFERÊNCIAS

- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Editora Blucher, 2006. 392 p.
- DIXON, S. N.; DARLINGTON, G. A.; DESMOND, A. F. A competing risks model for correlated data based on the sub distribution hazard, **Lifetime Data Analysis**, Boston, v. 17, p. 473-495, 2011.
- KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. 2. ed. New York: Springer, 2003. 536 p.
- R CORE TEAM. **R: A language and environment for statistical computing**. Viena, Áustria, 2015. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.
- SAS/STAT© Software: **Enterprise Guide , 7.1 Copyright**, SAS Institute Inc. Cary, NC, USA, 2016.
- BERKSON, J.; GAGE, R Survival cure for cancer patients following treatment, **Journal of the American Statistical Association** v. 47, p. 501-515, (1952).
- BERKSON, J. Application to the logistic function to bio-assay. **Journal of the American Statistical Association**, v. 39, n. 227, p. 357-365, 1944.
- COX, D.R. Regression models and life-tables, **Journal of the Royal Statistical Society**. Series B. v. 34, p. 187-220. 1972. Disponível em: < <http://hydra.usc.edu/pm518b/literature/cox-72.pdf> >. Acesso em: 12 Jun. 2017.
- TOMAZELA, S.M.O. **Avaliação de desempenho de modelos de Credit Score ajustados por Análise de Sobrevivência**. Dissertação de Mestrado. São Paulo: Instituto de Matemática e Estatística, Universidade de São Paulo, 2007.
- ROCHA F. C. **A inadimplência de crédito no setor bancário brasileiro: um estudo de caso**. Monografia (Graduação em Economia). Florianópolis: Universidade Federal de Santa Catarina, 2010.
- ROSENBERG, E.; GLEIT, A. Quantitative Methods in Credit Management: A Survey. **Operations Research**, v. 42, n. 4, p. 589-613, 1994.
- QUIDIM, I. L. **Análise de sobrevivência com fração de fidelizados: uma aplicação na área de marketing**. Dissertação (Mestrado em Estatística) São Paulo: IME - Instituto de Matemática e Estatística, Universidade de São Paulo, 2005.
- GRANZOTTO, D.C.T; LOUZADA-NETO, F; PERDONÁ, G.S.C. Modelos de sobrevivência com longa duração: uma aplicação a grandes bancos de dados financeiros. **Revista Brasileira de Biometria**, v. 24, n. 4, p.102-116, 2010.
- KAPLAN, E.L; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, p. 457-81, 1958.