

Modelos de regressão logística para risco de inadimplência de doadores filantrópicos: variáveis determinantes e predição¹

Thalita Paula Monteiro²

Luciano Vereda³

Adrian Pizzinga⁴

Resumo: Este artigo oferece uma solução para o seguinte problema: entender e otimizar as taxas de inadimplência de pessoas físicas quanto às suas prometidas doações para instituições filantrópicas. Adota-se uma solução muito utilizada na literatura, qual seja, a de ajustar um modelo de regressão logística para desvendar as variáveis que influenciam a probabilidade de ocorrer inadimplência e que predigam novos casos de inadimplência. O estudo de caso envolve dados de pessoas físicas doadoras da Junta de Missões Nacionais da Convenção Batista Brasileira, oriundos das cinco regiões do Brasil. Os principais resultados obtidos: (i) as variáveis Idade, Estado Civil, Gênero, Região, Tipo de Doador, Tipo de Cobrança e Vencimento (para o pagamento bancário da doação) se mostraram como os determinantes da inadimplência; e (ii) a taxa de acerto de predição mostrou que o modelo final é capaz de excelente antecipação dos casos de inadimplência.

Palavras-chave: doadores; inadimplência; instituição filantrópica; modelo de regressão logística; predição.

Códigos JEL: C250; D640; G210; L310.

¹Os autores são muito gratos ao Professor Jaime Orrillo e aos Pareceristas que, com suas diversas correções e sugestões, fizeram do artigo originalmente submetido um trabalho muito mais bem escrito. Também merecem reconhecimento os Professores Marco Aurélio Sanfins e Valentin Sisko por comentários e sugestões para a primeira versão do artigo. Em particular, agradecemos o Professor Sisko pela excelente sugestão (assim consideramos na ocasião) de criarmos a que foi chamada de Figura 9, para melhor entendimento do leitor sobre duas importantes variáveis explicativas. Adrian Pizzinga agradece o apoio financeiro da CAPES (Bolsa PNPd), concedido para Pós-Doutoramento na Escola de Economia de São Paulo (FGV EESP).

² Junta de Missões Nacionais (thalitamonteiro@id.uff.br).

³ Faculdade de Economia – Universidade Federal Fluminense (UFF) (lvereda@gmail.com).

⁴ Instituto de Matemática e Estatística – Universidade Federal Fluminense (UFF) (adrianhpster@gmail.com) – Endereço: Rua Mário Santos Braga S/N Valonguinho, 24.020-140. Niterói, Rio de Janeiro Brasil).

Logistic regression models for default risk of philanthropist: determinant variables e prediction

Abstract: This article discusses a solution for a specific problem: the understanding and improvement of default rates among people who engage on donating commitments with philanthropic institutions. We rely on a typical tool of the financial and economic literature, i.e., the selection and estimation of an adequate logistic regression model for unveiling the variables that explain the probability of default and that make possible the prediction of upcoming default events. The empirical example involves a data set about donators of a philanthropic non-governmental institution, including people from the five regions of the country. The main findings: (i) the variables Age, Civil Status, Region, Donator Type, Charging Type and Donation Deadline have proved to be the most relevant to predict default events; and (ii) out-of-sample validation has confirmed that the estimated model is capable of excellent prediction of new default cases.

Key-words: default; donator; logistic regression model; philanthropic institution; prediction.

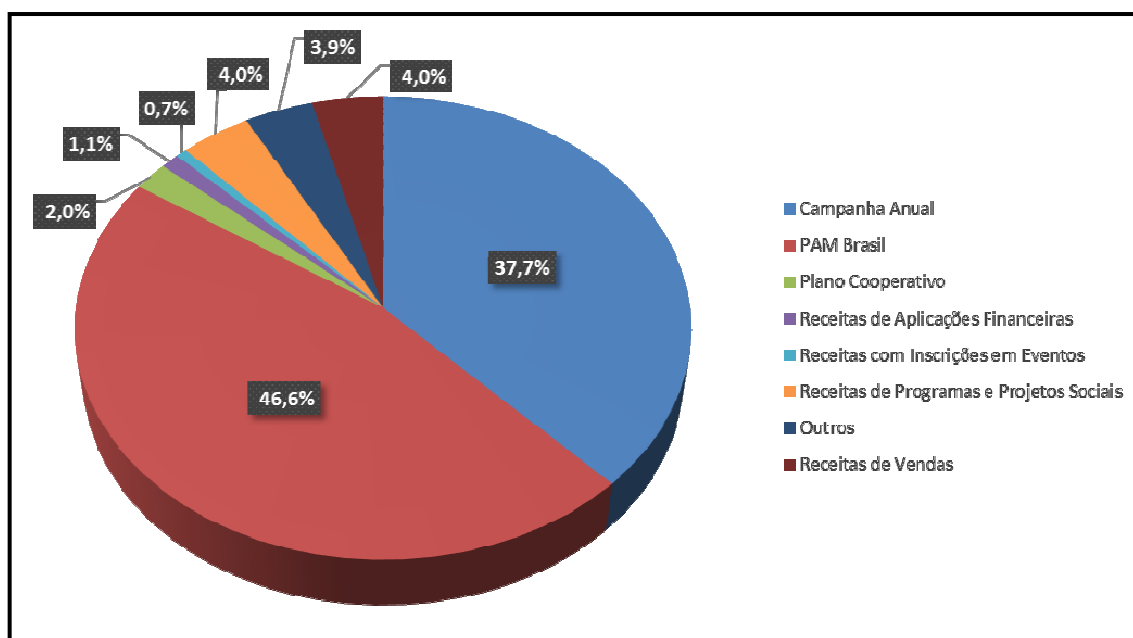
JEL codes: C250; D640; G210; L310.

1. Introdução

A Junta de Missões Nacionais (JMN) é uma agência missionária filantrópica que pertence às Igrejas Batistas da Convenção Batista Brasileira (CBB). De acordo com Oliveira (2007), a JMN iniciou suas atividades no dia 25 de junho de 1907 durante a primeira assembleia da CBB na cidade de Salvador na Bahia. Por existir uma grande preocupação de caráter social por parte da JMN, seu trabalho consiste em sustentar vários projetos sociais. Dentre estes, destaca-se a Cristolândia, um programa de recuperação de dependentes químicos. Somente neste projeto, existem 34 unidades distribuídas em 7 Estados do Brasil, acolhendo 1.202 contemplados – com testemunhos, encontrados em Santos (2015), de pessoas ex-usuárias de drogas e atendidas no projeto. Mais detalhes sobre o trabalho da JMN podem ser encontrados em www.missoesnacionais.org.br.

Ao ser analisada na Figura 1 a distribuição das fontes de receita no período de 2014 da JMN, fica evidenciada uma concentração de receitas em duas fontes, quais sejam, o Programa de Adoção Missionária (PAM) e a Campanha Anual.

Figura 1 – Distribuição de receitas globais da JMN.

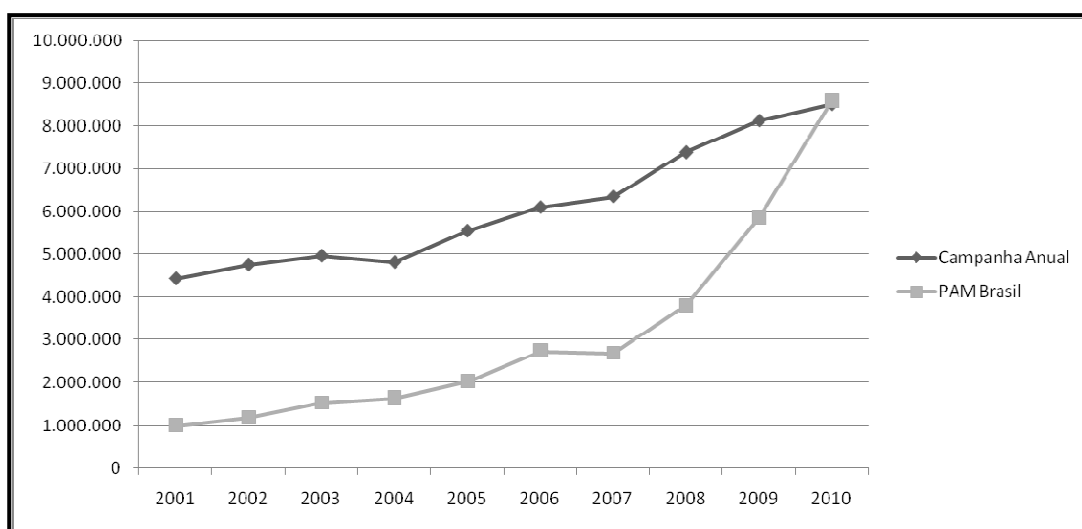


Fonte: Sistema Financeiro e Contábil da JMN 2015.

Dentre as duas fontes mencionadas, destaca-se, tanto pelo volume da receita quanto pelo crescimento histórico, a fonte de receita advinda do PAM. Ademais, se a análise considerar o

crescimento histórico, detalhado na Figura 2, das receitas das duas maiores fontes, fica ainda mais evidente que o PAM é uma estratégia de captação de receitas que se tornou fator de importância central para a sobrevivência da JMN.

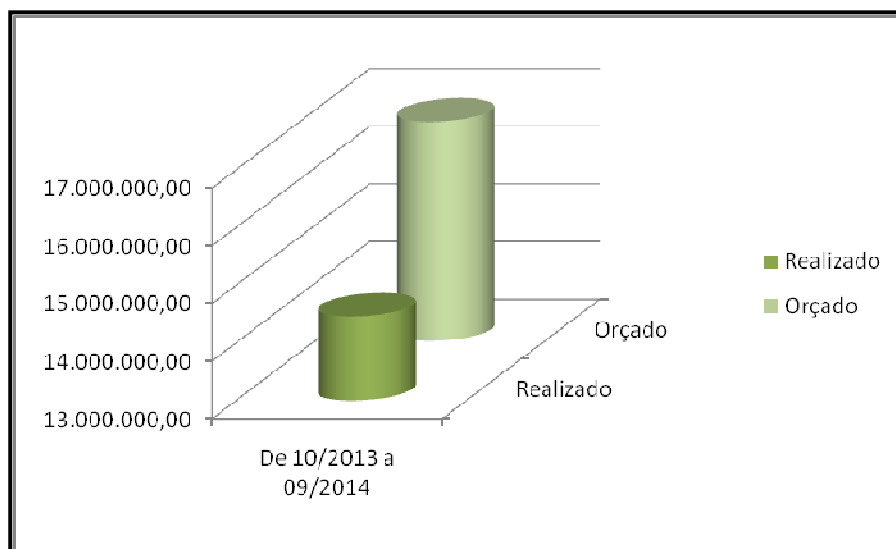
Figura 2 – Crescimento histórico das receitas (em reais) do PAM Brasil e da Campanha Anual.



Fonte: Sistema Financeiro e Contábil da JMN 2015.

No entanto, pela Figura 3, e considerando a importância do PAM, os resultados não se mostraram satisfatórios no período mais recente, pois as correspondentes doações (realizadas efetivamente) não alcançaram patamares próximos ao que fora orçado pela JMN. Este resultado se deu por conta do alto índice de inadimplência dos doadores em seus respectivos projetos, o que trouxe, por conseguinte, uma preocupação quanto à estabilidade financeira da JMN.

Figura 3 – Dados orçamentários (em reais) do PAM Brasil.



Fonte: Sistema Financeiro e Contábil da JMN 2015.

Neste artigo, modelos de regressão logística são estimados visando-se a dois objetivos: (i) identificar as variáveis explicativas estatisticamente significantes que determinem as probabilidades de ocorrerem inadimplência pelos doadores do PAM; e (ii) como tarefa principal, prever / antever casos de inadimplência a partir do conhecimento dos valores das variáveis explicativas previamente identificadas, correspondentes a novos doadores. Com estes dois objetivos atingidos, arrecadações e ações de mobilização de pessoas serão ampliadas. Com isso, a receita da JMN deverá aumentar e, por conseguinte, haverá mais eficiência na implementação dos projetos sociais. O foco recairá sobre dados com informações de *peças físicas*, pois elas representam quase 70 por cento das contribuições ao PAM (cf. Amaral, 2013). Além do mais, cerca de 7 em cada 10 pessoas físicas que assinam um PAM o deixam após três meses (cf. Souza, 2013).

O artigo assim se organiza. A seção 2 oferece uma contextualização econômica do problema, com uma analogia entre a inadimplência de doadores filantrópicos e os elementos essenciais do risco de crédito. A seção 3 é inteiramente reservada aos pormenores essenciais da teoria estatística dos modelos de regressão logística (aqui visto como um modelo de risco de crédito), com ênfase na definição geral do modelo, na estimação dos parâmetros e nas estratégias de predição. Ao longo de toda a seção 4, apresenta-se um estudo de caso com os dados sobre pessoas físicas doadoras da JMN – serão abordados: o banco de dados, estatísticas descritivas (gráficos de frequência), o ciclo percorrido de modelagem, as

interpretações dos resultados e a análise do poder preditivo com o modelo definitivo. A seção 5 destina-se à conclusão e às sugestões de estudos futuros dentro do tema.

2. Inadimplência de doadores filantrópicos: analogia com o risco de crédito

A definição mais reconhecida de risco de crédito se baseia na relação estabelecida entre emprestadores e tomadores de recursos, que é inerente à atividade de intermediação financeira. De acordo com o Banco Central do Brasil (BCB), órgão brasileiro responsável por ordenar e supervisionar esse tipo de relação, o risco de crédito é definido como “... a possibilidade de ocorrência de perdas associadas ao não cumprimento pelo tomador ou contraparte de suas respectivas obrigações financeiras nos termos pactuados, à desvalorização de contrato de crédito decorrente da deterioração na classificação de risco do tomador, à redução de ganhos ou remunerações, às vantagens concedidas na renegociação e aos custos de recuperação” (Resolução nº 3721, artigo 2º, BCB).

Dentre os vários elementos presentes nesta definição, está a existência de um contrato de crédito que disciplina o serviço de natureza financeira prestado pela parte credora, a qual transfere recursos financeiros para a parte devedora em troca da devolução futura do montante emprestado acrescido dos respectivos juros, que constituem o preço do serviço. O problema estudado neste trabalho, porém, é a quebra da promessa de doação de recursos por parte de um indivíduo, que não honra o compromisso assumido com uma entidade filantrópica. Não se trata, portanto, de um problema “clássico” de risco de crédito porque não há a prestação de um serviço financeiro, tampouco um contrato jurídico celebrado entre as partes e que regula a transação.

Apesar da apontada diferença, há uma analogia entre o caso tradicional de risco de crédito e o caso específico que estudamos. Com efeito, em ambos os casos existe a expectativa de transferência de recursos financeiros de uma parte para outra, que deve ser efetuada em prazo livremente pactuado. Adicionalmente, a falha em transferir esses recursos enseja o evento de *default*, que ocorre com uma determinada probabilidade. Esses dois elementos estão presentes tanto na noção “histórica” de crédito citada por Blatt (1999) (que enfatiza que a palavra crédito se origina da palavra latina *credere* – crer, confiar, acreditar – e do substantivo *creditum* – ou seja, confiança) quanto na definição de risco de crédito sugerida por Caouette et al. (1999), que se baseia na probabilidade de que a expectativa de transferência não se cumpra. É, portanto, bastante razoável tratar os dois casos de modo

semelhante, inclusive utilizando as ferramentas propostas na imensa literatura de risco de crédito para modelar a ocorrência de *default* – aqui entendido como *a falha do doador filantrópico em honrar as doações prometidas para a empresa filantrópica*.

Sob uma perspectiva ampla, os modelos de risco de crédito são ferramentas estatísticas usadas para mensurar o risco de tomadores e suas transações. Entre estes, citam-se os modelos de previsão de insolvência, que têm como objetivo calcular/estimar a probabilidade de uma empresa ou indivíduo cair em um evento de *default* em um determinado período de tempo. Esses modelos se baseiam em amostras históricas com dados sobre tomadores de crédito (pessoas jurídicas e físicas). Com a análise da amostra coletada, buscam-se: (i) a identificação das variáveis que melhor discriminam/classificam as empresas e pessoas em insolventes ou solventes no período da análise, e (ii) com base nas variáveis selecionadas, a classificação das empresas e pessoas, nas suas futuras operações de créditos, como prováveis solventes ou insolventes. Estes modelos de previsão de insolvência se baseiam em técnicas estatísticas de Análise Multivariada, tais quais a análise discriminante e classificatória (cf. Johnson e Wichern, 2007), e de Análise de Regressão, cujos possíveis exemplos são os modelos de regressão logística (cf. Hosmer e Lemeshow, 1989; Dobson, 1990; e Neter et al., 1996).

O objetivo deste trabalho é justamente buscar um modelo adequado de regressão logística para (i) desvendar as variáveis que influenciam a probabilidade de ocorrer inadimplência na base de doadores do PAM, e (ii) prever a ocorrência de casos de inadimplência à medida que novos doadores são captados. Há uma contribuição relevante para a literatura econômica brasileira, visto que só encontramos dois artigos com preocupações semelhantes⁵ nos seus principais veículos: Maldonado e Orrillo (2006) e Raad e Guedes (2015), ambos tratando de heranças transmitidas de pais para filhos através de modelos teóricos. A analogia que fazemos com um problema típico de risco de crédito permite aplicar uma abordagem quantitativa que já foi adotada na literatura internacional⁶, mas que ainda não foi aplicada a dados brasileiros⁷.

⁵ A saber, o estudo do ato de doar que, segundo Andreoni (1989), reflete o altruísmo dos agentes “... *and include charity and intergenerational transfers*”.

⁶ Desde Andreoni (1989) e o seu modelo de altruísmo impuro, vários autores produziram um sem número de contribuições. Bons exemplos são os trabalhos de Rose-Ackerman (1996), Brekke et al. (2003) e Bekkers e Wiepking (2010). Uma excelente coletânea de artigos pode ser encontrada no “*Handbook of the Economics of Giving, Altruism and Reciprocity*”, obra editada por Serge-Christophe Kolm e Jean Mercier Ythier e publicada em 2006 pela editora Elsevier.

⁷ É necessário destacar a contribuição da Pesquisa Doação Brasil 2015, que em sua fase quantitativa utilizou os resultados de 2230 entrevistas realizadas junto à população urbana brasileira (com idade acima de 18 anos e renda familiar mensal acima de 1 salário mínimo) para definir o perfil clássico do doador institucional em dinheiro. A pesquisa foi coordenada pelo Instituto pelo Desenvolvimento do Investimento Social (IDIS) e visou

3. Modelos de regressão logística

Uma frequente aplicação da teoria dos Modelos Lineares Generalizados (MLGs) – cf. Cordeiro (1986), McCullagh e Nelder (1989), Dobson (1990) e Hardin e Hilbe (2007) – é sua utilização na modelagem de variáveis dependentes do tipo binário 0-1. Existem alguns MLGs para uma variável resposta que só assume dois valores possíveis, indicando a ocorrência ou não de um determinado evento de interesse.

De acordo com Cordeiro (1986), McCullagh e Nelder (1989), Dobson (1990), Neter et al. (1996) e Hardin e Hilbe (2007), um MLG é definido por um trinômio de componentes básicos, sendo um *aleatório*, outro *sistemático* e, por fim, a *função de ligação*. As principais características destes componentes para os MLGs destinados à modelagem de respostas binárias são apresentadas abaixo:

□ *Componente aleatório*: Devido à natureza dos dados, admite-se que:

$$Y_i \sim \text{Ber}(p_i),$$

na qual p_i é a probabilidade de que ocorra sucesso na observação da i -ésima unidade amostral (isto é: $Y_i = 1$), $i = 1, \dots, n$.

□ *Componente sistemático*: Define-se uma transformação linear $T : R^p \longrightarrow R^n$ do vetor de parâmetros $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1})'$. A correspondente matriz de transformação é dada por uma matriz X , cuja primeira coluna é um vetor de 1's (para contabilizar um efeito de intercepto β_0) e as suas demais $p-1$ colunas são formadas por valores de variáveis explicativas. O vetor-imagem em R^n desta transformação é composto dos *preditores lineares* $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1}$.

□ *Função de ligação*: Nota-se que a média de Y_i é dada pela probabilidade de sucesso p_i , a qual deve pertencer ao intervalo $(0,1)$, e o preditor linear η_i é um número real. Uma maneira conveniente de se interligar os dois é escolher uma função de ligação $g(\cdot)$, que tenha como domínio o intervalo citado e como imagem a reta real. Matematicamente, $g : (0,1) \longrightarrow R$ ou, equivalentemente, $g^{-1} : R \longrightarrow (0,1)$, posto que tal função deve ser monótona e diferenciável,

mapear os hábitos de doação dos indivíduos no Brasil. A metodologia adotada na ocasião, porém, não coincide com a utilizada neste artigo.

para não gerar conflitos com o processo de estimação do vetor de coeficientes β . Escolhida a função de ligação, tem-se $g(p_i) = \eta_i \Leftrightarrow p_i = g^{-1}(\eta_i)$, $i = 1, 2, \dots, n$.

Uma opção frequentemente adotada para a função de ligação para MLGs é a dada pela função de ligação *logit*, qual seja $\log_e \left(\frac{p_i}{1-p_i} \right) = \eta_i$, $i = 1, \dots, n$. Invertendo-se esta relação, obtém-se $p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{1}{1 + \exp(-\eta_i)}$, $i = 1, 2, \dots, n$.

O modelo logit apresenta diversas propriedades que o tornam bastante atrativo. Por exemplo, o processo de estimação não envolve maiores complicações computacionais. Outra vantagem deste modelo é a sua maior amplitude de interpretação, já que pode ser de interesse não só analisar a probabilidade p_i , mas também a *chance* definida por $\frac{p_i}{1-p_i}$. No modelo logit, esta chance é facilmente obtida tomando-se anti-logaritmos nos dois lados da função de ligação. Cada um dos coeficientes fornece informação sobre o efeito da variável explicativa correspondente.

Para facilitar a notação, admite-se que o modelo possui uma variável independente (regressão logística *simples*), sendo que as conclusões podem ser estendidas diretamente para o caso com mais variáveis (regressão logística *múltipla*). Também será omitido o sub-índice i , correspondente à identificação da unidade amostral. Logo, utilizar-se-á a escrita $\eta = \beta_0 + \beta_1 x$.

Inicialmente, considere-se a expressão do modelo logit que evidencia a chance:

$$\frac{p}{1-p} = \exp(\eta) = \exp(\beta_0 + \beta_1 x) = \exp(\beta_0) \exp(\beta_1 x). \quad (1)$$

Convenciona-se, a partir de agora, que a expressão em (1), quando aplicado em x , é chamado de Chance 1 e, quando aplicado em $x+1$, é chamado de Chance 2. Ou seja:

$$\text{Chance1} = \exp(\beta_0) \exp(\beta_1 x), \quad (2)$$

$$\begin{aligned} \text{Chance2} &= \exp(\beta_0) \exp(\beta_1 (x+1)) \\ &= \exp(\beta_0) \exp(\beta_1 x + \beta_1) \\ &= \exp(\beta_0) \exp(\beta_1 x) \exp(\beta_1). \end{aligned} \quad (3)$$

Das expressões em (2) e (3), obtém-se a relação desejada entre as duas chances:

$$\text{Chance 2} = \text{Chance 1} \left[\exp(\beta_1) \right]. \quad (4)$$

Como pode ser averiguado em (4), a exponencial de β_1 representa uma constante de proporcionalidade entre as Chances 1 e 2 – e, como esta constante também é recuperada por uma divisão entre as Chances 1 e 2, é muito frequente a atribuição a esta da expressão *razão de chance*. Assim, podem ser tecidas interpretações a respeito da relação entre as chances – ou probabilidades, ou ainda, os valores observados das respostas – e os diferentes níveis de x .

Foi visto, portanto, que o parâmetro $\exp(\beta_1)$ desempenha um papel fundamental na interpretação de um modelo ajustado via regressão logística. Seu estimador de máxima verossimilhança é dado por $\exp(B_1)$, no qual B_1 é o estimador de máxima verossimilhança de β_1 . Cordeiro (1986), McCullagh e Nelder (1989) e Dobson (1990) demonstram que, se o conjunto de dados for relativamente grande, B_1 terá distribuição aproximadamente Normal, implicando que $\exp(B_1)$ terá uma distribuição aproximadamente Log-Normal. Esta última distribuição assintótica é assimétrica. Logo, inferências geralmente são feitas, inicialmente, para β_1 e, após isto, obtêm-se as conclusões para $\exp(\beta_1)$ – vide, por exemplo, Hosmer e Lemeshow (1989) para mais detalhes.

A seleção do modelo estatístico visa encontrar a expressão que envolva o mínimo de parâmetros possíveis a serem estimados e que explique de maneira eficaz a incerteza da variável resposta. Para isso, são estabelecidas estratégias de seleção: testes de significância individuais, teste da razão de verossimilhança e critérios de informação de Akaike. Estes procedimentos serão brevemente revisados a partir de agora.

Geralmente, um *teste de significância individual* (Dobson, 1990) é utilizado para se ter uma primeira percepção de quais são as variáveis explicativas relevantes em um modelo de regressão. As hipóteses do teste, para o caso de se testar a significância da j -ésima variável explicativa, são: $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$. A estatística para este teste é a divisão entre o estimador de máxima verossimilhança do coeficiente correspondente e seu erro padrão; ou seja,

$$t = \frac{B_j}{s.e.(B_j)}. \quad (5)$$

Uma vez que a estatística de teste em (5) tem, sob a hipótese nula H_0 , distribuição assintótica Normal padrão, a hipótese nula é rejeitada se o módulo do valor observado da estatística é superior ao quantil de probabilidade $1 - \frac{\alpha}{2}$ de uma distribuição Normal padrão, sob um nível de significância α .

O teste da razão de verossimilhança, por sua vez, é adequado quando o objetivo é comparar dois modelos (Kleinbaum e Klein, 2010). Estes devem ser *encaixados*, isto é, todas as q variáveis explicativas do *modelo reduzido* (MR) estão também entre as p variáveis explicativas do *modelo completo* (MC), sendo que $q < p$. Supondo sem perda de generalidade que os coeficientes cuja significância conjunta está sendo testada são os k últimos, as hipóteses são: $H_0 : \beta_q = \beta_{q+1} = \dots = \beta_{p-1} = 0$ versus $H_1 : \beta_j \neq 0$ para pelo menos um $j = q, q+1, \dots, p-1$.

A estatística de teste usada é

$$LR = 2[\log L_{MC} - \log L_{MR}], \quad (6)$$

a qual, sob H_0 , tem distribuição assintótica Qui Quadrado com $p-q$ graus de liberdade. Dessa forma, a hipótese nula é rejeitada se o valor observado da estatística de teste em (6) for superior ao quantil de probabilidade $1 - \alpha$ da referida distribuição, sob um nível de significância α .

Por fim, o critério de informação de Akaike (AIC) (Hardin e Hilbe, 2007) se baseia na função de log-verossimilhança, com introdução de um fator de correção como modo de penalização da complexidade do modelo. A estatística deste critério é:

$$AIC = -2[\log L - p], \quad (7)$$

na qual L é o valor da verossimilhança maximizada para o modelo estimado e p é o número de parâmetros do modelo. Na comparação entre dois modelos, aquele que apresentar menor valor da estatística em (7) será aquele considerado mais adequado aos dados.

Quanto ao processo de predição com um MLG para dados binários, elemento que está no âmbito deste artigo, há dois enfoques sobre o grau de dificuldade encontrado. Em outras palavras, a predição de novas observações de uma variável aleatória de Bernoulli modelada por MLG é extremamente fácil por um lado, e um tanto difícil por outro. A predição pode ser tida como fácil, se for levado em consideração que não são demandados construções de intervalos de

predição, sendo somente necessário o ato de observar a probabilidade estimada e reportar o valor predito $\hat{Y}_h = 1$, se for predito que ocorrerá sucesso; e $\hat{Y}_h = 0$, se for predito o contrário. Mas existe a necessidade de se encontrar um *limiar* dentro do intervalo $(0,1)$, a partir do qual obtêm-se os valores preditos da variável resposta, através de comparações, respeitando o seguinte critério:

- Se a probabilidade estimada ultrapassar o limiar, prediz-se $\hat{Y}_h = 1$;
- Se a probabilidade estimada não ultrapassar o limiar, prediz-se $\hat{Y}_h = 0$.

A obtenção deste limiar é totalmente heurística, o que exige a investigação e comparação de procedimentos alternativos. Abaixo, são explicados brevemente dois destes procedimentos:

1. *Utilizar o limiar 0,5 para o processo de predição:* Este é o método mais fácil de ser aplicado, além de ser muito intuitivo. Possui a desvantagem de, possivelmente, não ser o ponto que minimiza a porcentagem de predições erradas.
2. *Buscar o limiar ótimo para minimização da taxa de predições incorretas:* A partir do conjunto de dados utilizado no ajuste, é possível encontrar um limiar que torna o processo de predição o mais consistente possível, no seguinte sentido: encontrado tal limiar ótimo, obtêm-se os valores preditos de acordo com a regra de predição descrita anteriormente e comparam-se estes como os valores observados. O número de coincidências entre os valores observados e seus respectivos preditos (predições corretas) deve ser o maior possível. Uma maneira de se obter o limiar ótimo pode ser a partir de um candidato inicial (exemplos: 0,5, média amostral da variável dependente etc.) e, em seguida, praticar-se uma análise de sensibilidade para valores próximos, sempre visando maximizar as predições corretas. Outra forma de implementar essa busca é mais intensa, contudo muito mais exaustiva/abrangente: criar uma grade de valores no intervalo $(0,1)$ e, para cada valor da grade, fazer do mesmo um candidato a limiar e praticar a predição para cada valor da variável dependente dentro do banco de dados. Aqueles valores da grade que oferecerem taxas percentuais de acertos mais altas podem passar a um “segundo turno” de escolha, o qual levará em consideração as duas taxas de acertos, a dos sucessos (os 1’s) e a dos não-sucessos (os 0’s).

4. Estudo de caso

Ao longo desta seção, serão apresentados os resultados obtidos e as análises de uma regressão logística feita com a base de dados sobre pessoas físicas doadoras da JMN. Os objetivos são: entender quais variáveis do banco de dados têm efeitos estatisticamente significantes para descrever a chance de um doador ser inadimplente; interpretar os efeitos estatisticamente significantes considerando a magnitude e sinal dos correspondentes coeficientes estimados; e prever a variável binária que caracteriza um doador inadimplente (sim ou não) a partir do melhor modelo logístico estimado e selecionado.

O banco de dados, fornecido pelo Sistema Financeiro e Contábil da JMN no decorrente ano de 2015, possui diversas informações a respeito de 77.498 doadores, dentre eles pessoas físicas e jurídicas. Ao analisar o banco de dados fornecido, foram detectados alguns problemas, tais como registros duplicados do mesmo doador e valores faltantes. Após a eliminação das linhas correspondentes aos problemas detectados e a seleção apenas de registros referentes a doadores classificados como pessoas físicas, o banco passou a ter 32.416 linhas.

Quadro 1 – Variáveis dependente e explicativas utilizadas na análise.

Nome	Significado	Descrição
Inadimplente (variável dependente)	Doador com pelo menos 90 dias sem efetuar pagamento do título no valor comprometido	(0) Não
		(1) Sim
Idade	Idade do doador	1 a 102 anos
Estado civil	Estado civil do doador	(C) Casado ou amasiado
		(D) Divorciado
		(S) Solteiro
		(V) Viúvo
Gênero	Gênero do doador	(0) Masculino
		(1) Feminino
Região	Região do Brasil em que reside o doador	(1) Norte
		(2) Nordeste
		(3) Centro-Oeste
		(4) Sudeste
		(5) Sul
Classificação	Código interno da empresa para classificação do doador de acordo com o valor da sua doação	(1) Bronze 1
		(2) Bronze 2
		(3) Bronze 3
		(4) Prata 1
		(5) Prata 2, Ouro 1 e Ouro 3
Cobrança	Forma de pagamento do título de cobrança escolhida pelo doador	(1) Boleto
		(2) Crédito
		(3) Débito
Vencimento	Dia do vencimento do título de cobrança escolhido pelo doador	De 1 a 30

Fonte: Elaboração própria.

Considerando que neste artigo se utilizará o método da validação *fora da amostra* do modelo, no qual é especialmente importante quando a finalidade é a predição de futuros casos de inadimplência, foram selecionados de forma aleatória 10% dos dados para ser usado na validação. Portanto, os 90% restantes, que corresponde a 29.174 registros dos doadores, estão sendo considerados no banco de dados base para os resultados a seguir.

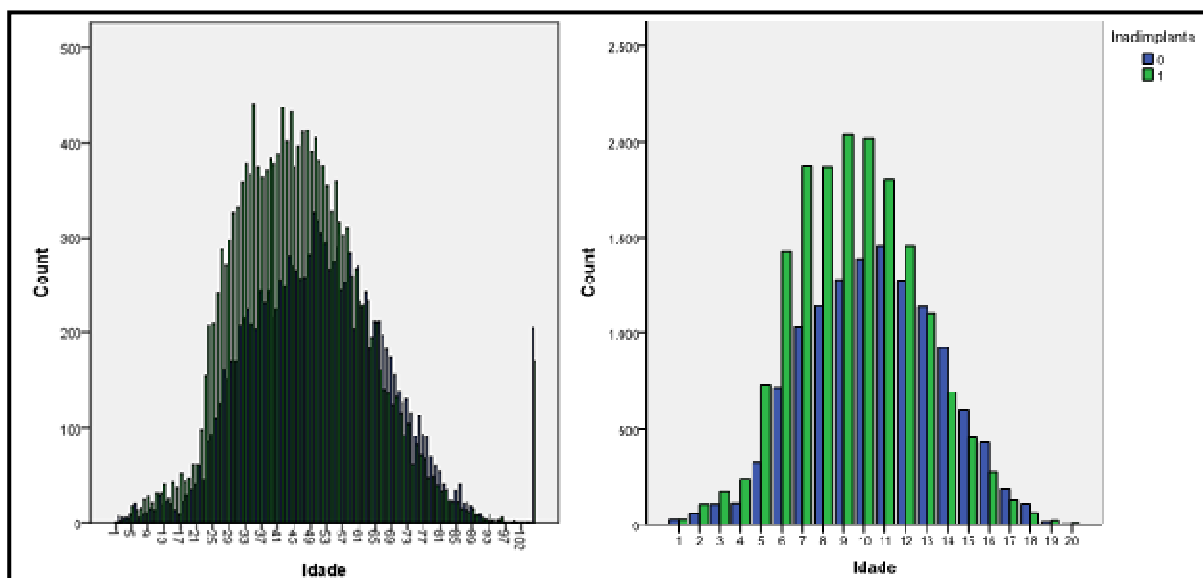
As análises incluem, além da variável dependente “Inadimplente” (variável binária que assume 1 se o doador estiver sem honrar seus compromissos por um período igual ou superior ao de 90 dias), as seguintes variáveis explicativas: Idade, Estado Civil, Gênero, Região, Classificação (tipo de doador), Cobrança e Vencimento. Todas as variáveis estão detalhadamente descritas no Quadro 1. As demais variáveis originalmente disponíveis no

banco de dados foram descartadas pela forte incidência de campos nulos e/ou por se tratarem de variáveis não-categorizáveis trivialmente (textos livres).

4.1 Estatísticas descritivas: análise de frequência dos inadimplentes

Ao longo desta subseção, são apresentadas estatísticas descritivas das variáveis explicativas através dos gráficos de barras, agrupados pela variável dependente Inadimplente. O objetivo é o de reunir elementos iniciais (por serem tão somente visuais) a respeito das influências/efeitos de cada uma das variáveis explicativas na magnitude da probabilidade de inadimplência.

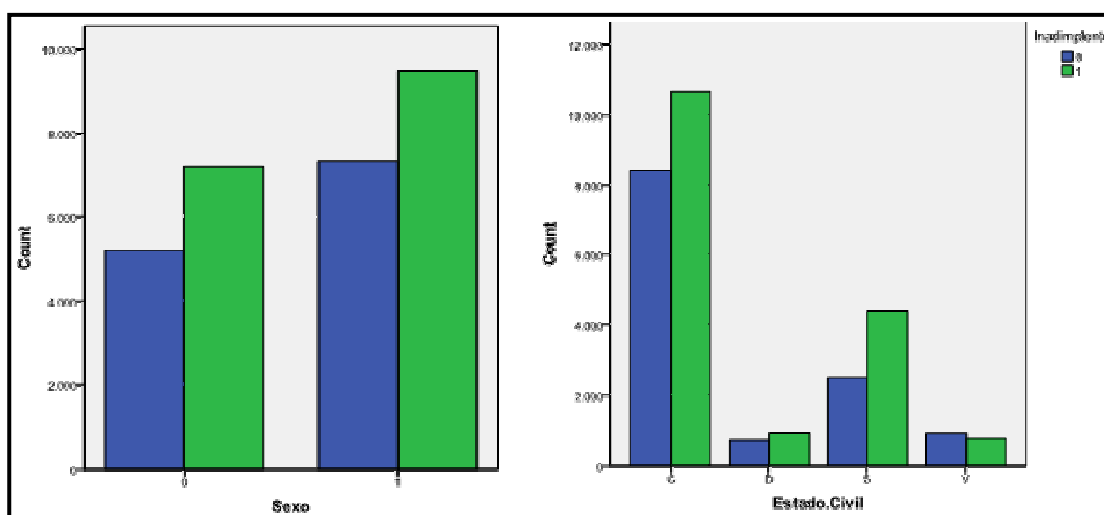
Figura 4 – Gráficos de barras para a variável “Idade” (com diferentes categorizações).



Fonte: Elaboração própria.

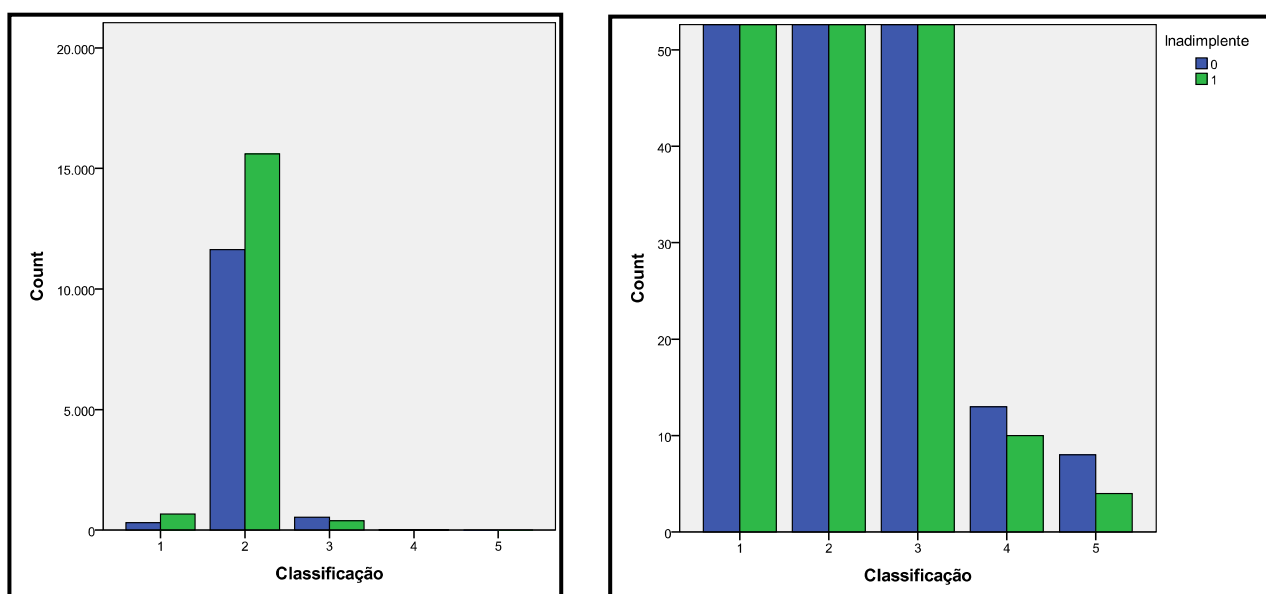
Com base na Figura 4, é possível perceber que grande parte dos doadores se concentram entre as idades de 25 a 70 anos. Os doadores com mais de 65 anos tendem a ser mais adimplentes, enquanto que, dentre os mais novos, observa-se maior frequência de inadimplência. O gráfico da direita na Figura 4 corresponde à variável categorizada de 5 em 5 anos, e do da esquerda corresponde a todos os valores, sem tratamento da variável.

Figura 5 – Gráficos de barras para as variáveis “Gênero” e “Estado Civil”.



Fonte: Elaboração própria.

Figura 6 – Gráficos de barras para a variável “Classificação” (em diferentes escalas).



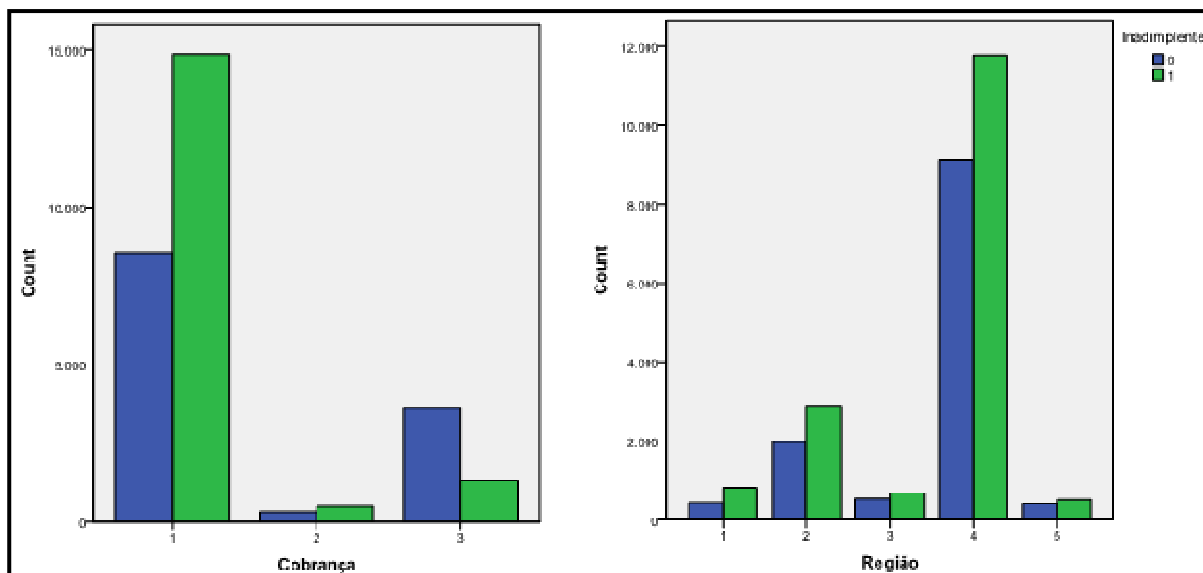
Fonte: Elaboração própria.

Já a Figura 5 sugere que a maioria dos doadores é do gênero feminino e, para ambos os gêneros, a frequência de doadores inadimplentes é maior, quando comparada aos adimplentes. Além disso, a maioria dos doadores são casados e a frequência de inadimplência é maior nos doadores casados, divorciados e solteiros, o que se inverte quando o doador é viúvo.

A Figura 6 apresenta a variável “Classificação” em dois formatos: o segundo painel tem a escala (as contagens), do primeiro painel, ampliadas, visto que as categorias 4 e 5 possuem frequências (tanto de inadimplentes quanto de adimplentes) muito menores do que as demais.

É possível perceber que a grande maioria dos doadores se concentra na classificação Bronze 2 e que a frequência de inadimplência é maior para este grupo. Esta realidade se inverte quando o doador é classificado como Prata 1, Prata 2 e Ouro 1, quando, desta vez, a frequência dos doadores adimplentes é maior.

Figura 7 – Gráficos de barras para as variáveis “Tipo de Cobrança” e “Região”.

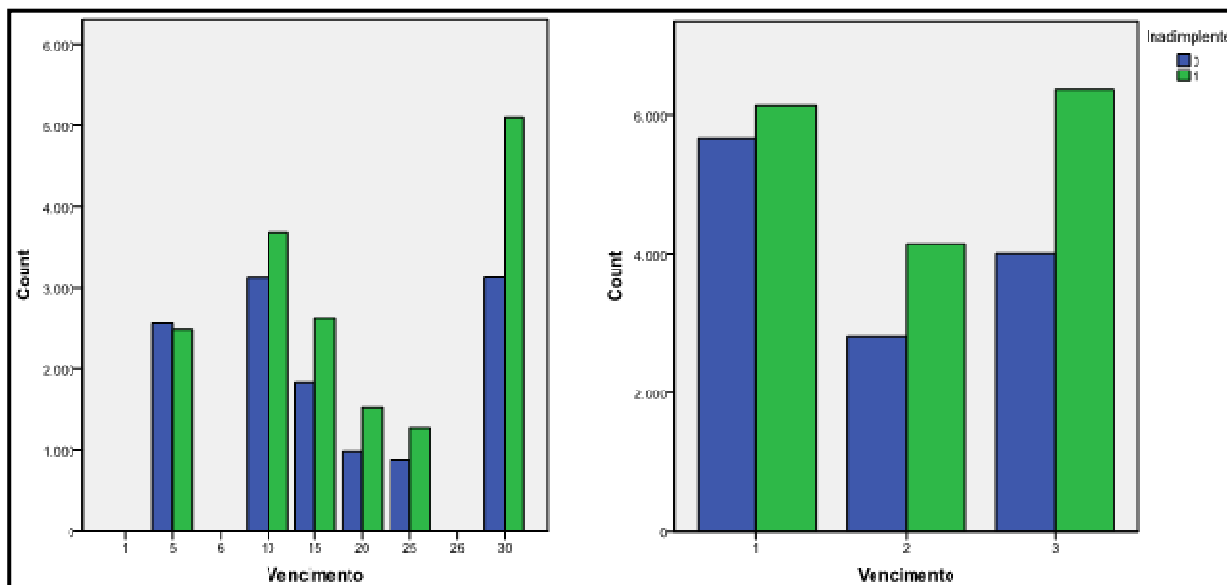


Fonte: Elaboração própria.

Com base na Figura 7, é possível observar que a grande parte dos doadores se concentra no tipo de cobrança por boleto e que é grande a frequência de inadimplentes neste grupo, o que se opõe quando observado os doadores com tipo de cobrança em débito, no qual a grande maioria é adimplente. Além disso, a maior parte dos doadores está na Região Sudeste do Brasil e, em seguida, na Região Nordeste. Para as 5 regiões a frequência de doadores inadimplentes se mostrou maior, quando comparada aos adimplentes; porém, para as duas supracitadas regiões com maior quantidade de doadores, a diferença entre as quantidades de adimplentes e inadimplentes aumenta.

Por fim, na Figura 8 nota-se que os doadores estão bem distribuídos entre os vencimentos no início e final do mês. Porém, a frequência de inadimplência muda, pois no dia 5 a frequência de inadimplência é menor e no dia 30 foi consideravelmente maior, quando comparado a frequência dos doadores adimplentes.

Figura 8 – Gráficos de barras para a variável “Vencimento” (com categorizações no segundo painel: 1 = “do dia 1º ao dia 10”; 2 = “do dia 11 ao dia 20”; e 3 = “do dia 21 até o final do mês”).



Fonte: Elaboração própria.

4.2 Modelagem

Ao longo desta subseção, serão apresentados o ciclo percorrido da modelagem, as interpretações dos resultados e a análise do poder preditivo com validações dentro e fora da amostra.

Para cada variável explicativa categórica, dentre todas aquelas já pormenorizadas no Quadro 1, foram criadas variáveis do tipo dummy. No Quadro 2 estão indicadas as suas categorias de referência – ou seja: as categorias para as quais *não foram criadas* variáveis dummy e com relação às quais os efeitos das demais categorias (dentro de uma variável categórica específica) devem ser comparados (veja explicação, dada na seção 3, sobre a interpretação dos coeficientes do modelo).

No Quadro 3, estão os resultados consolidados da estimação do modelo de regressão logístico com todas as variáveis explicativas (Modelo 1). É possível observar que os teste de significância individual, considerando um nível de significância de 5%, sugerem que as variáveis Classificação Bronze 2, Prata 1 e Cobrança (Crédito) não são significativas para este modelo.

Quadro 2 – Categorias de referência para as variáveis do tipo dummy.

Variável Explicativa Categórica	Categoria de Referência
Estado Civil	Viúvo
Gênero	Masculino
Região	Nordeste
Classificação	Bronze 1
Cobrança	Boleto

Fonte: Elaboração própria.

Quadro 3 – Modelo com todas as variáveis.

Variáveis	Estimativa do Coeficiente	Erro Padrão	Estatística <i>t</i>	<i>p</i> -valor	Razão de Chance
Idade	-,019	,001	521,573	,000	,981
Estado.Civil(Casado)	,116	,057	4,220	,040	1,123
Estado.Civil (Divorciado)	,258	,076	11,572	,001	1,294
Estado.Civil (Solteiro)	,148	,064	5,319	,021	1,159
Gênero Feminino	-,101	,026	14,875	,000	,904
Região (Norte)	-,302	,071	18,275	,000	,739
Região (Centro-Oeste)	-,356	,088	16,284	,000	,700
Região (Sudeste)	-,337	,065	26,644	,000	,714
Região (Sul)	-,375	,095	15,656	,000	,687
Classificação (Bronze 2)	-,020	,075	,072	,788	,980
Classificação (Bronze 3)	-,502	,103	23,862	,000	,605
Classificação (Prata 1)	-,469	,452	1,078	,299	,625
Classificação (Prata 2 ou mais)	-1,289	,624	4,264	,039	,276
Cobrança (Crédito)	-,142	,074	3,648	,056	,868
Cobrança (Débito)	-1,638	,036	2080,789	,000	,194
Vencimento	,015	,001	132,210	,000	1,015
Constante	1,538	,124	152,889	,000	4,656

Fonte: Elaboração própria.

Quadro 4 – Modelo com todas as variáveis e efeitos polinomiais.

Variáveis	Estimativa do Coeficiente	Erro Padrão	Estatística <i>t</i>	<i>p</i> -valor	Razão de Chance
Idade	,077	,009	75,294	,000	1,080
Idade ²	-,002	,000	144,860	,000	,998
Idade ³	,00001	,000	164,037	,000	1,000
Estado.Civil (Casado)	-,042	,058	,520	,471	,959
Estado.Civil (Divorciado)	,135	,077	3,076	,079	1,145
Estado.Civil (Solteiro)	,019	,065	,082	,775	1,019
Gênero Feminino	-,110	,026	17,497	,000	,895
Região (Norte)	-,294	,071	17,176	,000	,746
Região (Centro-Oeste)	-,350	,089	15,565	,000	,705
Região (Sudeste)	-,320	,065	23,969	,000	,726
Região (Sul)	-,374	,095	15,458	,000	,688
Classificação (Bronze 2)	-,241	,079	9,197	,002	,786
Classificação (Bronze 3)	-,726	,106	46,531	,000	,484
Classificação (Prata 1)	-,712	,452	2,487	,115	,491
Classificação (Prata 2 ou mais)	-1,545	,626	6,094	,014	,213
Cobrança (Crédito)	-,142	,075	3,623	,057	,867
Cobrança (Débito)	-1,655	,036	2098,699	,000	,191
Vencimento	,083	,026	9,719	,002	1,086
Vencimento ²	-,004	,002	5,509	,019	,996
Vencimento ³	,00007	,000	4,748	,029	1,000
Constante	,144	,204	,501	,479	1,155

Fonte: Elaboração própria.

No Quadro 4, estão os resultados consolidados da estimação do modelo de regressão logística com todas as variáveis explicativas mais efeitos polinomiais para as variáveis numéricas Idade e Vencimento (Modelo 2). É possível observar que os testes de significância individual, considerando um nível de significância de 5%, inicialmente indicam que as variáveis Estado Civil Casado, Divorciado e Solteiro, Classificação Prata 1, Cobrança (Crédito) e a Constante não são significativas para este segundo modelo.

Para acessar a significância conjunta dos termos polinomiais adicionais, foi feito o teste da razão de verossimilhança, com o objetivo de comparar os Modelos 1 e 2 estimados. No Quadro 5, é possível observar a estatística do teste da razão de verossimilhança e o *p*-valor

deste teste, o que leva a concluir que o modelo completo com todas as variáveis explicativas e os efeitos polinomiais (Modelo 2) é o mais adequado. O critério de informação de Akaike confirma essa conclusão, pois o valor da estatística para o Modelo 2 é menor.

Quadro 5 – Comparação entre os Modelos 1 e 2.

	Estatística LR	p-valor	AIC
Modelo 1	-	-	36535,93
Modelo 2	180,250	$2,2 \times 10^{-16}$	36363,68

Fonte: Elaboração própria.

Quadro 6 – Comparação entre modelos.

	Estatística LR	p-valor	AIC
Modelo 2 ⁸	-	-	36363,68
Modelo 3	2,503	0,1136	36364,18
Modelo 4	11,513	0,0092	36369,19
Modelo 5	3,590	0.0581	36365.27
Modelo 6	17,900	0.0031	36371.58

Fonte: Elaboração própria.

No Quadro 6, é feita a comparação do modelo vigente, que é o Modelo 2, com outros modelos sem as variáveis que não foram significativas de acordo com os testes de significância individuais. O Modelo 3 se refere ao modelo com todas variáveis explicativas do Modelo 2, retirando somente a variável Classificação Prata 1. O Modelo 4 se refere ao modelo com todas as variáveis explicativas do Modelo 2, retirando somente as variáveis de Estado Civil Casado, Divorciado e Solteiro. O Modelo 5 se refere ao modelo com todas as variáveis explicativas do Modelo 2, retirando somente a variável Cobrança Crédito. Já o Modelo 6 se refere ao modelo com todas as variáveis explicativas do Modelo 2, retirando as 5 variáveis (Classificação Prata 1, Estado Civil Casado, Estado Civil Divorciado, Estado Civil Solteiro e Cobrança Crédito) que não foram significativas nos testes de significância individual. Ao observar a Tabela 6, é possível concluir, com base nos testes da razão de verossimilhança e nos critérios de informação de Akaike, que o modelo mais adequado é o

modelo completo (Modelo 2). Apesar de os testes da razão de verossimilhança, que comparam o Modelo 3 e Modelo 5 com o Modelo 2, indicarem, sob um nível de significância de 5%, que os modelos reduzidos (Modelos 3 e 5) são tão adequados quanto o modelo completo (Modelo 2), o critério de informação de Akaike aponta o Modelo 2 como sendo o modelo mais adequado. Portanto, o modelo selecionado para este artigo é o Modelo 2 com todas as variáveis explicativas e os efeitos polinomiais.

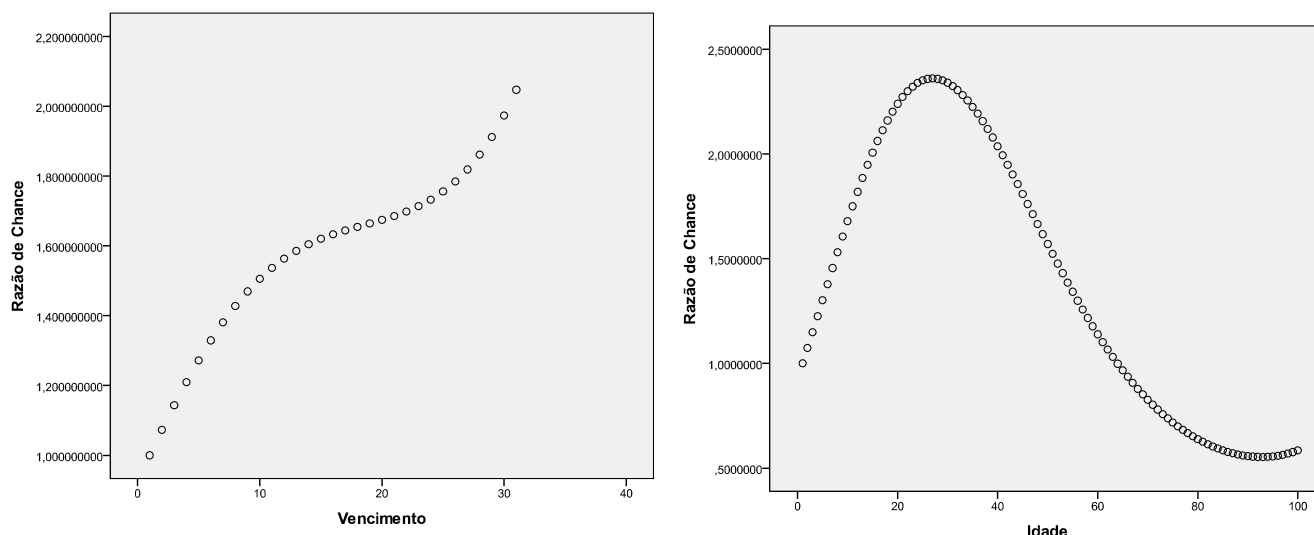
Os resultados para o modelo selecionado, observados no Quadro 4, propiciam as seguintes interpretações de razão de chance (cf. seção 3) para as variáveis categóricas. Note que os acréscimos ou decréscimos nas chances, na medida em que os níveis das variáveis categóricas mudam, vão ao encontro das evidências iniciais sugeridas na subseção 4.1, com os gráficos de barra:

- Em comparação com os doadores viúvos, é estimado para os doadores divorciados um acréscimo de quase 15% na chance de ocorrer inadimplência, e um acréscimo de 2% para doadores solteiros. Já para os doadores casados, é estimado um decréscimo de 4% na chance de ocorrer inadimplência.
- Em comparação com os homens, é estimado para as mulheres um decréscimo de 10% na chance de ocorrer inadimplência.
- Em comparação com os doadores residentes da Região Nordeste, é estimado para os doadores das demais regiões um decréscimo na chance de ocorrer inadimplência, destacando-se a Região Sul com o decréscimo de 31%.
- Em comparação com os doadores de classificação Bronze 1, é estimado para os doadores das demais classificações um decréscimo na chance de ocorrer inadimplência, destacando-se expressivamente os doadores de Classificação Prata 2, Ouro 1 ou Ouro 3, com o decréscimo de quase 79%.
- Em comparação com os doadores cujo tipo de cobrança se dá em forma de boleto, é estimado para os doadores com o demais tipo de cobrança um decréscimo na chance de ocorrer inadimplência, destacando-se fortemente o tipo de cobrança em débito com um decréscimo de mais de 80%.

⁸ Modelo completo para os testes da razão de verossimilhança para os demais modelos (vide seção 3).

Para as variáveis numéricas “Idade” e “Vencimento”, foram feitos gráficos das razões de chance conforme pode ser observado na Figura 9.

Figura 9 – Gráficos de dispersão para as Razões de Chance e as variáveis “Idade” e “Vencimento”.



Fonte: Elaboração própria.

Na sequência, são tecidas algumas interpretações que complementam a informação da Figura 9 (também em forte sintonia com o que se viu anteriormente nos gráficos de barra da subseção 4.1):

- Em comparação com doadores com o Vencimento para o dia 5, é estimado para os doadores com Vencimento para o dia 30 um acréscimo de 55% na chance de ocorrer inadimplência.
- Em comparação com os doadores com Idade de 65 anos, é estimado para os doadores com Idade de 27 anos um expressivo acréscimo de 144,22% na chance de ocorrer inadimplência.

Esses resultados podem ser comparados com aqueles gerados pela Pesquisa Doação Brasil 2015, muito embora a amostra utilizada nessa pesquisa fora construída de maneira diversa da utilizada para formar a amostra deste artigo. Com efeito, a amostra, aqui usada, é composta de pessoas que já assumiram o compromisso de contribuir financeiramente (e de forma sistemática) para uma obra social ligada à Convenção Batista Brasileira (CBB); enquanto que a amostra utilizada na Pesquisa Doação Brasil 2015 englobava pessoas que

fizeram algum tipo de doação (bens, trabalho voluntário ou dinheiro) para organizações sociais, obras sociais de igrejas ou comunidades, grupos organizados e campanhas de captação para calamidades ao longo do ano de 2015 – mas também inclui aquelas que não o fizeram. Falando informalmente, o problema modelado neste artigo trata de decisões condicionais a um compromisso já assumido, enquanto que a Pesquisa Doação Brasil 2015 trata de decisões incondicionais.

Apesar dessas diferenças, as duas pesquisas confirmam que renda, gênero, idade e região do país na qual o indivíduo vive são fatores importantes. Na Pesquisa Doação Brasil 2015, a propensão a fazer uma doação em dinheiro cresce com a renda e com a idade, é maior para indivíduos do gênero feminino e depende da região de residência, sendo maior no Nordeste e no Sudeste⁹. Os nossos resultados indicam que a probabilidade de se manter um compromisso de doação em dinheiro assumido com uma entidade filantrópica ligada a CBB é maior se o indivíduo é mulher e/ou se a sua idade supera 60 anos. A influência das *dummies* de região e classificação (que depende do valor monetário das contribuições prometidas) é tal que o *default* é mais provável na região Nordeste e menos provável se o indivíduo assume um compromisso de maior valor. Esses resultados são compatíveis com a influência do fator renda identificada na Pesquisa Doação Brasil 2015, pois tanto o valor do compromisso assumido (que pode ser proporcional à renda do indivíduo e da sua família), quanto a região na qual ele/ela reside (que define uma expectativa incondicional da renda que ele/ela recebe¹⁰), são variáveis ligadas à renda do agente pesquisado.

Quanto aos resultados sobre o poder preditivo, no Quadro 7, é possível observar uma taxa de 66,4% de acertos globais dos casos *dentro da amostra* e uma excelente taxa de 83,1% de acerto quando os doadores são inadimplentes.

Para validar esses resultados, foi feita um estudo *fora na amostra* com dados previamente retirados do banco de dados (vide detalhes no começo desta seção 4). Conforme se observa no Quadro 8, todas as taxas praticamente mantiveram-se nos mesmos patamares daquelas observadas no Quadro 7.

Por fim, com o objetivo de tentar melhorar o poder preditivo dos casos de inadimplência, foram estimados alguns modelos usando as interações entre as seguintes variáveis: Idade e Estado Civil; Idade e Gênero; Idade e Vencimento; e Estado Civil e Gênero. Ao observarem-se os resultados de predição, não se notou uma melhora significativa;

⁹ Outros resultados interessantes levantados pela pesquisa é o fato de essa propensão depender da escolaridade, do sentimento religioso e do credo específico do indivíduo.

¹⁰ E, aqui, cumpre lembrar que a renda per capita da região Nordeste é a menor do país.

além do mais, modelos de regressão que abrangem variáveis com efeitos interativos em suas especificações são de interpretação mais difícil. Portanto, o modelo escolhido continua sendo o Modelo 2, com todas as variáveis explicativas e com termos polinomiais das duas únicas variáveis numéricas, e sem quaisquer efeitos interativos.

Quadro 7 – Validação dentro da amostra.

Observados		Estimados		
		Inadimplente		Porcentagem de Acerto
		0	1	
Inadimplente	0	5506	6995	44,0
	1	2812	13860	83,1
Porcentagem de Acerto Global				66,4

Fonte: Elaboração própria.

Quadro 8 – Validação fora da amostra.¹¹

Observados		Estimados		
		Inadimplente		Porcentagem de Acerto
		0	1	
Inadimplente	0	632	759	45,4
	1	328	1523	82,2
Porcentagem de Acerto Global				66,5

Fonte: Elaboração própria.

5. Conclusão

Este artigo visa ao entendimento de taxas de inadimplência de pessoas físicas quanto às suas comprometidas doações para instituições filantrópicas. O estudo de caso envolveu um conjunto de dados, da Junta de Missões Nacionais da Convenção Batista Brasileira, referentes a pessoas físicas doadoras e a forma de parceria estabelecida no instante que o doador adota um projeto da instituição e começa a contribuir. As variáveis analisadas que se mostraram como as principais determinantes da inadimplência foram: Idade, Estado Civil, Gênero, Região, Classificação (do tipo de doador), Cobrança e Vencimento. Além dessas, foram

¹¹ Estimação usando o limiar 0,510 (o mesmo limiar ótimo previamente encontrado).

introduzidos alguns efeitos polinomiais para as variáveis numéricas Idade e Vencimento. Os testes de significância individuais para cada uma dessas variáveis, considerando um nível de significância de 5%, inicialmente indicaram que algumas delas não seriam significativas para o modelo; porém, os testes de comparação de modelos (testes da razão de verossimilhança) e os resultados do critério de informação de Akaike indicaram que o melhor modelo seria com todas as variáveis, mais os efeitos polinomiais. Para o modelo selecionado, foram feitas validações dentro e fora da amostra, observando-se taxas de acerto de predição capazes de excelente antecipação dos casos de inadimplência, os quais são os mais importantes de serem preditos, levando-se em consideração a contextualização do problema como sendo análogo a um de risco de crédito (cf. seção 2).

Diante dos resultados obtidos, as recomendações são as de que se desenvolvam ações específicas para o perfil de doadores inadimplentes que podem ser caracterizados por doadores jovens, do gênero masculino, solteiros, com doações de pequenos valores, pagando em forma de boleto, na Região Norte, com datas de vencimento mais próximas do final do mês; já os adimplentes podem ser caracterizados por doadores com mais de 65 anos, do gênero feminino, sendo viúva ou casada, com doações de maiores valores, pagando em forma de débito, residente na Região Sul e com o vencimento próximo ao início do mês. Essas ações devem focar a conscientização dos doadores inadimplentes quanto à importância de tê-los como parceiros nos projetos da empresa e o que se poderia fazer com a soma dessas doações prometidas – mas, efetivamente, não-recebidas. Quanto ao perfil dos doadores adimplentes, podem ser feitas pesquisas exaustivas sobre onde está este público e o desenvolvimento de ações específicas para que se encontrem mais doadores com este perfil. As predições futuras serão fundamentais para o planejamento orçamentário da empresa, pois através dessas predições poder-se-á estimar com mais qualidade o montante de doações com as quais a empresa poderá contar.

Estudos futuros com o objetivo de melhorar os resultados podem ser feitos. Tais quais, sugere-se a estimação de modelos lineares generalizados baseados nas distribuições Poisson e/(ou Binomial Negativa) para contagem de meses de inadimplência de um doador. Desta forma, estaria sendo usada uma variável resposta mais completa, pois a variável resposta usada neste artigo é uma transformação não-biunívoca da contagem de meses de inadimplência, e este tipo de categorização implica inevitável perda de informação. Com um novo modelo baseado na contagem de meses de inadimplência (nenhum mês, 1 mês, 2 meses,

3 meses, 4 meses, 5 meses, 6 meses etc.), são tarefas viáveis fazer predição de inadimplência também com possível inclusão de efeitos interativos e buscar limiares ótimos. A predição, desta vez, pode melhorar, pois seria consequência de uma modelagem com mais informações na variável resposta.

Outra tarefa futura, tanto com um modelo de regressão logística quanto com um modelo para contagens de meses, é aprofundar a busca por efeitos interativos, testando efeitos com três variáveis, com o intuito de melhorar a predição dos adimplentes. Nesse processo, mostrar-se-ão muito úteis procedimentos mais refinados do que gráficos de barras, tais quais: (i) box plots e/ou histogramas, além de testes de diferença de média (entre os grupos de adimplentes e inadimplentes) para as variáveis explicativas numéricas; e (ii) tabelas e contingências e testes Qui-Quadrado para as variáveis categóricas, com o intuito de estudar as distribuições entre adimplentes e inadimplentes para cada categoria da variável. Com esses procedimentos, ficará bem mais factível desvendar o real poder de discriminação das variáveis explicativas escolhidas e dos possíveis efeitos interativos.

Outra extensão certamente de interesse é, com o melhor modelo, usar algum procedimento de otimização para minimizar a probabilidade de inadimplência. Com isso, seriam encontrados os patamares ótimos das variáveis explicativas e, assim, o perfil ótimo para os doadores de instituições filantrópicas no Brasil.

Referências

AMARAL, O. Á. Entrevista pessoal e por correspondência por David Allen Bledsoe entre 25 de setembro e 03 de outubro. Rio de Janeiro, RJ, 2013.

ANDREONI, J. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *The Journal of Political Economy*, 97 (6), p.1447-1458, 1989.

BEKKERS, R.; WIEPKING, P. A Literature Review of Empirical Studies of Philanthropy: Eight Mechanisms that Drive Charitable Giving. *Nonprofit and Voluntary Sector Quarterly*, 40 (5), p.924-973, 2010.

BESSIS, J. Risk management in banking. Chichester: John Wiley e Sons, 1998.

- BLATT, A. Avaliação de Riscos e Decisões de Crédito: um enfoque prático. Nobel, 1999.
- BLEDSON, D. A. Cooperação e Conexão Missionária dos Batistas Brasileiros, 1ª edição. Convicção, 2014.
- BREKKE, K. A.; KVERNDOKK, S.; NYBORG, K. An economic model of moral motivation. *Journal of Public Economics*, 87 (9-10), p.1967-1983, 2003.
- CAOQUETTE, J. B.; ALTMAN, E. I.; NARAYANAN, P. Gestão do risco de crédito: o próximo grande desafio financeiro. Qualitymark, 1999.
- CORDEIRO, G. Modelos Lineares Generalizados. Campinas: VII Simpósio Nacional de Probabilidade e Estatística, 1986.
- DOBSON, A. J. An Introduction to Generalized Linear Models. 2ª edição. Chapman e Hall/CRC, 1990.
- HARDIN, J. W.; HILBE, J. M. Generalized Linear Models and Extensions. 2ª edição. Stata Press, 2007.
- HANDBOOK OF THE ECONOMICS OF GIVING, ALTRUISM AND RECIPROCITY, editado por Serge-Christophe Kolm e Jean Mercier Ythier, Elsevier, 2006.
- HOSMER, D. W.; LEMESHOW, S. Applied Logistic Regression. John Wiley e Sons, 1989.
- JOHNSON, R. A.; WICHERN, D. W. Applied Multivariate Statistical Analysis, 6ª edição, Prentice Hall, 2007.
- KLEINBAUM, D. G.; KLEIN, M. Logistic Regression: A Self-Learning Text. 3ª edição. Springer, 2010.
- MALDONADO, W. L.; ORRILLO, J. Two-Sided Altruism in an OLG Model with Incomplete Markets and Default. *Brazilian Review of Econometrics*, 26 (2), p.291-308, 2006.

McCULLAGH, P.; NELDER, J. Generalized Linear Models. 2ª edição. Chapman and Hall, 1989.

MESQUITA, M. Missionários e Recursos. Betânia, 2011.

NETER, J.; KUTNER, M.; NACHTSHEIM, C.; WASSERMAN, W. Applied Linear Statistical Models. 4ª edição. Times Mirror Higher Education Group, 1996.

OLIVEIRA, Z. M. 100 anos da Junta de Missões Nacionais da CBB, 1ª edição. Convenção Batista Brasileira, 2007.

PESQUISA DOAÇÃO BRASIL 2015. Realização: Instituto pelo Desenvolvimento do Investimento Social (IDIS) e Instituto Gallup.

Endereço: <http://idis.org.br/pesquisadoacaobrasil/>.

RAAD, R. J.; GUEDES, G. R. Private Transfer Choices under Uncertainty in Human Capital. Revista Brasileira de Economia, 69 (1), p.105-124, 2015.

ROSE-ACKERMAN, S. Altruism, Nonprofits and Economic Theory. Journal of Economic Literature, 34 (2), p.701-728, 1996.

SANTOS, M. H. L. Duas Vidas Transformadas, Um Lar Edificado, Jesus Glorificado!. A Pátria Para Cristo, nº 268, 2015

SOUZA, S. O. Entrevista em áudio por David Allen Bledsoe em 17 de setembro. Rio de Janeiro, RJ, 2013.