

UM ESTUDO SOBRE AS VARIÁVEIS QUE IMPACTAM A INADIMPLÊNCIA NO CRÉDITO CONCEDIDO PARA PROJETOS IMOBILIÁRIOS

Fabio Cesar Kothe Jannuzzi

Universidade Estácio de Sá – Av. Pres. Vargas, 642 – sl. 2207 – Centro – Rio de Janeiro (RJ)
fabioj@uol.com.br

Marco Aurélio Carino Bouzada

Universidade Estácio de Sá – Av. Pres. Vargas, 642 – sl. 2207 – Centro – Rio de Janeiro (RJ)
marco.bouzada@estacio.br

Resumo

O presente estudo de caso procura identificar as variáveis que impactam a inadimplência no crédito concedido para projetos imobiliários de reforma ou ampliação de unidade residencial própria. A revisão de literatura versou sobre os temas de Crédito, Risco, Análise de Crédito e *Credit Scoring*. O uso da técnica de regressão logística permitiu verificar a relação entre a inadimplência e as características do tomador e, desta forma, a definição de um modelo de decisão - *credit scoring* - com média de acurácia de 74%. Tal modelo considerou tanto variáveis quantitativas como qualitativas (como faixa de renda, faixa etária e forma de pagamento), que precisaram ser quantificadas com uso de variáveis *dummy*. Entretanto, a reprovação estatística do modelo de escoragem do pedido remete à necessidade de ampliação na coleta de informações acerca dos tomadores de crédito da instituição em estudo. Adicionalmente, foi possível verificar a rejeição das duas hipóteses de pesquisa: a de que cada variável, isoladamente, seria capaz de impactar a inadimplência; e a de que as variáveis, em conjunto, seriam capazes de compor um modelo de *credit scoring* estatisticamente validado.

Palavras-chave: crédito, concessão de crédito, administração de risco, *credit scoring*, regressão logística.

Abstract

This case study tries to identify the variables that impact delinquency rates on credit granted for real estate projects, concerning residential reform or expansion. The literature about Credit, Risk, Credit Analysis and Credit Scoring was reviewed. The use of logistic regression allowed to verify the relationship between delinquency and the borrower characteristics and thus to obtain a decision model - *credit scoring* - with an average accuracy of 74%. This model considered both quantitative and qualitative variables (such as income level, age and payment method), which needed to be quantified through the use of dummy variables. However, credit scoring model statistical failure implies the need to expand the information collection about borrowers within the institution being studied. Additionally, we could reject both research hypotheses: (i) that each variable alone would be able to impact the delinquency, and (ii) that the variables, together, would be able to compose a credit scoring model statistically validated.

Key-words: credit, credit grant, risk management, credit scoring, logistic regression.

1 INTRODUÇÃO

De acordo com Machado (2010), no primeiro trimestre de 2010, o IBGE (Instituto de Geografia e Estatística) informou que a economia brasileira voltou a crescer esse ano com alta de 9%. Os Indicadores Serasa Experian de Inadimplência do Consumidor também apresentaram crescimento, com variação de 4,3%, entretanto, em queda quando comparado com os primeiros meses de 2009, um momento crítico da crise no país.

Ainda segundo a autora, a crise financeira não é a única responsável pelo aumento da inadimplência; há muitos fatores que influenciam nesta situação temporária do consumidor, tais como falta de condições financeiras de arcar com seus débitos, desemprego, doença ou morte de familiares, entre outros.

O aumento das transações no mercado imobiliário, dito aquecido, pelo aumento da disponibilidade de crédito imobiliário e a redução de taxas de juros praticadas no mercado financeiro incentivam a procura por crédito para a realização de reformas e ampliações em unidades residenciais.

Mas as incertezas estão sempre presentes no mundo real e este fato concorre para dificultar o acerto em uma tomada de decisão via intuição. Desta forma, é relevante o uso de técnica estatística para identificar as variáveis que impactam a inadimplência no crédito à pessoa física – ainda mais quando as informações relevantes estão disponíveis, só carecendo de um método sistemático para contemplá-las, como no caso em questão - a fim de evitar o comprometimento da continuidade da operação da instituição concedente e estabelecer uma vantagem competitiva para essa organização.

Para o presente estudo foram consideradas as situações de (in)adimplência ocorridas em financiamentos imobiliários ativos em 2010, ou seja, aqueles em vigência no período.

A base de dados da instituição em estudo permitiu que fossem consideradas, apenas, as seguintes variáveis disponíveis: a renda do tomador do crédito, sua idade e endereço residencial, o valor do crédito concedido, o valor da prestação e o prazo e a forma de pagamento.

Em face de peculiaridades afetas aos servidores públicos integrantes da Autarquia Federal, principalmente, a estabilidade, não foram consideradas variáveis macroeconômicas na pesquisa.

Foi estabelecido para a condução do estudo o objetivo de identificar as variáveis que impactam a inadimplência no crédito concedido para projetos imobiliários, ou seja, para situações em que o tomador do crédito pertencente à Autarquia Federal estudada na pesquisa deseja crédito para realizar reforma ou ampliação de unidade residencial própria.

As seguintes hipóteses de pesquisa foram formuladas no estudo:

- I. as seguintes variáveis, consideradas isoladamente, impactam a inadimplência no crédito concedido para financiamentos imobiliários a pessoas físicas pertencentes à Autarquia Federal: a renda, a idade e o endereço residencial do tomador do crédito; o valor do crédito e da prestação; e o prazo e a forma de pagamento;
- II. as variáveis listadas acima, consideradas em conjunto, permitem estabelecer uma ferramenta de *credit scoring* que estime, satisfatoriamente, a probabilidade de ocorrência de inadimplência futura no crédito concedido para financiamentos imobiliários a pessoas físicas pertencentes à Autarquia Federal.

2 REFERENCIAL TEÓRICO

O crédito é um ato de vontade no qual alguém cede parcela de seu patrimônio a outrem na expectativa de que retorne a sua posse integralmente, após decorrido o prazo estipulado (SCHRICKEL, 2000).

O crédito é importante propulsor da economia, pois possibilita que pessoas, ainda que com recursos modestos, adquiram casas, carros e bens de consumo e isso, por sua vez, cria empregos e aumenta o volume de oportunidades econômicas (CAOUILLE, 2009).

Para as pessoas físicas, o crédito visa suprir necessidades imediatas de caixa ou para antecipar consumo ou investimento (SECURATO, 2007). Para o financiamento de suas necessidades, os clientes podem recorrer à obtenção de uma linha de crédito pontual (SANTOS, 2009), o contrato de crédito, que indica o direcionamento que o cliente dará para os recursos financeiros, seja para gastos com moradia (ampliação ou construção de imóveis) saúde, educação ou aquisição de bens (imóveis, eletrodomésticos, veículos etc).

Entretanto, ao conceder um crédito, o gerente de crédito está comprando um risco (SANTOS, 2009): a qualquer momento, acontecimentos imprevistos e adversos podem afetar as empresas e pessoas físicas, reduzindo a probabilidade de recebimento do crédito.

A inadimplência significa a falta de cumprimento de uma obrigação (HOUAISS, 2009) e pode consistir, no risco de crédito, em um eventual atraso de pagamento (MATIAS, 2009).

A determinação do risco de inadimplência constitui-se em uma das principais preocupações dos credores, pela possibilidade de ocorrência de perdas financeiras que poderão prejudicar a liquidez (capacidade de honrar dívidas com os supridores de capital) e a captação de recursos nos mercados financeiros e de capitais (SANTOS, 2009).

A atividade de análise de crédito presta auxílio à decisão de concessão ou não de crédito em um cenário de incertezas, constantes mutações e informações incompletas, pela recomendação da melhor estruturação e tipo de empréstimo a conceder, à luz das necessidades financeiras do solicitante, dos riscos identificados e mantendo, adicionalmente, sob perspectivas, a maximização dos resultados da instituição (SCHRICKEL, 2000).

Os parâmetros básicos para orientar a análise para concessão do crédito à pessoa física seriam (SECURATO, 2007): Caráter (ou conceito), indica a determinação de pagar do tomador; Capacidade de pagamento, é a renda da pessoa física; Capital, é o patrimônio pessoal do solicitante; Condições, fatores macro ou microeconômicos que influenciam na concessão de crédito; Colateral, garantias que o solicitante coloca à disposição do credor como alternativa de saída para o caso de não vir a cumprir com as obrigações pecuniárias previstas contratualmente.

Nesse sentido, o dossiê básico de crédito de pessoa física seria composto por: a) ficha cadastral; b) comprovantes de renda (contracheques, declaração de Imposto de Renda ou outros documentos) e residência (contas de luz, água ou telefone); c) eventuais comprovantes relativos aos bens declarados na posição patrimonial, bem como, dependendo do valor do crédito concedido, certidões negativas de ônus e alienações; d) cópia autenticada de documentos pessoais (CPF, RG); e e) pela avaliação do risco de crédito.

Para realizar a análise de crédito, as empresas podem recorrer ao uso de técnica subjetiva, baseada no julgamento humano, ou de técnica objetiva, baseada em procedimentos estatísticos para pontuação de propostas de crédito (*credit scoring*) (SANTOS, 2009).

Os modelos de *credit scoring* (escoragem de crédito) recaem em modelos de escoragem do pedido (*application scoring models*) e modelos de escoragem comportamental (*behavioral scoring models*) (CAOINETTE, 2009).

Os escores de crédito baseiam-se em pesos obtidos estatisticamente para características financeiras e creditícias importantes, visando prever se um cliente pagará o crédito solicitado em dia. Ao final do procedimento resulta um escore que mede a solidez geral do cliente, sendo utilizado na tomada de decisão de aceitar ou rejeitar o pedido de crédito (GITMAN, 2004).

O Quadro 1 relaciona as características do tomador de empréstimo utilizadas em vários sistemas de *credit scoring* e as classifica como fatores de baixa ou alta pontuação.

A pressuposição nos modelos de *credit scoring* é de que exista uma métrica que separe os créditos bons dos maus, dividindo-os em duas distribuições distintas, conforme a Figura 1 (CAOINETTE, 2009).

A Figura 1 mostra a distribuição das contas “boas” e “más” a partir da pontuação de crédito, onde podemos verificar como trabalham as pontuações de corte.

Estabelecida uma pontuação mínima de 300 pontos, a pontuação de corte, a empresa

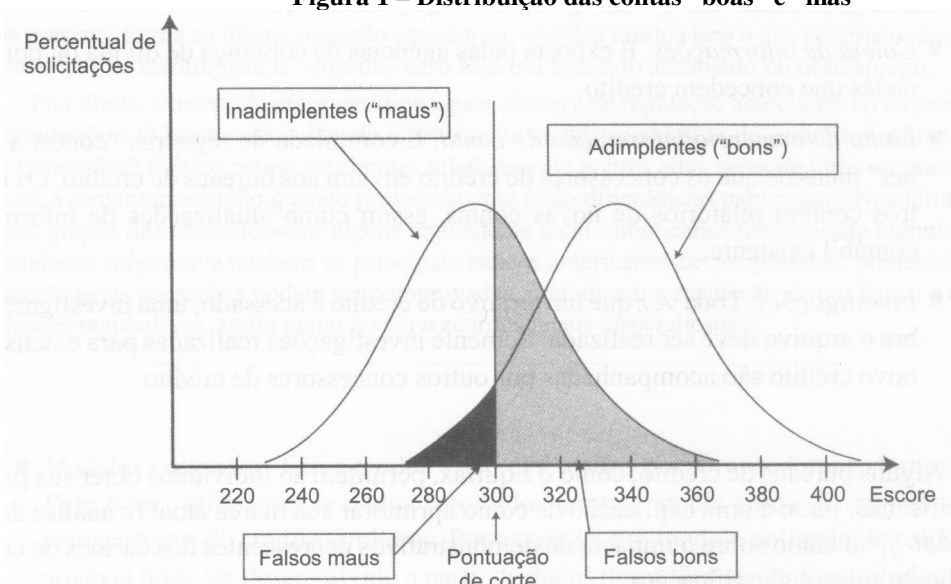
que usasse o sistema de escoragem evitaria emprestar dinheiro para o quadro de maus clientes à esquerda da linha vertical, mas renunciaria ao quadro menor de contas boas à esquerda da linha (CROUHY, 2007), ou seja, cálculos inadequados de pontos de corte podem resultar em rejeição de crédito a clientes potenciais e aprovação de crédito para clientes de alto risco (SANTOS, 2009).

Quadro 1 – Características do devedor e pontuação

Características do devedor	Baixa pontuação	Alta pontuação
Residência	Alugada	Própria
Tempo de residência na região	< 6 meses	> 10 anos
Nível de receita bruta anual	< U\$ 15,000	> U\$ 100,000
Ocupação	Baixa qualificação	Alta qualificação
Tempo de emprego	< 3 meses	> 10 anos
Número de cartões de crédito	Nenhum	5 ou mais
Empréstimos de empresas financeiras	Vários	Nenhum
Relação dívidas / receita bruta	> 30%	< 5%
Contas correntes ou de poupança mantidas	Nenhuma	Ambas
Idade	< 30 anos	> 50 anos

Fonte: Securato (2007, p. 36)

Figura 1 – Distribuição das contas “boas” e “más”



Fonte: Crouhy (2007, p. 188)

A seleção das melhores variáveis (aqueles atributos capturados no formulário de solicitação de crédito) e dos melhores pesos para determinar os dois grupos com a maior eficiência é realizada por intermédio da utilização de técnica multivariada (CAOINETTE, 2009), que pode ser otimização, análise discriminatória, análise *logit*, análise *probit* ou análise de sobrevivência.

Caouette (2009) afirma que os modelos de *credit score* oferecem muitas vantagens e alguns defeitos.

Como vantagens, o autor cita que os modelos são objetivos e consistentes, bastante simples e de fácil interpretação, as metodologias usadas para construir os modelos são comuns e bem entendidas e, com referência ao consumidor, os modelos podem eliminar práticas discriminatórias nos empréstimos e proporcionar um melhor serviço pela rapidez da aprovação ou não do crédito.

Como defeitos, o autor lembra que, em alguns casos, os modelos apenas automatizam as práticas vigentes, os testes estatísticos podem ser fracos e os modelos podem degradar-se pelo tempo se a população atual diverge da população original, motivo pelo qual os modelos de

credit scoring devem ser continuamente testados e avaliados para garantir que a performance real se aproxime das projeções iniciais.

3 ASPECTOS METODOLÓGICOS

3.1 Coleta de Dados

Os dados referentes às variáveis selecionadas para constituírem o modelo de decisão - *credit scoring* – foram obtidos mediante consulta às informações disponíveis no Sistema de Informações da Autarquia por ocasião da pesquisa de campo.

O período de coleta de dados durou cerca de cinco meses, encerrando-se no mês de julho de 2010. Durante esse período foram coletados dados de aproximadamente 48.000 contratos ativos referentes a cerca de 14.200 indivíduos, eliminando-se aqueles que não satisfaziam os parâmetros da amostra, quais sejam:

a) contratos celebrados até o limite de R\$ 40 mil: foram relacionados apenas contratos de crédito com valores inferiores a R\$ 40 mil, pois valores superiores foram assumidos como exceção ao procedimento da instituição;

b) contratos celebrados até o limite de 48 prestações: foram relacionados apenas contratos de crédito com número de prestações igual ou inferior a 48 prestações, pois valores superiores foram assumidos como exceção ao procedimento da instituição;

c) indivíduos com apenas um contrato ativo: foram relacionados apenas indivíduos com apenas um contrato ativo para evitar desvios decorrentes de multiplicidade de informações sobre o mesmo indivíduo.

Os dados coletados referem-se a operações de crédito concedidas a clientes da instituição e ativas no momento da observação. Todos os créditos concedidos nesse período foram analisados com intuito de verificar a condição de adimplência de cada um nos meses de fevereiro, abril e junho de 2010.

A discriminação das observações realizadas em 2010 é apresentada no Quadro 2.

Quadro 2 – Composição das observações realizadas em 2010

Mês de Referência	Contratos Ativos	Contratos Adimplentes	Contratos Inadimplentes
Fevereiro	15.573	13.047	2.275
Abril	15.814	13.551	2.022
Junho	16.235	14.300	1.784

Fonte: Dados da Pesquisa

Nessa pesquisa, os clientes inadimplentes são caracterizados como indivíduos com atraso superior a 90 dias no pagamento de suas operações.

Para selecionar os indivíduos da amostra foram utilizados métodos de aleatoriedade baseados em números gerados por computador (aplicativo Microsoft Excel).

Os clientes selecionados através do processo de amostragem aleatória foram numerados e tiveram suas informações segregadas e identificadas as suas características pessoais e do contrato, tanto dos adimplentes quanto dos inadimplentes.

Especificamente quanto à amostra de fevereiro de 2010, os indivíduos numerados pares compuseram a Amostra A e os indivíduos numerados ímpares compuseram a Amostra B.

Para a condução da pesquisa foram coletadas quatro amostras formadas por indivíduos selecionados aleatoriamente entre os tomadores de crédito pessoa física de uma instituição:

a) Amostra A-200 clientes (100 adimplentes e 100 inadimplentes) em fevereiro de 2010.

b) Amostra B-200 clientes (100 adimplentes e 100 inadimplentes) em fevereiro de 2010.

c) Amostra C-215 clientes (110 adimplentes e 105 inadimplentes) em abril de 2010.

d) Amostra D-203 clientes (110 adimplentes e 93 inadimplentes) em junho de 2010.

Para a elaboração do modelo de regressão logística foi utilizada a amostra de análise (Amostra A) e a eficiência do modelo foi testada através das amostras de validação (amostras B, C e D), ou seja, a comparação do comportamento previsto com o comportamento real. O

teste de validação mostra se o sistema de escoragem de crédito está funcionando, independentemente da amostra a partir da qual foi desenvolvido.

3.2 Tratamento de Dados

O tratamento dos dados consistiu em criar variáveis a partir dos dados coletados, identificar as variáveis estatisticamente significantes e desenvolver o modelo de escoragem do pedido por intermédio da técnica estatística chamada Regressão Logística.

Os dados observados na pesquisa de campo foram trabalhados pela técnica de regressão logística presente em *software* aplicativo de análise estatística, o SPSS.

A fase inicial da preparação dos dados consistiu em estruturar uma base agregando os valores numéricos referentes ao conjunto das possíveis variáveis explicativas disponíveis para utilização na construção do modelo.

O uso da técnica de regressão logística permitiu a criação de um modelo para verificar a relação entre a inadimplência (variável dependente) e as variáveis estudadas (independentes) e, desta forma, identificar as variáveis que impactam a inadimplência no crédito concedido para financiamentos imobiliários a pessoas físicas pertencentes à Autarquia Federal.

A estimação do modelo por intermédio da regressão logística foi realizada através do método *enter*, pois optou-se por incluir todas as variáveis preditoras simultaneamente, utilizando a Amostra A, composta por 200 clientes, sendo 100 inadimplentes e 100 adimplentes em fevereiro de 2010.

O uso da técnica de regressão logística permitiu a definição de um modelo de decisão - *credit scoring* - que estabelece a separação das pessoas em dois grupos (adimplentes e inadimplentes potenciais) e apresenta, satisfatoriamente, a probabilidade de que uma pessoa pertença a um grupo determinado, em função do comportamento observado das variáveis independentes.

A qualidade estatística das informações disponibilizadas pela ferramenta foi submetida às medidas de avaliação do modelo logístico.

A análise dos resultados obtidos pelo modelo de decisão - *credit scoring* - na medição da probabilidade de risco de crédito na Autarquia Federal foi realizada a partir do percentual de classificação correta dos tomadores de crédito como adimplentes ou inadimplentes.

Após o desenvolvimento do modelo, avaliou-se a sua capacidade preditiva pela construção de matrizes de classificação, como sugere Hair et al (2005), com intuito de verificar o grau de ajustamento da função construída.

A matriz de classificação consiste em uma tabela que compara a classificação realizada pelo modelo desenvolvido com a classificação original das observações da amostra. Essa matriz é elaborada através da análise de cada observação, visando conhecer se ela foi corretamente classificada pelo modelo. Os resultados dessa análise são os percentuais de acerto e erro de classificação do modelo.

Na pesquisa foi elaborada uma matriz de classificação a partir da Amostra B, a fim de realizar a validação do modelo, ou seja, verificação da possibilidade de generalização dos resultados encontrados na amostra de análise para a população referente ao mês de fevereiro de 2010.

Em acréscimo, foram elaboradas outras duas matrizes de classificação a partir das amostras C e D, com o intuito de verificar a validade do modelo ao longo do tempo, mais precisamente para os meses de abril (Amostra C) e junho (Amostra D).

3.2.1 Definição das Variáveis

A variável resposta (dependente) nos modelos de *credit scoring* é a qualidade de crédito (adimplência ou inadimplência) da operação de crédito.

Para classificar as observações de acordo com a qualidade de crédito, foram selecionadas variáveis explicativas ou independentes que pudessem influenciar a situação de adimplência dos clientes em suas operações de crédito. A identificação inicial das variáveis

explicativas foi baseada em estudos anteriores sobre o assunto e em informações disponíveis no sistema de informações da instituição.

As variáveis pré-selecionadas foram: renda do tomador do crédito, sua idade e endereço residencial, o valor do crédito concedido, o valor da prestação, o prazo de pagamento e a forma de pagamento, conforme elencado no Quadro 3 a seguir:

Quadro 3 – Discriminação das variáveis

Variável	Código da variável	Natureza/Formato
Renda	NF_1	Qualitativa / (dicotômica)
	NF_2	Qualitativa / (dicotômica)
	NF_3	Qualitativa / (dicotômica)
	NF_4	Qualitativa / (dicotômica)
Idade	IDADE_MENOR_25	Qualitativa / (dicotômica)
	IDADE_26_A_35	Qualitativa / (dicotômica)
	IDADE_36_A_55	Qualitativa / (dicotômica)
Endereço Residencial	REGIAO_S	Qualitativa / (dicotômica)
	REGIAO_N	Qualitativa / (dicotômica)
	REGIAO_NE	Qualitativa / (dicotômica)
	REGIAO_CO	Qualitativa / (dicotômica)
Valor do crédito	VALOR_FIN	Quantitativa
Valor da prestação	VALOR_PREST	Quantitativa
Prazo de pagamento	PRAZO_PAGTO	Quantitativa
Forma de Pagamento	FORMA_PAGTO	Qualitativa / (dicotômica)

Fonte: elaboração própria

As variáveis explicativas de natureza qualitativa foram inseridas na base de dados através de variáveis *dummy*, que consistem em variáveis construídas artificialmente para mensurar a presença ou ausência de algum atributo de natureza qualitativa. Elas assumem valor 0 ou 1, dependendo, respectivamente, da ausência ou presença de determinado atributo. Assim, foram inseridas N-1 *dummies* para representar as N categorias de cada variável qualitativa anteriormente explanada.

A variável renda foi dividida em 5 níveis funcionais, que refletem a progressão funcional e salarial dos tomadores de crédito, representados por 4 variáveis *dummy* – NF_1 a NF_4 - sendo NF_1 a faixa inicial e NF_4 a penúltima faixa de renda.

A variável idade foi dividida em 4 faixas, representadas por 3 variáveis *dummy* – 0 a 25 anos (IDADE_MENOR_25), 26 a 35 anos (IDADE_26_A_35) e 36 a 55 anos (IDADE_36_A_55).

A variável endereço residencial foi dividida em 5 regiões geográficas, representadas por 4 variáveis *dummy* – região Sul (REGIAO_S), região Norte (REGIAO_N), Região Nordeste (REGIAO_NE) e Região Centro-Oeste (REGIAO_CO).

4 APRESENTAÇÃO DOS RESULTADOS

No desenvolvimento do modelo de escoragem do pedido a partir do método de regressão logística foi utilizada a amostra A com 200 clientes, sendo 100 adimplentes e 100 inadimplentes em fevereiro de 2010.

Na condução da regressão logística com auxílio do software SPSS foi utilizada a opção *enter* como mecanismo de seleção das variáveis, pois optou-se por incluir todas as variáveis preditoras simultaneamente.

Para a situação de inadimplência foi atribuído o valor de 1 e para a situação de adimplência foi atribuído o valor 0.

Os valores apurados para as medidas de avaliação da capacidade preditiva do modelo logístico proposto foram: Log Likelihood (-2LL) – 168,918; Cox-Snell R^2 – 0,418; e Nagelkerke R^2 – 0,558. Os valores apurados para a medida de avaliação da capacidade preditiva do modelo logístico Qui-Quadrado referente ao modelo proposto foram: Qui-Quadrado – 108,341; df – 15; e Significância \approx 0,0000.

Na técnica de regressão logística, as previsões de pertinência ou classificação dos indivíduos em cada um dos grupos (adimplentes ou inadimplentes), são realizadas através da previsão direta da probabilidade do evento (inadimplência) acontecer.

Como visto, o valor resultante da soma do produto das variáveis selecionadas pelos seus respectivos pesos com a constante fornece uma pontuação (Z), que ao ser aplicada a função de distribuição logística fornece a probabilidade da ocorrência do evento (P).

Sendo assim, valores altos da probabilidade indicam que o modelo prevê a situação de indivíduos como inadimplentes. Por outro lado, baixos valores de probabilidade indicam um distanciamento da previsão de inadimplência, que será interpretada como uma aproximação da adimplência.

Tomando-se como referência que a probabilidade de 0,5 indica chances iguais de obtenção de um resultado quando se trata de duas escolhas, optou-se por atribuir o status de inadimplente às previsões com probabilidade superior a 0,5 e adimplentes, caso contrário.

As matrizes de classificação do modelo logístico aplicado às Amostras B, C e D apresentam médias de acurácia de 76%, 78% e 68%, respectivamente.

5 ANÁLISE DOS RESULTADOS

Os valores apurados para as medidas de avaliação da capacidade preditiva do modelo logístico Cox-Snell R^2 e Nagelkerke R^2 indicam que o modelo proposto é capaz de explicar cerca de 41,8% e 55,8%, respectivamente, das variações registradas na variável dependente.

Conforme Corrar (2007), a literatura especializada no assunto sugere que os pesquisadores usem essas estatísticas apenas como uma medida aproximada do poder preditivo do modelo, não atribuindo importância muito grande a cada um deles isoladamente.

Os valores apurados para as medidas de avaliação da capacidade preditiva do modelo logístico Qui-Quadrado indicam que é possível rejeitar a hipótese de que todos os parâmetros estimados são nulos.

No Teste de Hosmer e Lemeshow, seguindo uma distribuição Qui-quadrado, o cálculo resulta uma estatística de 8,3480 e um nível de significância de 0,4002. Isso indica que os valores preditos não são significativamente diferentes dos observados. Portanto, há um indício de que o modelo pode ser utilizado para estimar a probabilidade de um determinado cliente se tornar inadimplente em função das variáveis independentes.

A matriz de classificação do modelo proposto, contendo a comparação entre a classificação realizada pelo modelo desenvolvido e a classificação original das observações da amostra, indica uma média de acurácia de 80%. Nota-se, portanto, uma melhoria considerável em comparação à situação inicial de 50% de acerto, ou seja, na presunção de que todos os indivíduos seriam adimplentes.

A distribuição de frequência das probabilidades de ocorrência do evento de interesse em relação aos pontos de corte apresentada na Figura 2 sugere que o ponto de corte estabelecido para o modelo, ou seja 0,5, poderia ser ampliado para algo em torno de 0,70, com reduzido efeito sobre a capacidade preditiva do modelo. Além disso, é possível observar a pouca ocorrência de erros de previsão nos casos extremos (modelo indicando clientes com muito baixa ou muito alta probabilidade de inadimplência).

Até o momento, todos os testes sugerem que, de forma geral, o modelo pode ser utilizado para estimar a probabilidade de um cliente assumir a condição de inadimplente.

Entretanto, segundo Corrar (2007), não basta apenas verificar a avaliação do modelo como um todo: é necessário verificar se cada coeficiente pode ser utilizado como estimativa de probabilidades, por intermédio da estatística Wald.

A estatística de Wald apresentada no Quadro 4 sugere que não é possível afirmar que as variáveis são significativamente diferentes de 0 (zero), ou seja, nenhuma variável pode ser aproveitada na composição do modelo, a um nível de significância de 0,05.

A variável que se mostrou mais relevante foi o prazo de pagamento, com 0,12 de significância, o que não seria suficiente para negar a hipótese de não influência dessa variável

com os níveis de significância mais comumente utilizados nos testes de hipóteses (0,01, 0,05 e 0,10). No entanto, tal registro parece válido, já que este nível de significância revelou-se muito mais baixo do que o encontrado para as outras variáveis (todos maiores do que 0,50).

A matriz de classificação elaborada a partir da amostra de validação B para verificar a possibilidade de generalização dos resultados encontrados para a população referente ao mês de fevereiro de 2010 apresenta uma média de acurácia de 76%, inferior, portanto, a média de acurácia do modelo, 80%.

As duas matrizes de classificação elaboradas a partir das amostras de validação C e D para verificar a validade do modelo ao longo do tempo, mais precisamente para os meses de abril (Amostra C) e junho (Amostra D), apresentaram médias de acurácia de 78% e 68%, respectivamente.

Considerando a totalidade das amostras B, C e D, a média de acurácia do modelo seria de (457/618), ou seja, 74%.

Na regressão logística, cada coeficiente deve ser interpretado como estimativa do efeito que uma variável independente produz sobre a dependente quando as demais se mantêm inalteradas, sendo que o sinal do coeficiente é que vai determinar a direção da mudança. Os coeficientes negativos indicam que uma variação positiva em tais variáveis contribui para diminuir a probabilidade de um cliente se tornar inadimplente. No caso de sinal positivo, ao contrário, a variação positiva nessa variável concorre para aumentar a probabilidade do cliente se tornar inadimplente.

Para compreensão e interpretação dos resultados são apresentadas, a seguir, algumas considerações acerca das variáveis, a partir das estimativas dos coeficientes apresentados no Quadro 4. Cabe ressaltar, no entanto, que tais refletem apenas tendências de impacto dos fatores analisados sobre a probabilidade de inadimplência, não sendo conclusivas, uma vez que nenhum dos fatores apresentou significância estatística.

- renda - os coeficientes apresentados pela regressão logística para a variável indicam que quanto menor o nível funcional do tomador de crédito, maior o efeito positivo sobre a inadimplência;
- idade - os coeficientes estimados para a variável indicam um efeito negativo sobre a inadimplência até os 35 anos de idade, quando comparados aos idosos e um efeito positivo em idades intermediárias quando comparadas às mais avançadas;
- endereço residencial - os coeficientes apresentados para a variável indicam um efeito negativo – em comparação a moradores da região Sudeste – sobre a inadimplência o fato de o indivíduo possuir residência declarada nas regiões Sul e Nordeste, ocorrendo de modo inverso nas demais regiões;
- valor do crédito - o coeficiente nulo - até décimo de milésimo (10^{-4}) - apresentado pela regressão logística indica que a variável não interfere, estatisticamente, na situação de adimplência ou inadimplência do indivíduo;
- valor da prestação - o coeficiente estimado indica que uma variação positiva nessa variável contribui para diminuir a probabilidade de um cliente se tornar inadimplente;
- prazo de pagamento - o coeficiente estimado indica que a variação positiva nessa variável concorre para aumentar a probabilidade do cliente se tornar inadimplente;
- forma de pagamento - o coeficiente apresentado para a variável indica que sua existência (pagamento por boleto) tem efeito positivo sobre a inadimplência.

6 CONSIDERAÇÕES FINAIS

O aumento das transações no mercado imobiliário, dito aquecido, pelo aumento da disponibilidade de crédito imobiliário e a redução de taxas de juros praticadas no mercado financeiro incentivam a procura por crédito para a realização de reformas e ampliações em unidades residenciais.

A concessão de crédito compreende uma diversidade de situações e cada uma delas apresenta peculiaridades em termos de quais características dos tomadores de crédito e do

crédito em si são capazes de influenciar na (in)adimplência.

O presente estudo teve como objetivo identificar as variáveis que impactam a inadimplência no crédito concedido para projetos imobiliários, ou seja, para situações em que o tomador do crédito pertencente à Autarquia Federal estudada na pesquisa deseja crédito para realizar reforma ou ampliação de unidade residencial própria.

Para o alcance do objetivo, foi necessário, inicialmente, identificar, na literatura, as variáveis referentes aos financiamentos pessoais para projetos imobiliários com potencial para impactar a inadimplência.

Em seguida, foi construído um modelo de regressão logística para verificar o impacto das variáveis renda do tomador do crédito, sua idade e endereço residencial, o valor do crédito concedido, o valor da prestação, o prazo e a forma de pagamento na inadimplência. A variável “forma de pagamento”, não mencionada na literatura, mas presente na situação dos potenciais tomadores de crédito, foi inserida no modelo e espelha a possibilidade de pagamento consignado ou por intermédio de boleto de pagamento.

Os resultados dos testes para avaliação da capacidade preditiva do modelo logístico proposto, quais sejam, Qui-Quadrado, Cox-Snell R^2 , Nagelkerke R^2 , Teste de Hosmer e Lemeshow e matriz de classificação indicavam que o modelo seria uma opção válida para uso na solução do problema. Entretanto, a estatística de Wald impediu a aprovação estatística do modelo.

As matrizes de classificação elaboradas a partir das amostras de validação B, C e D apresentam médias de acurácia inferiores à média de acurácia do modelo. O fato da amostra referente ao mês de junho, a de maior distância temporal em relação a fevereiro, período considerado para a elaboração do modelo, possuir a pior média de acurácia alerta sobre a possibilidade da rápida degradação da capacidade preditiva do modelo ao longo do tempo.

Na medida em que se aproximou o alcance do objetivo – identificar as variáveis que impactam a inadimplência no crédito concedido para projetos imobiliários – surgiram algumas surpresas no confronto entre os resultados empíricos encontrados e a expectativa teórica, provavelmente pela inconsistência apontada pela estatística Wald.

O impacto, na inadimplência, das variáveis pré-selecionadas, renda do tomador do crédito, sua idade e endereço residencial, o valor da prestação e o prazo de pagamento, conforme sugerido por Securato (2007) e Santos (2009), e da forma de pagamento, não citada na literatura, foi confirmado no modelo empírico. Em sentido contrário, a variável “valor do crédito concedido” não impactou conforme o esperado.

O resultado de que a forma de pagamento por boleto e o prazo de pagamento aumentam a probabilidade da ocorrência de inadimplência, bem como de que as características de idade e endereço residencial alteram a probabilidade de inadimplência, como apresentado, também não causou surpresa, uma vez que se alinha com a expectativa teórica.

Entretanto, a interpretação dos resultados para a renda, valor do crédito e valor da prestação é dificultada na medida em que os coeficientes sinalizam em sentido diverso ao que seria suposto em uma avaliação teórica preliminar.

Ao final, a elaboração de um modelo de *credit scoring* com média de acurácia de 74% ao longo do período de observação, permitiria o alcance do objetivo de pesquisa, ou seja, identificar as variáveis que impactam a inadimplência no crédito concedido para projetos imobiliários, não fosse a negativa de relevância estatística apresentada pela estatística Wald.

Com relação às hipóteses formuladas, podemos concluir:

- a) pela rejeição da hipótese I, ou seja, as variáveis consideradas isoladamente não impactam a inadimplência;
- b) pela rejeição da hipótese II, ou seja, consideradas as variáveis em conjunto, as variáveis renda do tomador do crédito, sua idade e endereço residencial, o valor do crédito concedido, o valor da prestação, o prazo e a forma de pagamento não permitem estabelecer uma ferramenta de *credit scoring* que estime, satisfatoriamente, a probabilidade de ocorrência de inadimplência futura no crédito concedido para financiamentos imobiliários a pessoas

físicas pertencentes à Autarquia Federal.

Não obstante o modelo não estar aprovado estatisticamente, seu uso ainda é capaz de aumentar a acurácia da previsão em algo em torno de 25 pontos percentuais. Desta forma e em vista de todos os testes gerais terem sugerido que o modelo pode ser utilizado para estimar a probabilidade de um cliente assumir a condição de inadimplente, não parece sábio considerar o modelo totalmente dispensável. Talvez até ele possa ser usado, mas com ressalvas e cuidado. Certamente, ele não pode ser o único instrumento para determinar a concessão de crédito ou não ao tomador, mas provavelmente pode servir de apoio para isso.

Por causa da reprovação estatística do modelo, parece razoável supor que mais variáveis precisam ser indagadas no momento da concessão de crédito, como moradia (se própria ou alugada) e número de dependentes (SANTOS, 2009); algumas têm que ser mais bem coletadas (a renda, por exemplo, que deve ser tratada de forma quantitativa, ao invés de categórica, como no caso desse modelo, por força da maneira como a variável é coletada no momento da concessão do crédito) e que precisa haver um cuidado maior com a qualidade das informações.

REFERÊNCIAS

CAOQUETTE, J. B., ALTMAN, E. I., NARAYANAN, P. **Gestão do Risco de Crédito: O Grande Desafio dos Mercados Financeiros Globais**. 2. ed. Rio de Janeiro: Qualitymark, SERASA, 2009.

CORRAR, L. J., PAULO, E., DIAS FILHO, J. M. (Coord). **Análise Multivariada: para os Cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, 2007.

CROUHY, M., GALAI, D., MARK, R. **Fundamentos da Gestão de Risco**. Rio de Janeiro: Qualitymark; São Paulo: SERASA, 2007.

GITMAN, L. J. **Princípios da Administração Financeira**. 10. ed. São Paulo: Addison Wesley, 2004.

HAIR, J. F., ANDERSON, R. E., TATHAM, R. L. e BLACK, W. C. **Análise Multivariada de Dados**. 5. ed. Porto Alegre: Bookman, 2005.

HOUAISS. **Dicionário HOUAISS da Língua Portuguesa**. Disponível em <<http://houaiss.uol.com.br>>. Acesso em: 01 nov. 2009.

MACHADO, C. A. B. Avaliação de crédito: um estudo da carteira de inadimplentes do Banco Finasa. In: XXI Encontro Anual da Associação Nacional dos Cursos de Graduação em Administração (ENANGRAD), 2010, Brasília. **CD-Anais da XXI ENANGRAD...** Brasília: ENANGRAD, 2010. 18 p.

MATIAS, A. B., DAUBERMANN, E. C., RICI, E. T. G. Inadimplência não é Insolvência: Um Estudo Conceitual e Empírico com uma Abordagem Macrofinanceira. **Revista FACEF Pesquisa / Faculdades de Ciências Econômicas, Administrativas e Contábeis de Franca**, v. 12, n. 1, p. 52-62, 2009.

SANTOS, J. O. **Análise de Crédito: Empresas, Pessoas Físicas, Agronegócios e Pecuária**. 3. ed. São Paulo: Atlas, 2009.

SCHRICKEL, W. K. **Análise de Crédito: Concessão e Gerência de Empréstimos**. 5. ed. São Paulo: Atlas, 2000.

SECURATO, J. R. (Org.). **Crédito: Análise e Avaliação do Risco - Pessoas Físicas e Jurídicas**. São Paulo: Saint Paul Editora, 2007.