

Desenvolvimento de um modelo de *credit scoring* para uma cooperativa de crédito brasileira

PRISCILA ALMEIDA DE MAGALHAES
UNIVERSIDADE FEDERAL DE MINAS GERAIS

POURI DO CARMO MARIO
UNIVERSIDADE FEDERAL DE MINAS GERAIS

RESUMO

Os modelos de *credit scoring* são modelos empregados comumente por instituições financeiras e de crédito para a mensuração e previsão do risco de crédito, tendo uso consolidado no processo de concessão do mesmo. Este trabalho avaliou a possibilidade de aplicação de modelos de *credit Scoring* em uma cooperativa de crédito de Minas Gerais. Para isso, foram coletados dados relativos a uma amostra de cooperados, os quais foram utilizados para desenvolver o modelo, com o uso de regressão logística (*logit model*). As principais dificuldades encontradas na realização deste trabalho decorreram dos dados do sistema computacional utilizado pela instituição, que é atualmente a única fonte das operações de crédito da instituição. Esse sistema não engloba alguns dados relevantes pela literatura consultada em modelos de *credit scoring* e consequentemente estes não puderam ser considerados neste estudo. Foi utilizada uma amostra aleatória para o grupo de solventes e uma amostra intencional (acessibilidade) para o grupo de insolventes. Após o processo de análise das variáveis, restaram dois modelos, que obtiveram um percentual de acerto de 89,2% e 85,8%, além de alta significância estatística para as variáveis inseridas. As principais variáveis dos modelos são: estado civil, tempo no último emprego, gênero, instituição empregadora e existência de aval, todas se comportando conforme a hipótese prevista.

Palavras-chave: *credit scoring*, risco de crédito, cooperativas de crédito.

1 INTRODUÇÃO

As cooperativas de crédito surgiram da necessidade de se obter melhores condições nos serviços de intermediação financeira, com menores taxas de juros e de serviços e têm grande relevância no cenário mundial, como é possível notar pelos dados a seguir.

Na Alemanha, as cooperativas de crédito possuem cerca de quinze milhões de associados e, segundo Meinen (2002), respondem por cerca de 20% de todo o movimento financeiro bancário do país. Na Holanda, o banco cooperativa Rabobank atende a mais de 90% das demandas financeiras rurais. Nos EUA, há mais de doze mil unidades de atendimento cooperativo apenas no sistema CUNA (Credit Union Nacional Association), que reúne cooperativas de crédito mútuo do tipo Desjardins. Além dessas, os bancos cooperativos agrícolas, por sua parte, respondem por mais de um terço dos financiamentos agropecuários norte-americanos. Ainda segundo Meinen, estima-se que 25% dos norte-americanos sejam associados a uma cooperativa de crédito. Schar Dong (2002), citando dados fornecidos pela Agência de Estatística da União Européia, de 2000, menciona que 46% do total das instituições de crédito da Europa eram cooperativas, participando com cerca de 15% da intermediação financeira. (PINHEIRO, 2008, P.7)

De forma ampla a Associação Internacional de Bancos cooperativos (International Co-operative Banking Association) informa que:

As cooperativas de créditos nas suas diversas formas servem a mais de 621 milhões de pessoas apenas nas nações integrantes do G-20. Essas fornecem 3,6 trilhões de

dólares em empréstimos, possuem depósitos no valor de 4,4 trilhões de dólares e têm 7,6 trilhões de dólares em ativos.¹

No Brasil elas ainda não possuem a expressividade que têm nos países desenvolvidos, mas vêm crescendo nos últimos tempos, como se pode notar na tabela 1.

Tabela 1 Participação das cooperativas de crédito nos principais agregados financeiros do segmento bancário brasileiro.

Posição (dez)	Quantidade	% Patrimônio Líquido	% Ativos Totais	% Depósitos	% Operações de Crédito
1997	1 120	1,4	0,3	0,3	0,7
1998	1 198	1,4	0,4	0,5	0,8
1999	1 253	1,6	0,6	0,6	1,0
2000	1 311	1,7	0,7	0,8	1,1
2001	1 379	1,8	0,8	1,0	1,4
2002	1 430	2,0	0,9	1,1	1,5
2003	1 454	2,0	1,1	1,4	1,8
2004	1 436	2,4	1,3	1,4	2,0
2005	1 439	2,6	1,3	1,3	2,1
2006	1.450	2,4	1,4	1,5	2,0
2007	1.462	2,3	1,3	1,3	2,1

Fonte: www.cooperativismodecredito.com.br

Diferentemente das instituições financeiras, as cooperativas são constituídas como sociedades de pessoas, com forma e natureza próprias, de natureza civil, não sujeitas à falência, sem fins lucrativos e com objetivo de propiciar crédito e prestar serviços aos associados. Ainda assim, pela natureza dos serviços prestados de intermediação financeira, elas estão sujeitas aos mesmos riscos inerentes às demais instituições desse segmento. Dentre esses riscos podem ser citados o risco de crédito, o risco de mercado, o risco país e o risco operacional.

A necessidade de controle e gerenciamento do risco de forma eficaz e com baixo custo fez com que as instituições financeiras passassem a primar pelo aperfeiçoamento das técnicas utilizadas para essa finalidade. Entre os instrumentos utilizados estão os *credit scoring* que são modelos estatísticos usados como ferramenta na previsão da solvência de pessoas físicas e jurídicas.

Tendo em vista que os modelos quantitativos já têm uso consolidado na análise de crédito de instituições financeiras tradicionais, considerou-se a possibilidade do desenvolvimento de um modelo de *credit scoring* para uma cooperativa de crédito de Minas Gerais a partir de seu próprio banco de dados. Sendo que a pesquisa parte do seguinte questionamento: quais são as principais variáveis a serem usadas em um modelo de *credit scoring* para prever o comportamento de crédito de futuros e atuais cooperados durante o desenrolar de um empréstimo?

O objetivo geral deste trabalho foi desenvolver um modelo de *credit scoring* baseado nos dados de uma cooperativa de crédito de Minas Gerais e seus objetivos específicos foram: (1) Propor um modelo de *credit scoring* a partir dos dados cadastrais de clientes; (2) Verificar a consistência do modelo proposto e as principais variáveis utilizadas, bem como analisar os relacionamentos indicados com a realidade da cooperativa.

¹ Tradução livre do inglês: *Financial cooperatives in their diverse forms serve over 621 million people in the G-20 nations alone. They "provide US\$3.6 trillion in loans, hold US\$4.4. trillion in savings and have US\$ 7.6 trillion in total assets*

É possível perceber pelos dados apresentados que o grau de representatividade do cooperativismo de crédito brasileiro ainda está muito aquém daquele obtido por outros países como França, Holanda, Estados Unidos. Porém, eles mostram que essa forma de intermediação financeira pode ser muito bem sucedida e tem grande potencial de crescimento no Brasil.

Entretanto, é necessário que o crescimento neste mercado extremamente competitivo seja feito de forma sustentável e isso inclui uma gestão apropriada dos riscos. Deve-se considerar também que a Lei 5.764/74 determina que assim como as sobras serão divididas entre os associados o mesmo ocorrerá na existência de perdas, o que deveria gerar um comprometimento ainda maior na sua gestão.

Outro aspecto importante é que as cooperativas são um elemento importante no desenvolvimento de alguns setores da economia que não encontram facilidades na concessão de crédito em outras instituições como os micro-empresários e a população de baixa renda.

Além disso, esse trabalho se justifica pelas possíveis contribuições que seus resultados podem fornecer ao processo de concessão e análise do crédito na cooperativa.

3. *CREDIT SCORING*

Métodos tradicionais de análise usam julgamento humano para definir o risco da concessão do crédito, baseado na experiência de decisões anteriores. Entretanto, pressões econômicas resultantes do crescimento da demanda pelo crédito, aliada à maior competição no setor financeiro e desenvolvimento de tecnologia computadorizada levou ao desenvolvimento de modelos estatísticos para auxiliar o processo de concessão de crédito. A esses modelos foi dado o nome de *credit scoring*. A função do modelo de *credit scoring* é medir o risco, sendo, portanto uma ferramenta que auxilia na decisão de fornecer ou não o crédito ao solicitante.

Os modelos tradicionais de *credit scoring* atribuem pesos estatisticamente predeterminados a alguns atributos dos solicitantes para gerar um escore de crédito. Se esse escore é favorável, quando comparado a um valor de corte, então a solicitação é aprovada. (CAOINETTE, ALTMAN, NARAYANAN 1999, p.182)

Credit scoring é um procedimento estatístico que possibilita a previsão do comportamento futuro baseado em informações sobre o desempenho passado. “É essencialmente uma forma de reconhecer grupos diferentes em uma população quando não é possível ver as características que os separam, mas apenas características relativas àquelas.” (THOMAS, 2000, 151)² Uma das primeiras necessidades na construção de um modelo desse tipo é escolher as variáveis explicativas advindas principalmente dos cadastros preenchidos pelos proponentes e das centrais de informações de crédito (SERASA, SPC). Os dados que respondem a essas variáveis serão colhidos de uma amostra dos proponentes para os quais o crédito já foi concedido.³ Posteriormente os clientes componentes da amostra serão classificados em riscos bons e ruins o que deve ser definido de acordo com um ponto de corte definido pela instituição financeira segundo suas necessidades de mercado.

Segundo Hand e Henley (1997) existem três formas mais comuns de se escolher quais características serão utilizadas: baseando-se no conhecimento de especialistas, os quais por

² Tradução livre do inglês

³ É importante lembrar que apenas os proponentes aceitos para o empréstimo requisitado podem ser acompanhados durante o pagamento do mesmo e classificados como bons ou maus clientes de acordo com a definição adotada. Isso implica em uma amostra tendenciosa já que não é possível saber o que aconteceu com os proponentes recusados. Tentativas de lidar com esse viés usando as informações que se tem do proponente, mas não sua classe são chamadas de *rejected interference*. (HAND e HENLEY 1997, THOMAS 2000)

trabalharem com as informações diariamente, tem uma boa visão de quais dados serão mais úteis; usando o procedimento chamado *stepwise*, o qual consiste na adição de novas características ou grupo delas a cada passo⁴, de forma a se achar o conjunto que obtém melhor resultado; a terceira forma é analisar cada característica individualmente usando uma função para determinar o poder discriminante da característica, sendo que as que tivessem baixo poder discriminante não seriam consideradas para inclusão no modelo.⁵ Vale lembrar que normalmente são usados os três métodos em conjunto.

As variáveis (frequentemente chamadas de características no meio bancário) das quais será obtida a predição são normalmente uma mistura de variáveis contínuas (ex. idade, tempo no atual endereço, receita mensal) e categóricas (proprietário do imóvel ou locatário, tipo de conta bancário, tipo de ocupação). As contínuas podem ser agrupadas em intervalos, tornando-as variáveis categóricas. Apesar disso não ser necessário para vários procedimentos estatísticos, o agrupamento oferece a vantagem de colocar todas as variáveis sobre a mesma forma, além de vantagens interpretativas (por exemplo, um cliente que pertence ao grupo X das idades pode ser caracterizado como mais arriscado do que um que pertence ao grupo Y). Para que as variáveis sejam categorizadas é preciso anteriormente escolher o número de categorias e as posições dos pontos de corte. (HAND,2001, p.144)⁶

Uma questão que se origina da existência de variáveis categóricas é a de como transformá-las em numéricas. Hand e Henley (1997, p.527) colocam que apesar de o alcance das técnicas estatísticas para lidar com dados categóricos multivariados terem ampliado dramaticamente nos últimos 15 anos, quase todos os sistemas *credit scoring* comerciais usam variáveis auxiliares (*dummy variables*). Ressaltam ainda que, entretanto, as técnicas alternativas para codificar as variáveis categóricas em formas numéricas passivas de serem usadas nos modelos de dados contínuos estão se tornando mais comuns.⁷

Existem vários métodos estatísticos usados para o desenvolvimento de *credit scoring* sendo que, nesta monografia, serão sucintamente explicadas a análise discriminante linear, a regressão logística e as redes neurais. Hand e Henley (1997) ainda acrescentam as árvores de decisão e Saunders e Altman (1997) o modelo probit como métodos comumente utilizados no desenvolvimento desses modelos.

Entretanto, independentemente do método estatístico utilizado, o objetivo final é prever quem irá apresentar problemas de insolvência e não oferecer explicações sobre as razões desta ou responder hipóteses sobre a relação entre insolvência e outras variáveis econômicas ou sociais. (THOMAS, 2000)

Uma forma de chegar a uma função discriminante linear é definir a variável Y igual a 1 se o proponente for bom, e igual a 0 se o proponente for ruim. A equação de regressão linear da variável Y resultante das respostas para características (variáveis X) fornece um conjunto de pesos para as variáveis preditivas que minimizem a soma dos quadrados entre o valor

⁴ Esse método apresenta falhas caso sejam usadas variáveis auxiliares (*dummy variables*) para as características categóricas.

⁵ O autor define que resultados acima de 0,1 na função $\sum_{j=1}^{n_{atrib}} (p_{ij} - q_{ij}) w_{ij}$ onde p_{ij} é o numero de bons cliente do atributo j da característica i dividido pelo total de bons clientes da amostra, q_{ij} é o numero de cliente ruins do atributo j da característica i dividido pelo total de cliente ruins da amostra e $w_{ij} = \ln(p_{ij}/q_{ij})$

⁶ Tradução livre do inglês

⁷ A formula apresentada pelos autores para realizar essa transformação é $w_{ij} = \ln(p_{ij}/q_{ij})$, sendo que W é o valor numérico para determinado atributo da característica (*weight of evidence*), ou seja, a resposta da variável categórica.

calculado e o valor real de Y(0 ou 1). A partir desse ponto, novos proponentes terão seus Y estimados comparados ao ponto de corte.

Uma fraqueza da utilização da regressão linear na construção de um modelo desse tipo é tal que a soma resultante da aplicação do método dos mínimos quadrados não fica restringida entre os valores 0 e 1, o que seria bastante atrativo para as instituições financeiras já que possibilitaria a interpretação dos resultados como probabilidades. (SAUNDERS E ALTMAN, 1997).

A regressão logística, por sua vez, usa um grupo de variáveis preditoras para definir a probabilidade de insolvência do cliente, assumindo que esta probabilidade é logisticamente distribuída e a probabilidade acumulada assume a forma de uma função logística a qual por definição fica restrita entre 0 e 1. (SAUNDERS E ALTMAN, 1997)

Redes neurais são sistemas de inteligência artificial desenvolvidos para simular os processos humanos de raciocínio e aprendizado. Os algoritmos de redes neurais são um conjunto de dados (neste caso, variáveis usadas na solicitação de crédito) que são transformadas matematicamente através de uma função de transferência para gerar um resultado (no caso, se o crédito de um solicitante seguirá curso normal ou será cancelado).

Essencialmente, análise de redes neurais é similar a análise discriminante não linear, na qual se abandona a idéia de que as variáveis usadas na função predição de insolvência são linearmente e independentemente relacionadas. Mais especificamente, modelos de redes neurais aplicados ao risco de crédito exploram correlações “escondidas” entre as variáveis preditoras, as quais são então adicionadas ao modelo como variáveis explicativas.(ALTMAN e SAUNDERS, 1998, p.1726)⁸

Uma das críticas a este modelo é a de que:

Os reguladores vêm a opinião qualitativa de um analista de crédito como uma base aceitável para a decisão de conferir ou não o crédito, mas ainda assim esperam que decisões de crédito baseadas em dados de máquinas venham a satisfazer o padrão de tratabilidade, ou seja, a capacidade de explicar quais variáveis produziram uma decisão de crédito adversa. Por isto, redes neurais são mais úteis na retaguarda da gestão de crédito.(CAOINETTE, ALTMAN E NARAYANAN, 1999, p.188)

Independentemente do método estatístico utilizado, a classificação perfeita dos bons e maus proponentes seria impossível, até porque, às vezes, as mesmas respostas para as características são dadas por proponentes bons e ruins. Entretanto, procura-se obter uma regra que classifique erroneamente tão poucos proponentes quanto possível e ainda satisfaça um razoável requerimento de continuidade (THOMAS, 2000)

Caouette, Altman, Narayanan (1999) destacam alguns dos pontos positivos dos sistemas de *credit scoring*:

- São objetivos e consistentes, características desejáveis para qualquer instituição, e especialmente para aquelas que não possuem uma forte cultura de crédito;
- Se desenvolvidos apropriadamente, eles podem eliminar práticas discriminatórias nos empréstimos;
- Eles tendem a ser relativamente inexpressivos, bastante simples e de fácil interpretação;

⁸ Tradução livre do inglês: Essentially, neural network analysis is similar to non-linear discriminant analysis, in that it drops the assumption that variables entering into the bankruptcy prediction function are linearly and independently related. Specially, neural network models of credit risk explore potentially “hidden” correlations among the predictive variables which are then entered as additional explanatory variables in the non-linear bankruptcy prediction function

- A instalação de tais modelos é relativamente fácil;
- As metodologias usadas para construir esses modelos são comuns e bem entendidas, assim como as abordagens usadas para estatísticas;
- Uma instituição é capaz de proporcionar melhor serviço ao consumidor pela sua habilidade de aprovar ou negar um pedido de empréstimo rapidamente.

No entanto, esses modelos apresentam alguns defeitos, na maior parte dos casos:

- Automatizam as práticas de crédito prevaletentes dos bancos. Em outras palavras, eles pouco ajudam a eliminar as distorções históricas de seleção de uma instituição;
- Se as variáveis não satisfazem presunções subjacentes, como a distribuição normal multivariada, a validade estatística dos modelos pode ser questionável;
- É muito comum os testes estatísticos usados para ajustar esses modelos serem fracos e confundirem o usuário, superestimando sua eficácia.

É importante mencionar que, uma vez que a instituição financeira (IF) opte por usar um modelo de *credit scoring*, ela terá que decidir por um que seja genérico ou um customizado, sendo que este último apesar de possuir custo mais elevado, tem, normalmente, performance superior. Isso acontece porque, a população sobre a qual o modelo genérico é construído pode diferir da população real da instituição. Thomas (2000, p. 161) menciona experimento feito por Overstreet, Bradley, Kemp em 1992 no qual um *scorecard* genérico foi montado para diversas cooperativas de crédito, as quais normalmente possuem poucos clientes. Nesse caso foi verificado que o sistema desenvolvido não era competitivo com os customizados, entretanto era melhor do que não usar nenhum. Caouette, Altman, Narayanan (1999, p.182) declaram que normalmente sistemas genéricos são usados pelas IF's como um substituto até ter amostragem suficiente para gerar um sistema baseado em sua própria clientela.

5 METODOLOGIA

A pesquisa tem caráter empírico-descritivo e propõe-se a coletar e analisar dados reais e, com base nessa análise, propor um modelo de *credit scoring* estimado por meio de regressão logística. Esse modelo de aprovação de crédito tem como objetivo principal servir de auxílio na avaliação de decisão do analista sobre a concessão de crédito. Os resultados encontrados foram analisados com base nos procedimentos utilizados e nas limitações encontradas durante a pesquisa.

5.1 O Modelo da Regressão Logística

A regressão logística é comumente utilizada para análise de dados com resposta binária ou dicotômica e consiste em relacionar, por meio de um modelo, a variável resposta (variável dependente binária) com fatores que influenciam ou não a probabilidade de ocorrência de determinado evento (variáveis independentes).

Supondo um evento dependente em que a variável Y é uma variável binária que assume valores 0 ou 1, e variáveis independentes X_1, X_2, \dots, X_k , a função de distribuição logística é dada por:

$$P_i = \Pr[Y_i=1/X=X_i] = 1/1+e^{-z}$$

Onde:

- P_i representa a probabilidade de um evento ocorrer e
- $Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_iX_i$

Esta função possui a propriedade de que quando $Z_i \rightarrow -\infty \Rightarrow P_i \rightarrow 0$ e $Z \rightarrow +\infty \Rightarrow P \rightarrow 1$. Além disso, essa passagem não se dá de forma linear, mas sim de acordo com a forma não linear da função logística.

Analisando-se o significado da função de distribuição logística no contexto de risco de crédito, tem-se que a variável dependente consiste na situação de solvência do cliente que assumirá valores 0 ou 1, a depender de os dados procederem de um cliente adimplente ou inadimplente. As variáveis independentes representam os fatores que se supõe influenciarem a insolvência como, por exemplo, dados pessoais e financeiros. A probabilidade de insolvência de uma determinada pessoa é dada por P_i , que é a probabilidade condicional de Y assumir o valor 1. Os coeficientes estimados por B representam medidas das variações na proporção das probabilidades.

5.2. Amostra e coleta de dados

Os dados utilizados nessa pesquisa foram cedidos por uma cooperativa de crédito de Minas Gerais, a qual possui mais de 6.000 cooperados ativos e uma carteira de crédito de aproximadamente R\$ 22.000.000,00.

A partir da lista de todos os cooperados fornecida pela instituição foi feita uma seleção aleatória de 150 cooperados que possuíssem empréstimos em andamento considerados solventes, dos quais foram usados os dados de 97 no modelo. Os cooperados insolventes da amostra foram selecionados por funcionário da cooperativa num total de 26, por conta da acessibilidade, sendo que foram usados os dados de 23 no modelo.

As informações dos cooperados selecionados foram preenchidas com base em uma planilha elaborada pelo autor. As informações requisitadas relativas às características pessoais e às condições econômico-financeiras dos cooperados foram definidas baseadas na literatura revisada. Algumas limitações foram encontradas nessa fase da pesquisa, Dentre elas:

- Os dados requisitados tiveram que ser limitados em função do sistema integrado usado pela cooperativa, o qual não possuía em sua base todas as informações solicitadas da primeira vez. Exemplo disso seriam as informações sobre restrição de crédito obtidas por meio do SERASA ou outro banco de dados similar.

- A pesquisa foi realizada com informações atualizadas a menos de um ano dos cooperados, o que pode interferir no resultado, uma vez que não possui um lapso de tempo entre o cadastro e a previsão a ser obtida pelo modelo de *credit scoring*.

- A intenção inicial da pesquisa era utilizar a totalidade dos cooperados, entretanto este número teve que ser limitado por causa do tempo necessário para realizar a coleta de dados que na maior parte dependia da disponibilidade dos funcionários da cooperativa, e devido à necessidade de buscar algumas informações individualmente (uma a uma), o que também tornou o processo bastante moroso.

5.3 Definição das variáveis

A variável resposta (dependente) nos modelos de *Credit Scoring* é a definição do crédito como solventes e ou insolventes. Nessa pesquisa foram determinados insolventes os clientes com 200 dias ou mais de atraso em uma ou mais parcelas. Cooperados com atrasos inferiores a 200 dias (aproximadamente 7 meses) foram considerados solventes. Também, para ser considerado adimplente o cooperado deveria ter efetuado o pagamento de pelo menos metade dos empréstimos.

Para classificar as observações de acordo com a qualidade de crédito, foram definidas 13 variáveis explicativas ou independentes a partir das características dos cooperados. Supõe-se que tais variáveis possam se relacionar à situação de solvência ou insolvência nas operações de empréstimos. O conjunto inicial de variáveis é mostrado a seguir:

- 1) Idade do cliente
- 2) Tempo no último emprego

- 3) Rendimento por dependentes
- 4) Estado civil
- 5) Alavancagem em relação a renda
- 6) Primeiro empréstimo na cooperativa
- 7) Instituição empregadora
- 8) Modalidade do empréstimo
- 9) Gênero do cliente
- 10) Comprometimento da renda
- 11) Atrasos em empréstimos anteriores
- 12) Possui avalista
- 13) Número de parcelas do empréstimo

5.4 Preparação dos dados, estimação e avaliação dos modelos

Após a coleta dos dados, as características foram transformadas em variáveis explicativas passíveis de serem usadas no modelo a ser proposto. Isso foi feito por meio de várias formas, dependendo da característica a ser utilizada. No quadro 2, é descrita cada variável e o seu respectivo tratamento e relacionamento.

NOME DA VARIÁVEL	DESCRIÇÃO DA VARIÁVEL	TRATAMENTO	EXPECTATIVA DO COMPORTAMENTO
A1	Idade do cooperado	(data atual)-(data de nascimento)	Espera-se uma relação inversa entre a idade e a probabilidade de insolvência
A2	Tempo no último emprego	(data do início do contrato) – (data de admissão)	Espera-se uma relação inversa entre o tempo de emprego e a probabilidade de insolvência
A3	Rendimento por dependentes	renda total/ número de dependentes	Espera-se uma relação inversa entre a renda por dependente e a probabilidade de insolvência.
A4	Estado civil	Solteiro = 1 Demais= 0	Espera-se que os cooperados solteiros tenham uma probabilidade maior de insolvência
A5	Alavancagem em relação a renda	valor total do empréstimo/ renda bruta	Espera-se uma relação direta entre a alavancagem e a probabilidade de insolvência
A6	Primeiro empréstimo na cooperativa	Sim = 1 Não = 0	Espera-se que os cooperados mais antigos possuam menor risco do que os mais novos
A7	Instituição empregadora	A = 0 B = 1 C = 2 D = 3 E = 4 F = 5 G = 6 H = 7 I = 8 J = 9 L = 10	A cooperativa participante desta pesquisa possui cooperados de várias instituições diferentes. Esta variável procura verificar se há alguma relação entre a instituição a qual o cooperado é ligado e o seu risco. Inicialmente não era possível ter uma previsão de como este indicador fosse se comportar.
A8	Modalidade do empréstimo	Consignado = 0 Outros = 1	Espera-se que os empréstimos consignados possuam uma probabilidade de insolvência menor
A9	Gênero do cliente	Homem = 0 Mulher = 1	Sem hipótese prévia de relacionamento entre variáveis
A10	Comprometimento da renda	valor consolidado das parcelas / renda bruta	Espera-se uma relação direta entre o comprometimento da renda e a probabilidade de insolvência.
A11	Atrasos em empréstimos anteriores	Tem atrasos anteriores=1 Não possui atrasos anteriores=0	Espera-se que cooperados com histórico de atrasos anteriores sejam mais propensos a serem insolventes
A12	Possui avalista	Possui avalista = 1 Não possui avalista = 0	Sem hipótese prévia de relacionamento entre variáveis
A13	Número de parcelas do empréstimo	Sem tratamento	Espera-se uma relação direta entre o número de parcelas e a probabilidade de insolvência.

Quadro 1 - Descrição e tratamento das variáveis

Fonte: Elaborado pelo autor

No caso da conversão das variáveis independentes categóricas foi utilizado um segundo método de tratamento baseado em Hand e Henley (1997), onde:

$$w_{ij} = \ln(p_{ij}/q_{ij})$$

- W(Weight of evidence) é o valor numérico para determinada evidência (resposta da variável categórica);
- p_{ij} é o número de bons clientes do atributo j da característica i dividido pelo total de bons clientes da amostra;
- q_{ij} é o número de clientes ruins do atributo j da característica i dividido pelo total de cliente ruins da amostra.

Os resultados dessa conversão e mostrada no quadro 3.00

NOME DA VARIÁVEL	DESCRIÇÃO DA VARIÁVEL	TRATAMENTO	EXPECTATIVA DO COMPORTAMENTO
B4	Estado civil	Solteiro = -0,785 Demais=0,535	Espera-se que os cooperados solteiros tenham uma probabilidade maior de insolvência
B6	Primeiro empréstimo na cooperativa	Sim = -0,954 Não =0,284	Espera-se que os cooperados antigos sejam menos arriscados que os novos
B7	Instituição empregadora	A = 0,243 B + L =-1,519 F = 0,959 DEMAIS = 1,995	Sem hipótese prévia de relacionamento
B8	Modalidade do empréstimo	Consignado = 0,17 Outros = -0,428	Espera-se que os empréstimos consignados tenham uma probabilidade de insolvência
B9	Gênero do cliente	Homem = - 0,523 Mulher = 0,658	Sem hipótese prévia de relacionamento
B11	Atrasos em empréstimos anteriores	Há atrasos anteriores=-2,132 Não possui atrasos anteriores= 0,081	Espera-se que cooperados com histórico de atrasos anteriores sejam mais propensos a serem insolventes
B12	Possui avalista	Possui avalista = -2,738 Não possui avalista = 0,619	Sem hipótese prévia de relacionamento entre variáveis

Quadro 2 Tratamento das variáveis categóricas

Fonte: elaborado pelo autor

6 RESULTADOS E ANÁLISES

Após a definição das variáveis, estas foram submetidas ao Teste-T: duas amostras presumindo-se variâncias diferentes, para que se pudesse verificar se os cooperados insolventes tinham variáveis explicativas estatisticamente diferentes daquelas dos cooperados solventes. Nesse processo foram eliminadas 6 variáveis (A7, A8, B8, A11, B11, A13). Também foi feita a matriz de correlação entre as variáveis para verificar se havia entre elas algumas que fossem altamente correlacionadas. Nessa etapa foi utilizado o Microsoft Excel para realizar os testes.

Foram, então, estimadas 12 equações usando os programas e-views e spss, sendo que duas delas tiveram melhor desempenho e serão analisadas em seguida.

As equações tiveram seu desempenho baseado nos seguintes indicadores, os quais serão explicados segundo Dias Filho e Corrar (2009).

1. Likelihood value: É uma das principais medidas de avaliação geral da Regressão logística, sendo que busca aferir a capacidade de o modelo estimar a probabilidade associada à ocorrência de determinado evento. Este é representado por -2LL (logaritmo natural do likelihood value multiplicado por menos 2). Por si só este índice não tem significado, mas ele serve como base de comparação entre as regressões estimadas, uma vez que não é possível o cálculo direto do R^2 . O likelihood value também pode ser avaliado pelos Índices Cox-Snell R^2 e Nagelkerke R^2 que servem como estimativa percentual do poder de explicação da regressão.

2. Teste Hosmer e Lemeshow: verifica se há diferença entre os valores preditos e os observados em certo nível de significância. Para tanto, dividem-se os casos em dez grupos, aproximadamente, e testa-se a hipótese de os valores preditos e observados não serem diferentes. Como o objetivo é aceitar a hipótese é necessário que o nível de significância fique acima de 0,05.

Por último foram feitas matrizes de classificação, que consistem em tabelas comparando a classificação realizada pelos modelos desenvolvidos com a classificação original das observações da amostra. Essas matrizes são elaboradas por meio da análise de cada observação, visando a conhecer se ela foi corretamente classificada pelos modelos. Os resultados dessa análise são os percentuais de acerto e erro de classificação dos modelos.

6.1 Análise do Modelo 1

O primeiro modelo a ser analisado obteve o melhor percentual de acerto dentre as estimativas feitas e partiu de uma equação inicial na qual utilizaram-se todos os índices até que estes ficaram restritos aos índices A2, B4, B7, A9 e A 12, atingindo os seguintes indicadores de ajustamento mostrados no quadro 4.

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	67,495 ^a	,561	,748

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Quadro 3 - Likelihood value do modelo 1

É possível verificar na tabela acima que a regressão obteve um índice de ajustamento de 56% pelo índice de Cox & Snell R^2 e de aproximadamente 75% pelo índice de Nagelkerke R^2 .

Step	Chi-square	df	Sig.
1	11,711	8	,165

Quadro 4 Teste de Hosmer and Lemeshow do modelo 1.

O teste de Hosmer and Lemeshow mostrado no quadro 5. obteve um valor chi-square de 11,711 e um nível de significância de 0,165, o que demonstra um razoável capacidade de predição do modelo.

No quadro 6. é descrito o modelo 1, sendo que na primeira coluna estão as variáveis, na segunda coluna o valor do coeficiente e na sexta coluna o valor de significância destes.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I. for EXP(B)		
							Lower	Upper	
Step 1 ^a	A2	-,126	,048	6,783	1	,009	,882	,802	,969
	B4	-,965	,500	3,728	1	,054	,381	,143	1,015
	B7	-,508	,280	3,286	1	,070	,602	,348	1,042
	A9	-1,921	,647	8,827	1	,003	,146	,041	,520
	A12	2,336	,913	6,547	1	,011	10,338	1,727	61,867

a. Variable(s) entered on step 1: A2, B4, B7, A9, A12.

Quadro 5 - Coeficientes do modelo 1.

Os coeficientes do modelo gerados tiveram, em sua maior parte, valores que ao nível de 5% foram significantes, com exceção dos índices B4 e B7 que são aceitos ao nível de 10%.

Ao analisar o efeito das variáveis independentes no modelo logístico, é necessário lembrar que a relação entre uma unidade a mais na variável e o impacto na probabilidade não é linear, sendo que, quanto maior o Z em $P_i = \text{Pr}[Y_i=1/X=X_i] = 1/1+e^{-z}$ maior será a probabilidade de insolvência do cooperado.

A2 (Tempo no último emprego) – O coeficiente proposto pelo modelo foi de -0,126, com isso verifica-se que quanto maior o tempo do vínculo empregatício, menor será o resultado de Z e conseqüentemente menor será a probabilidade do cooperado se tornar inadimplente. Esse fato corrobora a expectativa do comportamento dessa variável.

B4 (Estado civil) – O coeficiente proposto pelo modelo foi de -0,965 . Considerando que essa variável explicativa assume os valores -0,785 para os solteiros e 0,535 para os demais estados civis, percebe-se que, se o cooperado for solteiro, a multiplicação de valores negativos aumentará a função Z e, conseqüentemente, a probabilidade de insolvência. O inverso ocorre se o cooperado pertencer às demais classificações de estado civil, um vez que o coeficiente é negativo e a variável assume um valor positivo.

B7 (Instituição empregadora) – O coeficiente dessa variável é de - 0,508, logo nota-se que quanto maior for o valor assumido pela codificação da variável menor será a probabilidade de o cooperado se tonar inadimplente. Tendo isso em vista é possível verificar o seu comportamento. Como a codificação para as instituições é a seguinte:

- Instituição A = 0,243
- Instituições B e L = -1,519
- Instituição F = 0,959
- Demais instituições = 1,995

Percebe-se que os cooperados pertencentes à instituição B e L tem um risco maior de se tornarem insolventes, e que o conjunto formado pela instituição A e demais tem um risco bem inferior.

A9 (Gênero do cooperado) – Como o coeficiente proposto para essa variável é de - 1,921 e a variável assume os valores 0 para os homens e 1 para as mulheres, verifica-se que no conjunto observado as mulheres apresentam um risco menor do que os homens.

A12 (Existência de aval) = O coeficiente definido pelo modelo para essa variável foi de 2,336. Como a variável assume os valores 1 (possui avalista) e 0 (não possui avalista), conclui-se que o grupo de cooperados que possui avalista é mais arriscado do que o grupo sem avalista.

Classification Table^a

Observed		Predicted			
		Z		Percentage Correct	
		Adimplente	Inadimplente		
Step 1	Z	Adimplente	88	9	90,7
		Inadimplente	4	19	82,6
Overall Percentage					89,2

a. The cut value is ,300

Quadro 6 - Tabela de classificação do modelo 1.

Através do quadro 7 é possível verificar que o percentual de acerto foi bastante relevante, sendo que o modelo foi capaz de classificar corretamente 82,6% dos cooperados insolventes e 90,7% dos cooperados solventes, alcançando 89,2% de acerto total. O valor do cut off da probabilidade foi de 0,30 devido à proporção dos dados de cooperados solventes e insolventes.

6.2 Análise do modelo 2

Durante entrevista com funcionário da cooperativa foi questionado as situações em que o aval era solicitado, constatando-se que isso acontecia em duas hipóteses: no caso de o cooperado não ter margem em folha de pagamento para a concessão de um novo empréstimo ou pertencesse à instituição B. Como a divisão por instituição foi considerada para a definição do índice B7, optou-se por excluir o índice A12, verificando assim as modificações ocorridas no modelo. Também teria sido possível excluir a variável B7.

No quadro 8 é descrito o modelo 2, sendo que na primeira coluna estão as variáveis, na segunda coluna o valor do coeficiente e na sexta coluna o valor de significância desses.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	A2	-,100	,040	6,276	1	,012	,905
	B4	-,990	,471	4,408	1	,036	,372
	B7	-,846	,256	10,898	1	,001	,429
	A9	-1,655	,572	8,370	1	,004	,191

a. Variable(s) entered on step 1: A2, B4, B7, A9.

Quadro 7 - Coeficientes do modelo 2.

Neste modelo todos os coeficientes ficaram dentro do nível de 5% de significância.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	75,640 ^a	,530	,707

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Quadro 8 - Likelihood value do modelo 2

Em relação ao primeiro modelo tanto o R² Cox & Snell quanto o de Nagelkerke obtiveram valores semelhantes no segundo modelo como é verificado no quadro 9.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,098	8	,636

Quadro 9 - Teste Hosmer and Lemeshow do modelo 2

Classification Table^a

Observed		Predicted		
		Z		Percentage Correct
Step 1	Z	Adimplente	Inadimplente	
		87	10	89,7
		7	16	69,6
Overall Percentage				85,8

a. The cut value is ,300

Quadro 10 - Tabela de classificação do modelo 2

Nos dois quadros anteriores percebe-se que apesar do percentual de acerto ter ficado aquém da primeira equação, obtendo 69,6% de acerto dos insolventes e 85,8% de acerto total, o teste de Hosmer e Lemeshow obteve um nível de significância bem superior, corroborando a hipótese do modelo 2 ser mais bem ajustado do que o modelo 1.

7 CONSIDERAÇÕES FINAIS

O mercado de crédito é um fator que possibilita o desenvolvimento da economia em seu molde atual. As cooperativas de crédito inserem-se nesse contexto na tentativa de oferecer serviços diferenciados a taxas mais baixas, além de preencher lacunas deixadas por outras instituições financeiras, as quais não atendem mercados menos atrativos como os de micro-empresários.

Percebe-se, ao verificar o atual cenário das cooperativas, que a maioria ainda se encontra administrada de forma pouco eficiente, em relação ao seu real potencial de mercado. Uma das técnicas que pode ser implantada para melhorar seu desempenho é o modelo de *credit scoring*. Este consiste em uma ferramenta para tomada de decisão no processo de concessão de crédito. Isso é feito a partir de modelos estatísticos que buscam antever a condição futura de solvência ou insolvência de um tomador de crédito no momento em que este faz o pedido.

Esta pesquisa teve como objetivo principal o desenvolvimento de um modelo de *credit scoring*, por meio da regressão logística, para uma cooperativa de crédito de Minas Gerais, cujos dados necessários para sua realização foram fornecidos pela mesma.

Após sucessivos ajustes nas regressões, foram encontrados dois modelos que obtiveram melhor desempenho. As principais variáveis que apresentaram representatividade, em ambos os modelos, foram A2 (tempo no último emprego), B4 (estado civil), B7 (instituição ao qual o cooperado é vinculado), A9 (gênero do cooperado) e A12 (existência do aval).

Vale a pena notar que, apesar de não terem sido obtidas informações do endividamento inicial do cooperado e do SERASA devido a limitação do banco de dados, o

fato de o mesmo possuir ou não aval supre em parte essa informação. Isso acontece porque caso ele não possua um endividamento inicial alto e conseqüentemente não tenha margem na folha de pagamento para a concessão do empréstimo, um aval será pedido. Outro fator relevante é a relação existente entre os índices B7 e A12, já que os cooperados que pertencem à instituição B, com algumas exceções, precisam de aval.

Uma possibilidade para justificar o fato dos cooperados da instituição B serem mais arriscados, está no fato de muitos dos empregados dessa instituição serem contratados por períodos determinados.

Embora a pesquisa tenha limitações como a atualização anual do cadastro o que prejudica o lead *time*, provocando uma melhora fictícia do ajustamento do modelo; ainda sim, os dois modelos encontrados obtiveram um alto percentual de acerto total, especialmente o dos insolventes.

Durante a coleta de dados na cooperativa, notou-se que: quase nenhum dos cooperados insolventes haviam tido atrasos em empréstimos anteriores, dando um indício prévio da deterioração de sua situação financeira. Outro ponto interessante é o fato de que a maioria dos cooperados pesquisados não terminava um empréstimo antes do início do empréstimo seguinte, utilizando parte deste para quitar o anterior. Sendo que muitas vezes as parcelas tinham tendências ascendentes aumentando o grau do comprometimento da renda. Deste modo, é possível que um modelo de *behavioural scoring* tenha um melhor desempenho de predição dentro do grupo dos cooperados antigos.

É importante mencionar que por mais mecanizado e impessoal que o processo da análise de crédito se torne, isso nem sempre é o mais adequado ao modelo cooperativista. Logo, exceções podem existir, mas deve-se proceder com cautela, para que a tentativa de beneficiar alguns não acabe prejudicando a cooperativa como um todo.

De forma geral notou-se que a cooperativa de crédito estudada tem, atualmente, uma carteira de crédito saudável. Isso é confirmado tanto pela dificuldade de encontrar cooperados insolventes para serem usados no modelo como também pela entrevista feita com funcionário da cooperativa que apontou um percentual de inadimplência de aproximadamente 3% da carteira.

Nesse sentido é recomendável que em uma pesquisa posterior verifique-se a adequação dos controles internos da cooperativa, já que uma vez que o risco de crédito parece não ser uma preocupação imediata, o risco operacional se não gerenciado pode ser uma brecha para grandes perdas.

REFERENCIAS

ARAÚJO, Elaine aparecida. *Risco de crédito: desenvolvimento de modelo credit scoring para a gestão da insolvência de uma instituição de microcrédito*. IPEA-CAIXA-2006: monografias premiadas, p. 319-350. Brasília: Ipea, 2006.

BECKMAN, Theodore N; BARTELS, Robert. *Credits and collections in theory and practice*. 6. ed. New York: McGraw-Hill, 1955.

BRASIL. Lei 4595, de 31 de dezembro de 1964. Dispõe sobre a Política e as Instituições Monetárias, Bancárias e Creditícias, Cria o Conselho Monetário Nacional e dá outras

providências. Disponível em: <<http://www.planalto.gov.br/ccivil/leis/L4595.htm>> acesso em março de 2009

BRASIL. Lei 5764, de 16 de dezembro de 1971. Define a Política Nacional de Cooperativismo, institui o regime jurídico das sociedades cooperativas, e dá outras providências. Disponível em < http://www.planalto.gov.br/ccivil_03/Leis/L5764.htm> Acesso em março de 2009.

BRASIL. Resolução 3.380. Dispõe sobre a implementação de estrutura de gerenciamento do risco operacional. Disponível em: <

CAUOETTE John B., ALTMAN Edward I., NARAYANAN Paul. *Gestão do risco de crédito: o próximo grande desafio financeiro*; Rio de Janeiro: Qualitymark, 1999

DIAS FILHO, José Maria, CORRAR, Luiz J. Regressão Logística. In. CORRAR, Luiz J.,

PAULO, Edilson, FILHO, José Maria D.(Org.) *Análise Multivariada: para os Cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2009. p. 280-323

GONÇALVES, Eric B. *Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos*.2005, 96f. Dissertação (Mestre em Administração) – Programa de pós-graduação em Administração, Universidade de São Paulo, São Paulo,2005.

HAND, David J. Modelling consumer credit risk. *IMA Journal of Management Mathematics*. V.12, p. 139-155, 2001.

HAND, D. J., HENLEY, W. E. Statistical Classification Methods in Consumer *Credit scoring: A Review*. *Royal Statistical Society*. v. 160, p 523-541,1997

International co-operative alliance: uniting, representing and serving co-operatives worldwide: Statement on the co-operative identity. Disponível em:< <http://www.ica.coop/al-ica/>> acesso em março de 2008

International co-operative banking association Disponível em:< <http://www.icba.coop/>> acesso em março de 2008

KALAPODAS, Evangelos, THOMSON, Mary E. Credit risk assessment: a challenge for financial institutions, *IMA Journal of Management Mathematics*, vol. 17, p. 25-46, maio.2005.

PINHEIRO, Marcos Antonio Henriques. *Cooperativas de crédito: história da evolução normativa no Brasil*. 6 ed. Brasília :BCB, 2008.

SANTOS, Cláudia B. *Gestão de risco de crédito em instituições financeiras*. 2000. 213f. Dissertação (Mestrado em Controladoria e Contabilidade) – Curso de pós graduação em Controladoria e Contabilidade, Universidade de São Paulo, São Paulo,2000.

SAUNDERS, Anthony, ALTMAN, Edward I. Credit risk measurement: Developments over the last 20 years, *Journal of Banking & Finance*, vol.21 nº 11-12, p. 1721-1742, dez. 1997.