

## Mensuração do risco de crédito por meio de análise estatística multivariada

Claudio Silva Palmuti<sup>1</sup>  
Djair Picchiai<sup>2</sup>

**Resumo:** O presente trabalho tem como foco principal estimar e testar um modelo econométrico, aplicando a ferramenta estatística de análise multivariada por meio da Regressão Logística, utilizado na concessão de crédito. Para análise dos dados, foi adotado o procedimento de levantamento de dados, que teve como objeto de estudo a concessão de crédito pelo FAEP – Crédito Produtivo Popular – Crédito Produtivo Popular, uma instituição de microcrédito atuante no Estado de Minas Gerais. Após testar e validar o modelo gerado, concluiu-se, que as técnicas estatísticas multivariadas podem ser utilizadas como ferramentas complementares no gerenciamento do risco de crédito bancário. O modelo estimado classificou corretamente 87,6% dos clientes analisados.

**Palavras-chave:** Microcrédito. Crédito. Risco. *Credit scoring*. Regressão Logística

### Measurement of credit risk analysis using multivariate statistics

**Abstract:** This paper focuses primarily estimate and test an econometric model by applying statistical tool for multivariate analysis by logistic regression, used in granting credit. For data analysis procedure was adopted for data collection, which aimed to study the extension of credit by FAEP - Popular Productive Credit - , a microfinance institution operating in the State of Minas Gerais. After testing and validating the generated model, it was concluded that the multivariate statistical techniques can be used as complementary tools in the management of bank credit risk. The estimated model correctly classified 87.6% of clients analyzed.

**Keywords:** Microcredit. Credit. Risk. Credit scoring.-

**Classificação JEL:** M 1; M 19

### Introdução

O nível da atividade econômica em um país é determinante para o seu crescimento e para a geração de melhores condições de vida para sua população. Toda a produção dos diferentes setores da economia geram o produto nacional por meio do qual os agentes econômicos são remunerados. Esses agentes, famílias, empresas e Governo convertem sua renda em dispêndios para satisfazer suas necessidades de consumo de bens, serviços e investimentos sob a forma de demanda agregada pelo produto nacional, caracterizando um ciclo econômico em que a oferta agregada do produto nacional cria uma demanda agregada pelos bens e serviços desse produto (LEITE, 2000). Um desequilíbrio entre a oferta e a demanda agregada em que a procura se torna maior que o volume ofertado gera a necessidade do aumento da atividade produtiva para restabelecer o equilíbrio. O aumento da produção requer fatores como insumos e mão de obra que exigem capital para sua aquisição e remuneração. Nesse cenário, os agentes econômicos precisam de capital para iniciar ou ampliar sua capacidade produtiva e aproveitar a oportunidade de crescimento. No entanto, é comum a ausência total ou parcial de capital por parte dos agentes da economia em curto intervalo de

---

<sup>1</sup> Mestre em Administração de Empresas pelo Instituto de Ensino de Campo Limpo Paulista (Faccamp). Professor do Departamento de Exatas do Centro Universitário do Sul de Minas (Unis-MG). E-mail: palmuti@unis.edu.br

<sup>2</sup> Doutor e Mestre em Administração de Empresas pela EAESP - FGV. Professor do Programa de Mestrado em Administração de Empresas do Instituto de Ensino de Campo Limpo Paulista. E-mail: djair.picchiai@fgv.br

tempo, o que cria a necessidade de obtenção de crédito para financiamento imediato das atividades produtivas com prazo para pagamento.

Nesse contexto, o crédito aparece como instrumento essencial para o desenvolvimento da economia por financiar seus agentes em suas diferentes necessidades: o Governo demanda crédito para investimento público em infraestrutura e prestação de serviços; as famílias precisam de crédito para satisfazer suas necessidades de consumo de bens duráveis e não duráveis; as empresas necessitam de crédito para financiar suas atividades operacionais e investir em estrutura. Dessa forma, a disponibilidade de crédito pode impulsionar o produto nacional no sentido de atender as demandas e gerar renda para as famílias, lucro para as empresas e receitas tributárias para o Governo (BRUNI, 2005).

As instituições financeiras desempenham uma importante função na economia regional e nacional, ao proverem crédito para os agentes econômicos que necessitam de recursos financeiros. Os bancos comerciais se destacam nesse sentido, ao concederem empréstimos e financiamentos destinados ao consumo e à produção. Essa combinação estimula o crescimento da economia desencadeando efeitos positivos para a população e o Governo.

Para atender as necessidades de capital desse segmento de empresas, os bancos precisam desenvolver métodos específicos de análise do risco de crédito apropriados às suas características. É necessário identificar as variáveis de risco mais relevantes para classificar as MPE's – Micro e Pequenas Empresas — quanto ao grau de risco que oferecem como tomadoras de empréstimos de acordo com o setor de atuação e região em que se encontram e formular modelos que, por meio das variáveis de risco, possibilitem discriminar, com antecedência, empresas com perfil de boas pagadoras de empresas com perfil de más pagadoras.

Tendo em vista a necessidade de crédito apresentada pelas micro e pequenas empresas para financiar suas atividades operacionais de curto prazo, e de investimento a médios e longo prazo, de maneira que possam permanecer no mercado de forma competitiva, gerando postos de trabalho e renda para as comunidades, e a função dos bancos em fomentar o desenvolvimento destas empresas por meio da concessão de crédito de qualidade que também propicie a essas instituições fornecedoras de crédito a realização de seus objetivos, propõe-se aqui à seguinte questão: em uma instituição de microcrédito, pode-se melhorar a qualidade das decisões de concessão de crédito às MPE's — Micro e Pequenas Empresas — por meio da utilização das ferramentas estatísticas de análise multivariada?

## 1. Revisão da Literatura

### 1.1 O Microcrédito

Segundo Pochmann (2010), em 2009, o Brasil registrou a presença de 94,1 milhões de ocupados. Desse universo, 65,8% estavam submetidos ao emprego assalariado, sendo 66% com carteira assinada e 34% sem contrato de trabalho formal. No segmento de pequenos empreendimentos (de um até dez postos de trabalhos), as ocupações se dividiam em 48,7% de trabalhadores por conta própria, 43% de empregados assalariados e 8,3% de empregadores.

Frente à grande heterogeneidade das ocupações no País, torna-se inegável a necessidade de diversificação do sistema de proteção social e trabalhista. Dentro dessa diversidade de ocupações não assalariada, ganha importância a implantação da nova legislação que trata do micro empreendedor individual.

A partir do pagamento mensal de R\$ 51,15 ao Instituto Nacional do Seguro Social, acrescido do abatimento de impostos indiretos como ICMS estadual e ISS municipal, o trabalhador por conta própria com renda de até R\$ 36 mil ao ano poderá ser beneficiado pela regulamentação do micro empreendedor individual. Ou seja, a possibilidade concreta de formalização do negócio próprio, o acesso bancário e a produtos e serviços de créditos, além dos benefícios previdenciários, como a pensão por morte, o salário maternidade e a aposentadoria.

Trata-se, resumidamente, de um grande esforço da política pública de procurar ampliar a cobertura social e trabalhista para quase um quinto das ocupações do país. Mas esse esforço deve ser compartilhado pelo conjunto da sociedade brasileira e, principalmente, ser encampado pelo poderes públicos locais, com o desenvolvimento de ações integradas que promovam o desenvolvimento local, a geração de emprego e renda e a inclusão social e trabalhista.

O processo de inclusão social, para Coutinho *et al.* (2009) pode ser definido, pela redução dos níveis de pobreza. Além da redução da pobreza, assume-se que o acesso à educação e a bens públicos, além da ampliação da participação política e do direito ao trabalho, são fundamentais para que a inclusão social se dissemine. Outro fator é de suma importância: a inclusão financeira.

## 1.2. Intermediação Financeira

Os bancos comerciais atuam como agentes de intermediação financeira no mercado captando recursos junto aos investidores, pessoas físicas, empresas e Governo que possuem fundos excedentes, canalizando-os àqueles que necessitam de recursos para financiar suas atividades, pessoas físicas, empresas e Governo (GITMAN, 2002).

Silva (2003) explica que a intermediação financeira requer algumas condições básicas como a existência de moeda, bases institucionais para o funcionamento da intermediação e a existência de agentes econômicos deficitários e superavitários.

Por meio dessa intermediação, os bancos recebem depósitos à vista relativos a contas de movimentação, poupança, fundos e valores negociáveis de longo prazo tanto de empresas e pessoas físicas, quanto do Governo constituindo, assim, seu passivo, e alocando esses recursos sob a forma de empréstimos e financiamentos às empresas (comercial, industrial e serviços), pessoas físicas (crédito pessoal, imobiliário) e Governo (títulos da dívida pública).

Dessa forma, os bancos devem administrar os riscos inerentes às operações e à rentabilidade prevista visando a cobrir seus custos operacionais, administrativos e tributários e gerar lucro para os acionistas obtendo *spread* bancário (SAUNDERS, 2000).

## 1.3. Crédito e Risco

Para Schrickel (2000), crédito é todo ato de vontade ou disposição de alguém de destacar ou ceder, temporariamente, parte do seu patrimônio a um terceiro, com a expectativa de que essa parcela volte a sua posse integralmente, decorrido o tempo estipulado. Blatt (1999) enfatiza que a palavra crédito tem sua origem no vocábulo latino *credere*, que significa: crer, confiar, acreditar, ou ainda, do substantivo *creditum*, que significa, literalmente confiança. Diante do exposto, crédito corresponde à confiança que o credor deposita no devedor de que ele efetuará o pagamento combinado pelo uso ou aquisição de seu bem.

Em um sentido restrito, crédito corresponde a entregar determinado valor mediante a promessa de pagamento futuro (SILVA, 2003). Nas operações de crédito de uma instituição financeira, o risco é inerente e impossível de ser eliminado (DUARTE Jr.; VARGA, 2003). No entanto, o risco pode ser administrado desde que mecanismos de identificação, mensuração e classificação para este fim sejam desenvolvidos.

O risco deve ser entendido como uma medida da possibilidade de perda financeira ou, mais formalmente, como uma medida da variabilidade dos retornos esperados associados a ativos financeiros (GITMAN, 2002).

## 1.4. Análise de Crédito

Uma análise convencional utilizada pelas instituições e que serve para mensurar o grau de confiança de cada operação baseia-se nos tradicionais “Cs do crédito” - caráter, capital, condições, capacidade, colateral, segundo CAPE (2000 p. 93 – 98):

Para Saunders (2000), os “Cs” do crédito constituem o chamado sistema especialista de abordagem tradicional, e a decisão de crédito fica a cargo do gerente da instituição, cujo conhecimento especializado, julgamento subjetivo e atribuição de peso a certos fatores-chaves são,

implicitamente, as mais importantes determinantes na decisão de conceder crédito, ou não. Stell (2009) traz a ideia de que existe um sexto “C” na análise tradicional. Segundo o autor, este C – Change - representaria a mudança do ambiente econômico em geral.

Blatt (1999) explica que a situação ideal seria aquela em que, por meio de um sistema de computador, combinando as características do cliente e os indicadores financeiros da empresa, fossem simulados todos os cenários possíveis. Esse procedimento resultaria em decisões de crédito automáticas que eliminariam as inconveniências na solicitação de crédito baseadas somente em julgamento subjetivo.

Todavia, a realidade tem demonstrado que a análise de crédito não pode ser realizada totalmente por meio de programas de computadores. Isso quer dizer que modelos computacionais devem ser desenvolvidos e utilizados como ferramentas complementares na análise creditícia, mas não decisórios.

## 1.5 Modelagem de Crédito

Segundo Souza e Chaia (2000), vários pesquisadores têm modelado o comportamento do administrador e do tomador de recursos, utilizando informações que não a taxa de juros. Estudos realizados tanto no Brasil, quanto no exterior foram desenvolvidos objetivando explicar a probabilidade de perdas de crédito, envolvendo basicamente variáveis financeiras e focando o cliente apenas em segundo plano.

### 1.5.1. *Credit scoring*

*Credit scoring* pode ser definido como o processo de atribuição de pontos às variáveis de decisão mediante técnicas estatísticas. Trata-se de processo que estima a probabilidade de que um cliente com certas características, pertença ou não a um grupo possuidor de outras determinadas características consideradas desejáveis, hipótese em que se aprova um limite de crédito. Essa técnica estabelece uma regra de discriminação de um determinado cliente solicitante de crédito (VICENTE, 2001, p.49).

Os sistemas de pontuação de crédito estimam a probabilidade de um cliente vir a ser “bom pagador” ou “mau pagador” com base em suas características: “Existem vários fatores que estão associados à possibilidade de inadimplência. Um modelo de escoragem de crédito combina os fatores mais importantes associados à possibilidade de inadimplência, determina o inter-relacionamento entre eles e atribui números para gerar o *escore* final. A prática tem por objetivo produzir um modelo de escoragem de crédito no qual quanto maior for o *escore*, menor será o risco de perda com devedores duvidosos” (GHERARDI e GHIELMETTI, 1997).

Entre as metodologias tradicionais usadas para a construção de modelos de escoragem de créditos estão a análise discriminante, a regressão logística (Herszkowicz, 2000), as árvores de decisão (Joos *et al.*, 1998; Bursteinas e Long, 2000; Daubie *et al.*, 2002) e a programação linear e suas variações. Thomas (2000) destaca a regressão logística e a programação linear como os métodos mais utilizados atualmente.

### 1.5.2. Análise Discriminante

Os primeiros estudos acadêmicos sobre riscos de insolvência se baseavam em análises de quocientes, índices contábeis, e deduziam que empresas em processos de insolvência apresentavam quocientes bastante diferenciados das solventes. Nesses estudos, uma das principais questões levantadas consistia em se determinar quais índices seriam os mais importantes na predição e qual a ponderação ideal desses índices. Surgiu, então, a análise discriminante multivariada, capaz de solucionar este questionamento (BRUNI, FAMÁ; MURRAY, 1998).

O objetivo fundamental da análise discriminante é a alocação de novos solicitantes em grupos bem definidos, evitando todas as formas de superposição. Os dados de cada elemento de cada um dos grupos são coletados e, em seguida procura-se derivar uma função, que nada mais é que uma combinação linear para melhor discriminar os grupos entre si. O resultado almejado é a

obtenção de um conjunto único de coeficientes para cada uma das variáveis independentes e que classifique cada elemento observado nos grupos previamente definidos (GUIMARÃES 2000, p.33).

A função discriminante linear tem a forma:

$$Y_x = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \quad (1)$$

Onde:

$Y_x$  = Variável Dependente, reflete o valor do *escore*;

$\beta_0$  = Intercepto;

$\beta_1, \beta_2, \beta_3, \dots, \beta_n$  = Pesos (coeficientes) atribuídos a cada variável;

$X_1, X_2, X_3, \dots, X_n$  = Variáveis discriminantes;

A classificação de cada cliente é efetuada a partir da função discriminante estimada, e por meio do cálculo do valor dessa função para cada cliente (*escore*). De acordo com o critério de classificação de Fisher para dois grupos, o cliente deverá ser classificado como bom pagador se estiver mais próximo do centroide desse grupo do que do centroide do grupo dos maus pagadores, isto é, se a distância entre o seu *escore* discriminante e o centroide do grupo 1 for menor que a distância entre o seu *escore* e o centroide do grupo 2, e no grupo dos maus pagadores no caso contrário (COSTA, 1992, p.62).

De acordo com Silva (2003), Stephen C. Kanitz observando que a análise dos índices contábeis isoladamente não oferecia resultados tão bons quanto aqueles obtidos por meio da combinação e comparação um *escore* médio destes índices, desenvolveu o modelo que ficou conhecido como “Termômetro de Insolvência de Kanitz”, que é representado pela função:

$$FI = 0,05 X1 + 1,65 X2 + 3,55 X3 - 1,06 X4 + 0,33 X5$$

Onde:

FI: fator de insolvência;

X1: quociente entre lucro líquido e patrimônio líquido;

X2: quociente entre ativo circulante mais realizável em longo prazo e exigível total;

X3: quociente entre ativo circulante menos estoques e passivo circulante;

X4: quociente entre ativo circulante e passivo circulante;

X5: quociente entre exigível total e patrimônio líquido.

A análise discriminante permite descobrir as ligações que existem entre um caráter qualitativo a ser explicado e um conjunto de caracteres quantitativos explicativos. Também permite prever, por meio de um modelo, os valores da variável que derivam dos valores tomados pelas variáveis explicativas (GIMENES; URIBE-OPAZO, 2001, p. 67).

Saber se um cliente provavelmente honrará seus compromissos é uma informação imprescindível na hora de tomar uma decisão com vistas à concessão de crédito. Com isso, pode-se demonstrar que as instituições financeiras poderiam ter um acréscimo nos lucros se, na concessão de crédito, os critérios fossem mais precisos. De posse da classificação fornecida por um modelo de previsão de risco de crédito, a empresa pode ter um diagnóstico preliminar do provável comportamento do novo cliente, aprovando ou não a concessão do crédito (VASCONCELOS, 2004).

### 1.5.3. Regressão Logística

A técnica de regressão logística caracteriza-se por descrever a relação entre várias variáveis independentes ( $X_j$ ) e uma variável dependente dicotômica ( $Y$ ), representando a presença (1) ou ausência (0) de uma característica (HOSMER; LEMESHOW 1989; KLEINBAUM 1996).

O objetivo na análise de regressão logística é descrever o comportamento matemático de  $Y$  em função dos valores de  $X_j$ . Assim, utilizando o método de estimação da máxima verossimilhança, os parâmetros do modelo são estimados (HOSMER; LEMESHOW, 1989).

A expressão geral do modelo logístico é dada pelas Equações:

$$P(Y = 1) = \frac{1}{1 + e^{-z}} \quad (2)$$

onde:

$$z = \beta_0 + \sum_1^n \beta X \quad (3)$$

De acordo com Hosmer e Lemeshow (1989), a regressão logística tornou-se um método padrão de análise de regressão para variáveis medidas de forma dicotômica. Dessa forma, a principal diferença da regressão logística quando comparada ao modelo linear clássico é que a distribuição da variável resposta segue uma distribuição binomial, e não uma distribuição normal. A esse respeito, Hair *et al.* (2005) afirmam que a regressão logística se assemelha, em muitos pontos, à regressão linear, mas difere, no sentido de prever a probabilidade de um evento ocorrer. Para obter um valor previsto delimitado entre zero e um, usa-se uma relação assumida entre as variáveis independentes e a variável dependente que lembra uma curva em forma de “S”, a distribuição sigmoide.

Os modelos lineares de regressão não podem acomodar tal relação entre as variáveis, já que ela é inerentemente não linear. Por isso, a regressão logística foi desenvolvida para lidar especificamente com essas questões. A regressão logística deriva seu nome justamente dessa transformação logística utilizada com a variável dependente (HAIR *et al.*, 2005).

Camargos, Araujo e Camargos (2010), testaram um modelo de Regressão Logística, com base em uma amostra de 9.232 empresas que obtiveram crédito em uma instituição financeira pública de Minas Gerais no período de junho de 1997 a dezembro de 2005. Utilizaram como variável dependente, a qualidade do crédito (adimplente ou inadimplente). O modelo estimado pelos autores foi composto pela constante (intercepto) e cinco variáveis explicativas, cuja função matemática é apresentada a seguir.

$$P(Y = 1) = \frac{1}{1 + e^{-z}} \quad (4)$$

onde:

$$z = -2,15 + 0,0000245(X1) - 0,329(X20) - 0,0000176(X2) - 0,503(X11) + 0,00330(X12) \quad (5)$$

X1 = valor do financiamento

X20 = Proporção Bens do Avalista sobre o Valor do Financiamento

X2 = Valor dos Investimentos Fixos do Projeto

X11 = Tempo de Atividade da Empresa

X12 = Proporção do faturamento sobre o Valor Financiado

O modelo estimado por Camargos, Araujo e Camargos (2010) classificou corretamente 88,5% das empresas participantes da amostra.



## 2. Procedimentos

### 2.1. Metodologia

No presente trabalho, foi utilizado o método de pesquisa empírico-analítico. Esse método está relacionado a uma abordagem prática que envolve a coleta, o tratamento e análise dos dados com o intuito de investigar a existência de relações causais entre as variáveis em estudo. Os dados analisados possuem natureza predominantemente quantitativa e, por isso, técnicas estatísticas foram empregadas nas mensurações. A validação científica de estudos desenvolvidos mediante a aplicação desse método ocorre por meio de testes de significância dos instrumentos utilizados (MARTINS, 2002).

A escolha do método de *credit scoring* foi consequência da ampla abordagem que pode ser encontrada a seu respeito na literatura específica e pelo seu uso já consagrado pelas instituições financeiras que o implantaram (BLATT, 1999).

A técnica de Regressão Logística é uma derivada do *credit scoring*, em que os dados amostrais são utilizados para estabelecer um sistema de equações, por meio de combinação linear, entre as variáveis de risco que pressupõem-se explicar a inadimplência dos clientes.

Ao utilizar um sistema de pontuação de crédito de crédito tem-se por objetivo a classificação dos solicitantes de crédito de acordo com o seu risco ou probabilidade de inadimplência (DUARTE JR; VARGA, 2003).

Foi desenvolvido um modelo de *credit scoring*, por meio de Regressão Logística, para mensurar o risco de crédito na concessão de empréstimos realizados pelo FAEP - Crédito Produtivo Popular - uma instituição de microcrédito que concede crédito aos empreendedores formais e informais na região do Sul de Minas Gerais, onde utilizaremos como variável dependente, a variável dicotômica qualidade do crédito (adimplente =1 e inadimplente = 0). Este modelo visa a servir de balizador à avaliação e decisão do analista sobre a concessão, ou não, de crédito a um proponente. Os grupos definidos estão apresentados na Tabela 1:

Tabela 1 - Composição da amostra

Group	Count
Grupo 0: Inad.	330
Grupo 1: Adim.	1305

Fonte: Dados gerados pelo SPSS

GRUPO 0: constituído pelos clientes inadimplentes, no momento da coleta de dados, constituído por 330 clientes.

GRUPO 1: constituído pelos clientes adimplentes, no momento da coleta de dados, constituído por 1305 clientes.

### 2.2. Coleta de dados

A eficácia de um modelo de *Credit scoring* depende diretamente das informações utilizadas para avaliar os riscos dos clientes e das operações. A escolha correta dessas informações é o principal segredo para obtenção de um bom modelo (SICSÚ, 2010).

A variável resposta no modelo de *Credit scoring* é a qualidade do crédito (adimplência ou inadimplência) na operação de empréstimo.

As variáveis em análise foram coletadas diretamente no módulo FAC – BI, do sistema de gerenciamento de crédito da instituição FAEP - Crédito Produtivo Popular - que utiliza um *software* integrado de gestão, desenvolvido pela Fácil Informática especificamente para o mercado financeiro e totalmente em conformidade com as normas e resoluções do Bacen (Banco Central) e do Cosif (Contabilidade do Sistema Financeiro).

De acordo com Vazze (2005), variável é toda característica ou elemento de um fenômeno que pode ser observada e classificada ou medida.

Quanto à natureza, as variáveis podem ser qualitativas ou quantitativas. Variável qualitativa é aquela que expressa características associadas ao fenômeno que está sendo estudado, enquanto que variáveis quantitativas ou numéricas expressam mensurações de um determinado fenômeno.

Quanto à posição, as variáveis podem ser independente ou dependente. A variável independente explica as variações ocorridas na variável dependente. Nesta pesquisa, a variável dependente possui natureza qualitativa e está relacionada ao fato de o cliente pertencer ao grupo dos adimplentes ou inadimplentes.

As variáveis independentes utilizadas para explicar a qual grupo o cliente pertence foram divididas em dois grupos, qualitativas ou categóricas e quantitativas ou numéricas, conforme mostra a Figura 1.

#### QUADRO 1 - Variáveis utilizadas na pesquisa

Variáveis de avaliação do risco de crédito na concessão de empréstimos - FAEP - Crédito Produtivo Popular	
Variáveis Categóricas	Variáveis Numéricas
- Utilização	- Valor do Crédito
- Gênero	- Taxa de Juros
- Grau de Formalidade	- Idade
- Garantia Oferecida	- Prazo Pagamento
- Setor Atuação	- Renda Declarada
- Escolaridade	- Valor Prestação

Fonte: Elaborado pelo autor

Os passos referentes à operacionalização da pesquisa, foram os seguintes:

1. Coletaram-se os dados referentes às variáveis apresentadas no Quadro 4, para cada cliente da instituição;
2. Agruparam-se os clientes em dois grupos, de acordo com a sua situação (adimplente ou inadimplente) no momento da coleta de dados. Para clientes inadimplentes usou-se o critério de estar em atraso com os pagamentos há mais de 30 dias, e para clientes adimplentes usou-se o critério de estar em dia com os pagamentos ou com atraso de no máximo 30 dias;
3. Tabularam-se os dados coletados e calculou-se as estatísticas amostrais; e
4. Procedeu-se à análise dos resultados



### 3. Resultados

#### 3.1. Informatização dos dados

A tabulação das informações cadastrais com o uso dos *softwares* estatísticos SPSS versão 13.0 for Windows e BioEstat 5.0, proporcionaram a análise e o tratamento dos dados amostrais, bem como a construção da fórmula de escoragem que combinou os efeitos das variáveis mais relevantes na determinação do risco de crédito.

#### 3.2. Definição das variáveis relevantes no risco

Os valores médios das variáveis foram comparados com o propósito de se verificar se havia ou não diferença estatística significativa entre os grupos gerados. Esse teste é baseado na técnica de análise de variância (ANOVA), que consiste em utilizar a variância das variáveis em estudo dentro dos grupos e entre os grupos formados para comparar segundo um nível de significância pré-estabelecido, sua igualdade (TRIOLA, 1999).

TABELA 2 - Teste F para comparação das médias das variáveis de risco

	F	Sig.
VL_Lib	4.652	.031*
Uso	.695	.405
Garantia	.067	.796
Set_Atu	.306	.580
Formal	50.464	0.000*
Sexo	.033	.855
Renda	8.334	0.004*
Idade	2.688	.101
Escolar	1.202	.273
VL_Prest	24.978	0.000*
Prazo_Pgto	25.324	0.000*
TX_Juros	246.399	0.000*

Fonte: Dados gerados pelo *software* SPSS

\* Variáveis com relevância ao nível de significância 5%

A Tabela 1 fornece a estatística F, uma estatística da medida de igualdade entre as médias das variáveis correspondentes dos grupos 0 e 1. Quanto maior for o valor da estatística F maior é a desigualdade entre os grupos constituídos em relação à variável em questão.

A significância corresponde à probabilidade de se cometer um erro ao se afirmar, com base nos dados amostrais utilizados, que as médias das variáveis dos grupos são diferentes quando de fato elas não o são (NEUFELD, 2003). Dessa maneira, quanto menor for a significância, probabilidade de errar ao afirmar que os grupos provêm de populações diferentes, mais significativa é a variável para a distinção dos clientes que compõem os grupos. Pela tabela anterior, verifica-se que as variáveis mais relevantes para a formação dos grupos, com nível de significância a 5%, foram: Valor Liberado, Grau de Formalidade, Renda, Valor Prestação, Prazo de Pagamento e Taxa de Juros. As demais variáveis mostraram-se pouco significativas, ao nível de 5% de significância, para serem usadas enquanto medida de dissimilaridade entre os grupos.

### 3.3. Definição dos coeficientes das variáveis de risco

Para o presente trabalho, a classificação prévia do solicitante de crédito depende da existência de um modelo de previsão construído com base nas variáveis de risco mais relevantes identificadas no teste de comparação das médias.

A técnica estatística utilizada no desenvolvimento do modelo é a regressão logística, possibilitando a geração de uma função matemática cuja resposta permite medir a probabilidade de uma observação pertencer a um grupo previamente estabelecido, dado o comportamento de um conjunto de variáveis explicativas. Esse fator é particularmente importante nos modelos de risco de crédito, pois possibilita que seja medida a probabilidade de um determinado tomador de recursos assumir a condição de insolvente, em face das suas características.

No desenvolvimento do modelo, utilizou-se a técnica FORWARD WALD, com o intuito de identificar uma relação matemática entre as variáveis de entrada, aquelas com relevância no teste de comparação das médias, e os estados de inadimplente e adimplente dos participantes da amostra. Adotando um nível de significância é estabelecido um valor crítico ou fronteira até o qual o *score* fornecido pelo modelo é considerado como pertencente ao grupo em relação ao qual o cliente é agrupado.

O modelo final gerado pela técnica de regressão logística Forward Wald, foi composto por seis variáveis explicativas e um ponto de intercepto, conforme a tabela seguinte.

TABELA 3 - Variáveis de risco mais significativas para classificação dos clientes

	B	S.E.	Wald
Step 1 <sup>a</sup> VL_Lib	.006	.000	1.274
Formal	.557	.158	12.426
Renda	.115	.077	2.194
VL_Prest	.235	.108	4.764
Prazo_Pgto	-.262	.129	4.116
TX_Juros	19.392	5.129	14.295
Constant	-.853	.501	2.897

Fonte: Dados gerados pelo SPSS

As variáveis apontadas pela técnica de regressão logística, como sendo as mais relevantes para distinção dos clientes, entre adimplentes e inadimplentes, resultaram na seguinte equação:

$$z = -0,853 + 0,006(vl\_lib) + 0,557(formal) + 0,115(renda) + 0,235(vl\_prest) - 0,262(prazo\_pgto) + 19,392(tx\_juros) \quad (6)$$

### 3.4 Análise do poder preditivo das variáveis de risco

Ao analisar o cruzamento de uma variável explicativa potencial com a variável explicada (Adimplente ou Inadimplente) é possível ter uma noção do poder discriminador dessa variável. Essa análise pode ser realizada simplesmente analisando visualmente a tabela de contingência ou utilizando testes econométricos (SICSÚ, 2010).

Considerando as porcentagens por linha, ou seja, a distribuição “Grupo” dentro de cada categoria da variável, compara-se o percentual de cada categoria com o percentual total da amostra, representado na última linha da tabela.

Sicsú (2010, p. 77) enfatiza que se uma categoria não tiver relação com o “Grupo”, suas porcentagens serão próximas da porcentagem total da amostra, e essa variável parece não ter caráter discriminador significativo.

### 3.4.1. Análise Estatística – Qui-Quadrado

Constitui-se em um teste não-paramétrico com o objetivo de testar se as porcentagens observadas nas categorias de cada grupo são significativamente diferentes ou não.

Utiliza-se da seguinte expressão:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \quad (7)$$

Onde:

$\chi^2$  = é uma estatística de teste com distribuição Qui-Quadrado, com (r-1)(c-1) graus de liberdade;

oi: proporção observada;

ei: proporção esperada

Aplicando-se o teste Qui-Quadrado - com o auxílio do *software* estatístico BioEstat 5.3 - às tabelas de contingências apresentadas acima, encontramos os seguintes resultados:

TABELA 4 - Teste Qui-Quadrado aplicado às variáveis de risco

	VL_Liberado	Formal	Renda	VL_Prestação	Prazo Pgt	Tx_Juros
Soma	242,000	67,800	154,800	201,100	172,700	666,300
X <sup>2</sup>	38,25	31,416	33,502	34,202	33,220	33,954
GL	5	1	3	4	3	13
(p)	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	0,0012

Fonte: Dados gerados pelo BioEstat 5.3 – adaptados pelo autor

Analisando a Tabela 4, é possível observar que as porcentagens da cada categoria de cada variável considerada de risco, diferem significativamente da porcentagem total de cada grupo na amostra, sendo essa variável considerada relevante na discriminação de clientes, quanto à pontualidade nos pagamentos.

### 3.5. Análise do poder preditivo do modelo

A capacidade de previsão do modelo pode ser examinada por meio de uma matriz que mostra o percentual de clientes classificados correta e incorretamente. Conforme exibe a matriz na Tabela 5, o nível de acerto do modelo é de 87,4% com 1379 clientes da amostra classificados corretamente. Do grupo de clientes inadimplentes, 173 foram classificados corretamente e 157 classificados incorretamente, o que representa uma taxa de acerto de 52,4%. Do grupo de clientes adimplentes, há 1206 classificações corretas e 99 incorretas, o que corresponde a uma taxa de acerto de 92,4%.

TABELA 5 - Resultados da Classificação pelo modelo

Observed		Predicted			Percentage Correct
		Grupo		Correct	
Step 1	Grupo	Inad	Adim		
	Inad	173	157		52,4
	Adim	99	1206		92,4
Overall Percentage					87,4

a. The cut value is .500

Fonte: Dados gerados pelo SPSS

A significância conjunta das variáveis explicativas incluídas no modelo foi avaliada por meio do teste Qui-Quadrado, sob a hipótese nula de que todos os coeficientes são iguais a zero. O resultado do teste está representado na Tabela 6.

TABELA 6 - Análise da significância conjunta das variáveis incluídas no modelo

	Chi-square	df	Sig.
Step 1	39.387	6	.000

Fonte: Dados gerados pelo SPSS

Tendo em vista que o valor crítico para o teste Qui-Quadrado ao nível de significância de 5%, com grau de liberdade 6 é de 12,59, é inferior ao valor calculado pelo *software* estatístico SPSS, conforme representa a Tabela 6, apoiamos a afirmação de que as variáveis de risco incluídas no modelo, apresentam relevante poder de classificação dos clientes nos grupos ajustados.

### 3.5.1. Validação do nível de ajuste do modelo estimado

A avaliação do nível de ajuste de um modelo logístico pode ser realizada por meio do *Likelihood Value*, dos pseudos  $R^2$  e do teste de Hosmer e Lemeshow (HAIR *et al.*, 2005). O *Likelihood Value* é uma medida da qualidade geral do modelo e quanto menor o seu valor, melhor é o ajuste do modelo como um todo. O Nagelkerke  $R^2$  é uma medida que se assemelha ao coeficiente de determinação da regressão linear. Maiores valores dessa medida, que tem escala de zero a um, indicam melhor ajuste do modelo.

A Tabela 7 exibe os valores do *Likelihood Value* e do Nagelkerke  $R^2$  do modelo. O valor do Nagelkerke  $R^2$  indica que 83,6% da variância da variável dependente é explicada pelas variáveis independentes do modelo.

TABELA 7 - Avaliação de ajuste do modelo

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1327.879	.615	.836

Fonte: Dados gerados pelo SPSS

O teste de Hosmer e Lemeshow avalia as diferenças entre as classificações previstas pelo modelo e as observadas (HOSMER; LEMERSHOW, 1989). Se as diferenças forem significativas, o grau de acurácia do modelo não é bom. O teste divide os casos em classes e compara as frequências preditas e observadas em cada classe por meio de uma estatística Qui-Quadrado.

A Tabela 8 exibe os resultados do teste. Ao nível de significância de 5%, não se pode rejeitar a hipótese nula de que não há diferença significativas entre os valores preditos e observados, o que indica que o modelo é capaz de produzir classificações confiáveis.

TABELA 8 - Resultado teste Hosmer e Lemeshow

Step	Chi-square	df	Sig.
1	22.534	8	.003

Fonte: Dados gerados pelo SPSS

#### 4. Discussão

Observando que a distribuição dos *escores* dos grupos constituídos pelos clientes adimplentes e inadimplentes apresentam distribuição normal, define-se o intervalo de confiança para mensuração e análise do risco de crédito, por parte de um dado cliente, de acordo com seu perfil apontado pelas variáveis consideradas de risco.

A estatística descritiva referente ao grupo dos clientes adimplentes revela que o *escore* discriminante médio do foi de 31,179 com desvio-padrão de 16,382. O *escore* mínimo de 4,822 e o *escore* máximo de 62,390 confirmaram a ausência de *outliers* que tornaram a distribuição assimétrica positiva (0,316). A margem de erro para os clientes adimplentes igual a 0,890 ao nível de confiança de 95% implica que, mantidas as características dos clientes que constituem este grupo – clientes adimplentes - ao longo do tempo, há uma tendência de que os *escores* de futuros clientes, estejam situados no intervalo  $30,289 < Z < 32,069$  (TRIOLA, 1999).

A estatística descritiva referente ao grupo dos clientes inadimplentes revela que o *escore* discriminante médio do foi de 19,920 com desvio-padrão de 11,837. O *escore* mínimo de 3,579 e o *escore* máximo de 46,603 confirmaram a ausência de *outliers* que tornaram a distribuição assimétrica positiva (0,614). Assim, mantidas as características dos clientes que constituem este grupo – clientes inadimplentes - ao longo do tempo, em repetidas amostras 95% dos clientes prováveis de serem inadimplentes, obterão *escores* situados no intervalo  $18,638 < Z < 21,202$  (TRIOLA, 1999).

Aplicando-se a expressão da Regressão Logística ao *escore* padronizado de cada cliente, é possível estabelecer a probabilidade de que um determinado cliente venha a ser adimplente, ou seja, venha a pertencer ao grupo 1, sendo que maiores *escores* geram maiores probabilidades de o cliente vir a ser adimplente. Analisando os *escores* médios dos dois grupos gerados, é possível estabelecer um *escore* de corte igual a 25,55. Um proponente de crédito ao ter seu *escore* comparado com o *escore* de corte já pode ser previamente classificado. Se o *escore* do proponente for maior que 25,55 maior semelhança com o grupo dos clientes inadimplentes ele terá, e a probabilidade de vir a ser um bom cliente para a instituição será maior. Caso contrário, maior semelhança com o grupo dos inadimplentes ele terá, e a probabilidade de vir a ser um bom cliente para a instituição será menor.

#### Considerações finais

Por meio deste trabalho, verificou-se a possibilidade de complementar e aprofundar o processo de análise e concessão de crédito na instituição FAEP - Crédito Produtivo Popular –por meio do uso da técnica estatística multivariada de regressão logística.

Observa-se que há na literatura várias ferramentas de análise e mensuração do risco de crédito que vão desde as tradicionais técnicas subjetivas ou clássicas, como é o caso dos 5 Cs do crédito, até as técnicas estatísticas multivariadas. Cada ferramenta utilizada na mensuração, avaliação e controle do risco de crédito possui suas características particulares.

O modelo estimado nesta pesquisa identificou como sendo variáveis de risco relevantes para determinar a probabilidade de um solicitante de crédito vir a ser bom cliente, como sendo: valor liberado; grau de formalidade; renda do solicitante; idade do solicitante; valor da prestação assumida; prazo para pagamento e a taxa de juros cobrada.

As variáveis explicativas do risco de crédito, identificadas no presente trabalho, precisam ser confirmadas pelo analista de crédito em sua entrevista com o proponente, pois são decisivas no controle do risco envolvido na operação de crédito.

Reunir informações cadastrais, de natureza qualitativa e / ou quantitativa, acerca de um cliente, e resumi-las em um *escore* representativo de uma probabilidade de ocorrência em atrasos nos pagamentos, é uma medida que pode contribuir para tornar o processo decisório mais objetivo.

Cabe ressaltar que as variáveis de risco que se apresentaram mais relevantes para a análise de crédito nesta pesquisa são limitadas em função do tempo. A validade e eficiência do uso das técnicas multivariadas dependem essencialmente de atualização contínua dos dados relativos aos clientes.

A utilização de outras variáveis explicativas do risco de crédito, não incluídas nesse trabalho, poderá melhorar a significância do modelo de classificação, sendo, portanto, uma perspectiva para pesquisas posteriores.

A função estimada no presente trabalho, se aplicada à relação da regressão logística, apresenta 87,6% de acerto na classificação dos clientes quanto ao perfil de inadimplência ou inadimplência.

## Referências

- BLATT, A. **Avaliação de Riscos e Decisões de Crédito: um enfoque prático**. São Paulo: Nobel, 1999.
- BRUNI, A. L. **Mercados Financeiros**. São Paulo: Atlas, 2005.
- BRUNI, A. L.; FAMÁ, R.; MURRAY, A. D. **Modelos Brasileiros Preditivos de Risco de Crédito: um estudo exploratório atual sobre as suas eficácias**. São Paulo, p. 148-167, 1998.
- BURSTEINAS, B; LONG, J.A.; **Tree structured classifiers, interconnected data, and predictive accuracy**. Intelligent Data Analysis, Vol. 4 Issue 5, p397, 14p, 2000.
- CAMARGOS, M. A.; ARAUJO, E. A. T.; CAMARGOS, M. C. S.. **A Inadimplência na Concessão de Crédito de uma Instituição Financeira Pública de Minas Gerais: Uma Análise Utilizando Regressão Logística**. In: XIII SEMEAD – SEMINÁRIOS EM ADMINISTRAÇÃO. SÃO PAULO – SETEMBRO 2010.
- CAPE – Instituto Centro de Capacitação e Apoio Empreendedor. **Manual do Agente de Microfinanças**. Belo Horizonte: Sebrae, 2000.
- COSTA, Francisco C. **Avaliação de crédito pessoal**. Dissertação de Mestrado. Instituto Superior de Ciências do Trabalho e da Empresa – Portugal. Lisboa: Julho/1992.
- COUTINHO, L. G.; AMBROZIO, A. M. H. P.; SANT'ANA, A. A. e FRANCO MONTORO, G. C. **Inclusão Financeira no Brasil: o papel do BNDES. In Perspectivas e desafios para inclusão financeira no Brasil: visão de diferentes atores / coordenação : Luiz Edson Feltrim, Elvira Cruvinel Ferreira Ventura, Alessandra von Borowski Dodl - Brasília : Banco Central do Brasil, 2009, p. 53-66**
- DAUBIE, M.; LEVECQ, P.; MESKENS, N.. **A Comparison of the Rough Sets and Recursive Partitioning Induction Approaches: An Application to Commercial Loans. International Transactions in Operational Research**, Vol. 9 Issue 5, p681, 14p, Setembro de 2002.
- DUARTE JR, A. M.; VARGA, G. **Gestão de Riscos no Brasil**. Rio de Janeiro: Financial Consultoria. 2003.
- GHERARDI, C. e GHIEMETTI, S. – Revista “Tecnologia do Crédito” Ano 01, Nr. 02, Artigo: **Escoragem de Crédito: Metodologia que Identifica Estatisticamente o Risco de Crédito** – São Paulo: Serasa, Setembro, 1997.
- GIMENES, R. M. T.; URIBE-OPAZO, M. A. **Previsão de Insolvência de Cooperativas Agropecuárias, por meio de modelos multivariados**. Disponível em: [http://www.fae.edu/publicacoes/pdf/revista\\_da\\_fae/fae\\_v4\\_n3/previsao\\_de\\_insolvencia.pdf](http://www.fae.edu/publicacoes/pdf/revista_da_fae/fae_v4_n3/previsao_de_insolvencia.pdf) acessado em 05 de fevereiro de 2012
- GITMAN, L. J. **Princípios de Administração Financeira**. 7. Ed. São Paulo: Harbra. 2002.
- GUIMARÃES, Inácio A. **Construção e avaliação de uma regra de conhecimento e classificação de clientes de uma instituição financeira com base em análise multivariada**. (Dissertação de Mestrado em Economia) - Universidade Federal do Paraná. Curitiba: UFPR, 2000.



- HAIR, J. F. *et al.* **Análise multivariada de dados**. 5 ed. Porto Alegre: Bookman, 2005.
- HERSZKOWICZ, Fabio. **Credit scoring: A Aplicação de Métodos Estatísticos na Avaliação de Risco de Crédito**. 2000. 90p. Monografia – Faculdade de Economia, Administração e Contabilidade da USP, São Paulo.
- HOSMER, D. W.; LEMESHOW, D. W. **Applied logistic regression**. New York: John Wiley & Sons, 1989.
- JOOS, P., VANHOOF, K., SIERENS, N., OOGHE, H. **Credit classification: A comparison of logit models and decision trees**. Workshop on the Application of MachineLearning and Data Mining in Finance, Chemnitz, 1998.
- LEITE, J. A. A. **Macroeconomia: teoria, modelos e instrumentos de política econômica**. 2.ed. São Paulo: Atlas, 2000.
- MARTINS, Gilberto de Andrade. **Manual para Elaboração de Monografias e Dissertações**. 3.ed. São Paulo: Atlas, 2002.
- NEUFELD, J. L. **Estatística Aplicada à Administração usando Excel**. Tradução: José Luiz Celeste. São Paulo: Pearson Prentice Hall, 2003.
- POCHMANN, M. O Futuro das Ocupações em Pequenos Empreendimentos. **Jornal Valor Econômico**, São Paulo, 11 fev. 2010, p. A11.
- SAUNDERS, A.. **Administração de Instituições Financeiras**. São Paulo: Atlas, 2000.
- SCHRICKEL, W. K.. **Análise de Crédito. Concessão e Gerência de Empréstimos**. 5.ed. São Paulo: Atlas, 2000.
- SICSÚ, A. L.. **Credit scoring: Desenvolvimento – Implantação – Acompanhamento**. 1. ed. São Paulo: Blucher, 2010.
- SILVA, J. P.. **Gestão e Análise de Risco de Crédito**. 4.ed. São Paulo: Atlas, 2003.
- SOUSA, A. F. de; CHAIA, A. J. **Política de Crédito: uma análise qualitativa dos processos em empresas**. Caderno de pesquisas em administração, São Paulo, v.7, n° 3, julho/setembro 2000.
- STELL, Warren. Change: **The Sixth C of Credit**. The RMA Journal. Philadelphia: Sep 2009. Vol. 92, Iss. 1, p. 58-60 (3 pp.)
- THOMAS, L. C. A Survey of Credit and Behavioral Scoring: forecasting financial risk of lending to consumers, **International Journal of Forecasting**, Edinburg-U.K, v.16, p. 149-172, 2000.
- TRIOLA, M. F. **Introdução à Estatística**. 7 ed. Rio de Janeiro: LTC, 1999.
- VASCONCELOS, R. S. **Modelos de Escoragem de Crédito aplicados a Empréstimo Pessoal com Cheque**. Rio de Janeiro, 2004. Dissertação (Mestrado em Finanças e Economia Empresarial) – Escola de Pós-Graduação em Economia, Fundação Getúlio Vargas.
- VAZZE, A.. **Avaliação do Risco de Crédito em uma Instituição de Microcrédito**. 2005. 86 f. Dissertação (Mestrado em Administração). Faculdade Cenecista de Varginha, Varginha, 2005.
- VICENTE, E. F. R. **A estimativa do risco na constituição da PDD**. Dissertação (Mestrado em Administração). Universidade de São Paulo - USP. São Paulo: Maio/2001.

Recebido em 02.04.2012

Aprovado em 27.02.2013

