

O MODELO DE PREVISÃO DE SOLVÊNCIA EM INSTITUIÇÕES FINANCEIRAS ATRAVÉS DA REGRESSÃO LOGÍSTICA BINÁRIA

Rodrigo Pereira de Souza¹, Manuel Luiz Figueroa² e Marcos Santos Oliveira³

1 Aluno Graduação - DECAT – Universidade Federal de Sergipe

2 Prof. Dr.Associado - DECAT – Universidade Federal de Sergipe

3 Analista de Crédito – Fundação Gbarbosa

RESUMO: Com a entrada de novos bancos estrangeiros e conseqüentemente a criação de diversas instituições financeiras no Brasil, é necessária a criação de novos métodos de atuação para analisar o perfil dos clientes na área de concessão de empréstimos. Ao apresentar essa técnica, que é a regressão logística binária, contribuimos em grande parcela no momento de avaliar o risco de crédito. Com isso foi desenvolvido um modelo de regressão logística, capaz de estimar a probabilidade de solvência dos clientes, modelo o qual foi utilizado como ferramenta no processo decisão de concessão de empréstimos. Foi utilizado um banco de dados contendo 534 observações de clientes. Mediante os resultados apresentados, o modelo gerado satisfaz o propósito do estudo em questão. Foi utilizado o software SPSS 17.0 e o método de análise da regressão logística foi *Enter*, que além de selecionar as variáveis significativas, também nos mostra variáveis que servirão como indicadores de pontuação dos clientes.

PALAVRAS-CHAVE:Regressão Logística Binária, Credit Score e Concessão de Empréstimos.

1. INTRODUÇÃO

Mudanças ocorridas no cenário financeiro mundial, a partir dos anos 90, tais como, desregulamentação das taxas de juros e câmbio, aumento de liquidez e aumento da competição bancária, fizeram com que as instituições financeiras se preocupassem cada vez mais com o risco de crédito.

A necessidade de controle e gerenciamento eficaz do risco fez com que essas instituições passassem a primar pelo aperfeiçoamento das técnicas utilizadas para essa função, com o desenvolvimento de inúmeros modelos pelas mesmas e pelas empresas de consultoria.

Neste trabalho, construímos um modelo de regressão logística binária que será utilizado na avaliação de risco de crédito nessas instituições.

2. METODOLOGIA

Os modelos de Credit Score são sistemas que atribuem pontuações às variáveis de decisão de crédito de um proponente, mediante a aplicação de técnicas estatísticas. Esses modelos visam a segregação de características que permitam distinguir os bons dos maus clientes [1]. O escore pode ser utilizado para classificação de créditos como solventes ou insolventes, bons ou maus, de acordo com a pontuação obtida por cada cliente. Esta classificação, por sua vez, pode orientar a decisão do analista em relação à concessão ou não do crédito solicitado. [2]

2.1. Regressão Logística Binária

Em 1960, uma técnica foi desenvolvida para investigar a relação entre variáveis explicativas, métricas e não métricas e uma variável dependente categórica binária.[3]

Muitas variáveis dicotômicas, binomiais, podem ser estudadas como dependentes de outras variáveis contínuas ou categóricas. A situação da saúde de um paciente, curado ou não curado, a previsão de um freguês comprar ou não uma mercadoria, a previsão do sucesso ou não de um estudante, são exemplos de variáveis dicotômicas da regressão logística binária.[4]

A função logística é dada pela expressão:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} + \varepsilon \quad (I)$$

Onde:

$\pi(x)$ é a variável dependente denotada por y ;

x é a variável explicativa ;

e é a base do logaritmo natural;

β_0 e β_1 são os parâmetros a serem estimados;

ε é o erro aleatório que se supõe com média zero.

Com transformações algébricas convenientes de (I), obtém-se:

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (II)$$

2.2 Teste de Hormes-Lemeshow

Verifica o ajuste de um modelo de regressão logística gerado através das probabilidades previstas e observadas em faixas e através da aplicação de um teste de Qui-quadrado (X^2). O interesse é que não rejeite a hipótese nula para que não haja essas diferenças significativas. [3]

2.3 Curve Receiver Operating Characteristic (Curva ROC)

Trata-se de um gráfico elaborado quando se deseja efetivar uma análise preditiva do modelo, gerando uma Tabela de Classificação que se dá através do estabelecimento de um ponto de corte, cujos valores das probabilidades estimadas acima deste ponto indicam a ocorrência do evento de interesse e os valores abaixo indicam a não-ocorrência.[3] Disponibilizou também uma referência usual em relação à área da Curva ROC. Contida na tabela 1:

Tabela 1: Discriminação da área da Curva ROC

Área abaixo da Curva ROC (u.a.)	Interpretação
Menor ou igual a 0,5	Não há discriminação
Entre 0,7 e 0,8	Discriminação aceitável
Maior que 0,8	Discriminação excelente

3. RESULTADOS E DISCUSSÃO

As variáveis dummies que integrarão a equação foram categorizadas de outras variáveis e serão selecionadas através da análise técnica, o método *Enter*, pois este processo não desconsidera nenhuma variável mesmo que exista correlação entre elas. A análise técnica incorporou no modelo todas as variáveis, utilizando principalmente quando se tem certeza de que todas as variáveis são necessárias para se estimar os parâmetros do modelo. O método disponibilizou-nos 15 variáveis e a constante para compor o modelo final. São elas: Capital, ocupação1, ocupação2, ocupação3, ocupação4, ocupação5, ocupação6, ocupação7, Lim1, Lim2, idade1, idade2, idade3, sexF, pessoa, constante.

Tabela 2: Teste de Hosmer-Lemeshow

Passo	Qui-Quadrado	g.l.	Sig.
1	8,852	8	,355

$$H_0: \text{Clas. Obs.} = \text{Clas. Prev.}$$

$$H_1: \text{Clas. Obs.} \neq \text{Clas. Prev.}$$

Não se rejeita a hipótese nula, o que nos leva a concluir que não há diferenças significativas entre as probabilidades observadas e previstas.

Tabela 3: Tabela de classificação

Passo	Observado		Previsto		
			insolvente	solvente	Percentual de acerto
			0	1	
1	insolvente	0	164	30	84,5
	solvente	1	3	337	99,1
Percentual de acertos			93,8		

Pela Tabela de Classificação podemos compreender os conceitos de *Sensitividade* e *Especificidade* que irão gerar a Curva ROC. A *Sensitividade* corresponde aos acertos verdadeiros positivos, quando é *Solvente* nas classificações observadas e nas previstas simultaneamente, enquanto a *Especificidade* corresponde aos acertos verdadeiros negativos, quando é *Insolventes* nas classificações observadas e previstas simultaneamente. Logo abaixo temos a Curva ROC, que confirma os procedimentos de teste anteriores, e qualifica muito bem o modelo.

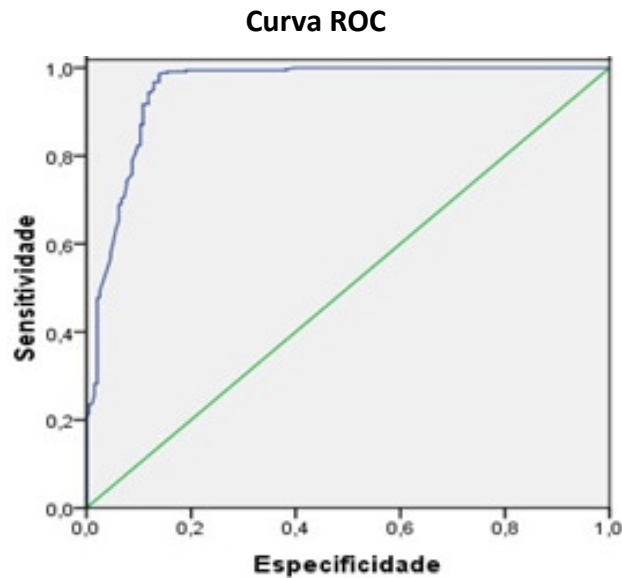


Figura 1: Área da Curva ROC

Área abaixo da curva ROC é de 0,952, dentro da faixa que a discrimina como excelente.

4. CONCLUSÕES

O modelo obtido,

$$\ln(\text{odds}) = 1,985 - 0,747 \text{capital} + 0,368 \text{ocupação 1} + 1,869 \text{ocupação 2} + 1,491 \text{ocupação 3} + 0,263 \text{ocupação 4} + 0,181 \text{ocupação 5} + 0,297 \text{ocupação 6} - 0,083 \text{ocupação 7} - 5,632 \text{Lim1} + 1,403 \text{Lim 2} - 0,486 \text{idad 1} - 0,398 \text{idad 2} - 0,575 \text{idad 3} - 0,270 \text{sexF} - 0,403 \text{pessoa}$$

apresentou resultados satisfatórios para poder classificar os clientes na hora de tomar a decisão de conceder o empréstimo. Portanto podemos concluir que realmente é possível a utilização da Regressão Logística Binária pelas instituições financeiras de empréstimos, como suporte no momento de classificar o perfil dos clientes.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] LEWIS, E. An Introduction to Credit scoring. Fair Isaac: San Rafael, California. 1992.
- [2] SAUNDERS, Anthony. Medindo o Risco de Crédito – Novas Abordagens para o Value at Risk e Outros Paradigmas. Rio de Janeiro: Qualitymark, 2000.
- [3] FÁVERO, L. P.; et al.. *Análise de Dados: modelagem multivariada para tomada de decisões*. Rio de Janeiro: Elsevier, 2009.
- [4] HOSMER, D. W. e LEMESHOW. S. (1998). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley Series in Probability and Statistics.
- [5] OLIVEIRA, M. S.. *Discriminação do Comportamento de Clientes: "Behaviour Score" Uma aplicação da Regressão Logística Binária*. Monografia, DECA - UFS. São Cristóvão - SE, 2009.
- [6] OLIVEIRA, F. E. M.. *SPSS Básico para Análise de Dados*. Rio de Janeiro, Ed. Ciência Moderna, 2007.