

## **Avaliação do Risco de Crédito: Modelos de Regressão Logística com amostras de diferentes proporções**

**Mariana Nolde Pacheco**<sup>1</sup>

**Lisiane Priscila Roldão Selau**<sup>2</sup>

**Resumo:** O objetivo do estudo é propor um modelo de risco de crédito, utilizando regressão logística binária, em que foram propostas duas amostragens de clientes: com proporção igual e desigual de clientes bons e maus, bem como diferentes pontos de corte para classificação desses clientes. O estudo foi realizado através da avaliação de dados reais de concessão de crédito. Nos modelos propostos pelo estudo, os valores do teste KS e da área abaixo da curva ROC foram muito semelhantes em ambas as amostras (igual e proporcional), bem como o percentual de acerto geral, com valores em torno de 60% de acerto para todos os pontos de corte. Já a avaliação da classificação das categorias (bons e maus), evidenciou resultados diferentes nas categorias. O estudo mostrou que ao aumentar o ponto de corte, melhoramos a classificação dos maus clientes, mas pioramos o percentual de acerto dos bons clientes. Dessa forma, os resultados do estudo sugerem que a proporção das categorias de clientes da amostra, bem como os pontos de corte da classificação do modelo devem ser considerados de acordo com o objetivo de classificação da empresa.

**Palavras-chave:** *Credit Scoring, Balanceamento Amostral, Regressão Logística.*

### **1 Introdução**

No Brasil, com o crescente desenvolvimento econômico dos últimos anos, houve um aumento da demanda e da concessão de crédito à população. Segundo Brito (2008), o conceito de crédito consiste na atividade de colocar um valor à disposição de um tomador de recursos, com o compromisso do pagamento do valor emprestado em um determinado período de tempo previamente estabelecido. Dessa forma, a concessão de crédito envolve diversos riscos, uma vez que há a possibilidade de não cumprimento das obrigações financeiras estabelecidas.

De acordo com Steiner et al. (1999), a análise correta da concessão de crédito é essencial para a sobrevivência das instituições financeiras, pois um erro na decisão de conceder o crédito pode significar um grande prejuízo financeiro dentro de uma única operação, gerando a perda do ganho obtido em outras diversas operações bem-sucedidas. Dessa forma, tem aumentado a necessidade de as empresas buscarem diferentes formas de identificar e diferenciar o bom e o mau pagador gerando,

---

<sup>1</sup> UFRGS - Universidade Federal do Rio Grande do Sul. Email: marinolde@yahoo.com.br

<sup>2</sup> UFRGS - Universidade Federal do Rio Grande do Sul. Email: lisianeselau@gmail.com

consequentemente, uma minimização do prejuízo obtido com transações malsucedidas bem como um acréscimo na rentabilidade da instituição.

As empresas utilizam diferentes técnicas para concessão de crédito aos seus clientes. Geralmente a avaliação é realizada através do uso de uma variedade de informações vindas de diferentes fontes, como os dados cadastrais do cliente na instituição. Embora os gestores consigam muitas vezes identificar fatores que diferenciam o bom e o mau pagador, esses critérios geralmente são subjetivos e errôneos, gerando prejuízos financeiros e morais não somente para as instituições como também aos clientes. Sendo assim, surge a necessidade da substituição desses critérios de avaliação subjetivos pelo uso de técnicas quantitativas que melhorem a tomada de decisão das empresas, não apenas diminuindo a concessão de crédito aos maus pagadores como também aumentando o crédito aos potenciais bons pagadores (SELAU, 2011).

Os modelos de risco de crédito, também conhecidos como *Credit Scoring*, são ferramentas de avaliação para classificação dos clientes, e apresentam como principal objetivo identificar previamente o bom e o mau pagador evitando transações financeiras equivocadas. Uma das técnicas utilizadas para avaliação do risco de crédito é a regressão logística, que analisa o efeito de uma ou mais variáveis explicativas (categóricas ou métricas) sobre uma variável resposta binária, que nesse estudo é o tipo de cliente (bom e mau). A regressão logística atribui diferentes pesos para cada uma das variáveis explicativas do modelo que juntas fornecem a probabilidade de o cliente pertencer ao grupo de interesse (HOSMER; LEMESHOW, 2013).

Nas instituições financeiras é comum observar um desbalanceamento na proporção de bons e maus pagadores. Como consequência, há um prejuízo no desenvolvimento de modelos estatísticos, visto que as proporções diferentes de bons e maus pagadores podem influenciar nas variáveis preditoras do modelo, ocasionando erros de classificação. Muitos estudos já utilizaram a comparação de diferentes técnicas estatísticas para análise dos dados de crédito, porém poucos estudos avaliam o impacto do desequilíbrio amostral na correta classificação e predição do modelo (BROWN; MUES, 2012). Dessa forma, o objetivo desse estudo é identificar o impacto do desbalanceamento amostral na predição do risco de crédito por meio da utilização de um modelo de regressão logística.

## **2 Método**

O estudo utilizou uma base de dados reais de uma instituição financeira, com informações do tipo: sexo, idade, escolaridade, profissão, etc. Tendo em vista a comparação dos modelos propostos, foram criadas duas composições amostrais com diferentes percentuais de bons e maus clientes. Primeiramente, foram tomadas amostras de análise e validação na proporção real de bons e maus pagadores da instituição. No segundo momento, foram criadas amostras de análise e validação com a

mesma proporção de clientes nas amostras (50% bons e 50% maus) e uma amostra de teste com as proporções reais do negócio para comparação com o primeiro modelo.

A construção dos modelos seguiu uma adaptação da sistemática proposta por Selau (2011), cujas etapas para desenvolvimento são: delimitação da população, seleção da amostra, análise preliminar, construção dos modelos, avaliação dos modelos. Para avaliação da qualidade dos modelos, foi observado o percentual de acerto de classificação, o valor de teste de Kolmogorov-Smirnov (KS), que revela a correta separação entre os grupos de bons e maus pagadores, e a área abaixo da curva ROC, que se baseia na sensibilidade do modelo em identificar os maus pagadores e na especificidade, que é a correta identificação de bons pagadores.

### 3 Resultados

Os modelos construídos foram avaliados utilizando a área abaixo da curva ROC e o valor do teste KS. As medidas foram realizadas tanto para as amostras de análise e validação do modelo de grupos proporcionais, quanto para as amostras de análise, validação e teste do modelo de grupos iguais, com a mesma proporção de bons e maus clientes. Os resultados obtidos foram bastante semelhantes nos dois modelos construídos e em todas as amostras, como pode ser observado na Tabela 1.

Tabela1. Avaliação dos modelos – Curva ROC e Teste KS.

Avaliação dos modelos – Teste KS e Curva ROC					
Avaliação	Grupos Iguais			Grupos Proporcionais	
	Análise	Validação	Teste	Análise	Validação
ROC	0,709	0,699	0,689	0,706	0,705
KS	30,26	30,11	27,18	30,70	30,04

O ponto de corte inicial para separação dos clientes bons e maus foi de 0,5. Dessa forma, clientes com probabilidade acima de 0,5 eram classificados como bons e abaixo de 0,5 eram classificados como maus. Para avaliação dos modelos propostos, foram fixados diferentes pontos de corte para classificação dos clientes (variando de 0,5 até 0,7). A Tabela 2 apresenta a proporção de acerto nos dois modelos, tomando como base os diferentes pontos de corte para classificação dos clientes.

Tabela 2. Percentual de acerto nos modelos de grupos iguais e proporcionais de clientes

Percentual de acerto geral (%)					
Ponto de Corte	Grupos Proporcionais		Grupos Iguais		
	Análise	Validação	Análise	Validação	Teste
0,5	65,4	66,1	65,3	65,6	64,7
0,55	65,3	65,5	64,7	64,7	63,6
0,6	63,9	63,7	64,2	63,6	60,4
0,65	62,0	61,5	62,6	62,7	59,2
0,7	59,3	59,0	59,9	59,5	55,9

Após a primeira avaliação geral do percentual de acerto dos modelos construídos, realizou-se a avaliação do percentual de acerto de cada categoria de clientes (bons e maus) nos modelos de grupos proporcionais e iguais de clientes. A Tabela 3 apresenta os resultados obtidos de percentual de acerto por categoria de cliente no modelo de grupos proporcionais de clientes, por meio do uso de diferentes pontos de corte para classificação dos bons e maus.

Tabela 3. Percentual de acerto das categorias no modelo de grupos proporcionais de clientes.

Percentual de acerto por categoria (%)				
Modelo de Grupos Proporcionais				
Ponto de Corte	Análise		Validação	
	Bons	Maus	Bons	Maus
0,5	75,7	50,9	80,2	46,7
0,55	66,7	63,2	66,0	63,2
0,6	57,1	73,5	59,4	70,0
0,65	49,2	80,3	47,2	79,4
0,7	40,2	86,4	38,7	85,9

Da mesma forma, como realizou-se a avaliação das categorias de clientes bons e maus no modelo de grupos proporcionais, avaliou-se também os percentuais de classificação correta das categorias no modelo de grupos iguais de clientes, cujo resultado pode ser visto na Tabela 4.

Tabela 4. Percentual de acerto das categorias no modelo de grupos iguais de clientes

Percentual de acerto por categoria (%)						
Modelo de Grupos Iguais						
Ponto de Corte	Análise		Validação		Teste	
	Bons	Maus	Bons	Maus	Bons	Maus
0,5	60,9	69,6	61,5	69,0	58,7	67,1
0,55	53,1	76,6	53,1	77,0	50,8	76,6
0,6	44,1	83,2	44,0	82,3	42,9	83,6
0,65	35,7	88,5	36,2	86,9	34,3	88,5
0,7	27,6	92,5	28,6	90,7	25,6	92,4

## 4 Discussão e Conclusões

Com da análise dos resultados obtidos com o modelo de regressão logística, pode-se observar que os valores do teste KS e da área abaixo da curva ROC foram muito semelhantes em ambos os modelos de grupos proporcionais ou iguais, o que sugere que os modelos conseguiram separar adequadamente os grupos de bons e maus clientes da empresa.

Os percentuais de acerto geral nos dois modelos construídos foram bastante semelhantes, com valores em torno de um acerto de 60%. Por meio da avaliação do percentual geral nos diferentes

pontos de corte observou-se que à medida que o ponto de corte aumenta, ocorre um percentual de acerto geral diminuído, o que gera uma piora na previsão do modelo proposto.

Já a avaliação do percentual de acerto das categorias de clientes (bons e maus) apresentam valores bem diferenciados de acordo com o ponto de corte. Embora de forma geral os modelos apresentem um percentual de acerto por volta dos 60% para ambos os pontos de corte, a avaliação do percentual de clientes por categoria sugere resultados diferentes. Observa-se que à medida que se aumenta o ponto de corte, há uma melhora na classificação dos maus clientes, mas há uma piora no percentual de acerto dos bons clientes. Da mesma forma, observa-se que no modelo de grupos iguais, a classificação inicial dos clientes foi feita de maneira equilibrada e assim que se aumenta o ponto de corte, o modelo foi melhorando a classificação dos maus clientes e piorando a classificação dos bons. Já o modelo de grupos proporcionais iniciou classificando melhor os clientes que estavam em maior proporção na amostra, os bons clientes, e à medida que o ponto de corte aumentou, passou a classificar melhor os maus clientes.

Os resultados sugerem que a proporção das categorias de clientes da amostra para a construções do modelo, bem como os pontos de corte da classificação devem ser considerados de acordo com o objetivo da empresa. Caso haja uma necessidade de classificação melhor dos maus clientes, tendo em vista que os mesmos causam um prejuízo maior para a empresa do que o lucro obtido com bons clientes, então o ponto de corte do modelo de regressão logística deve ser maior. Nesse sentido, uma sugestão de trabalho futuro é avaliar a construção de um modelo em que se tenham grupos com proporcionalidade maior de maus clientes.

## Referências

- BRITO, G. A. S.; NETO, A. A. Modelo de Classificação de Risco de Crédito de Empresas. *Revista Contabilidade e Finanças*, São Paulo, v. 19, n. 46, p. 18-29, 2008.
- BROWN, I., MUES, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert systems with Applications*, v. 39, n. 3, p. 3446-3453, 2012.
- HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. New York: Wiley, 3.ed., 2013.
- SELAU, L. P. R.; RIBEIRO, J. L. D. Systematic Approach to Construct Credit Risk Forecast Models. *Pesquisa Operacional*, v.31, n.1, 2011.
- STEINER, M. T. A.; CARNIERI, C.; KOPITKE, B. H.; STEINER NETO, P. J. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. *Revista de Administração*, São Paulo, v. 34, n. 3, p. 56-67, 1999.