

Regressão Logística aplicada a previsão de risco de crédito

Autor: José Luis de Fraga Lassakoski Júnior

Curso: Bacharelado em Estatística, UFRGS

Orientadora: Prof^a Lisiane Selau



INTRODUÇÃO

O aumento da concessão de crédito para pessoas físicas nos últimos dez anos no Brasil, ocasionou o aumento dos índices de inadimplência. Com o avanço tecnológico, principalmente na área da informática, foi possível recorrer a métodos mais técnicos e menos subjetivos para a escolha de quais clientes terão o seu crédito concedido. Tais métodos anteriormente poderiam ser consideradas inviáveis, em função dos recursos computacionais necessários.

Nesse sentido, o presente trabalho tem por objetivo apresentar a construção de um modelo para previsão de risco de crédito. A técnica multivariada utilizada em questão é a Regressão Logística. Foi utilizado um banco de dados com informações cadastrais de clientes, que serão utilizadas para classificar os futuros clientes como bons ou maus pagadores.

METODOLOGIA

Primeiramente foi delimitada a população alvo do trabalho, seguindo pela definição de bom e mau pagador, sendo definido como bom pagador aquele que possui atraso máximo de trinta dias no pagamento do seu crédito. Identificou-se as variáveis disponíveis no banco de dados, sendo validado o preenchimento dos dados. Em seguida foi selecionada a amostra, e separada em amostra de análise, de validação e de teste. O próximo passo foi a análise preliminar, composta pela criação de variáveis *dummies* e escolha das variáveis que entraram no modelo, como pode ser visto na Tabela 1. E finalmente a execução do modelo de regressão logística em si.

RESULTADOS

Foram calculadas três medidas, bastante utilizadas na literatura especializada, para avaliar a qualidade do modelo construído. As amostras de análise, validação e teste obtiveram mais de 60% de acerto na predição de bons e maus pagadores, o valor do Teste de Kolmogorov-Smirnov foi superior a 30% e a área sob a curva ROC foi superior a 70% nas 3 amostras, como pode ser visto na Tabela 2. Tais resultados mostram que o modelo obteve resultados satisfatórios, já que os valores obtidos para as três medidas são superiores aos indicados pela literatura. Na Figura 1 pode-se observar como a taxa de sinistro decai conforme aumenta a probabilidade estimada pelo modelo.

REFERÊNCIAS

SELAU, L. P. R. ; RIBEIRO, J. L. D. . Uma sistemática para construção e escolha de modelos de previsão de risco de crédito. Gestão & Produção (UFSCAR. Impresso), v. 16, p. 398-413, 2009.

DINIZ, C; LOUZADA, F. Modelagem Estatística Para Risco de Crédito. In: Minicurso no XX SINAPE, João Pessoa-PB, 2012.

Tabela 1: Variáveis no modelo

Variáveis	Descrição
DIDAD1	Até 20 anos
DIDAD2	De 21 a 25 anos
DIDAD3	De 26 a 30 anos
DIDAD6	De 41 a 50 anos
DIDAD7	De 51 a 60 anos
DIDAD8	Acima de 60 anos
DSEXOF	Sexo feminino
DFUN	Ensino fundamental
DSUP	Ensino superior
DCASADO	Casado
DGCEPRE1	CEP residencial péssimo
DGCEPRE2	CEP residencial muito mau
DGCEPRE3	CEP residencial mau
DGCEPRE7	CEP residencial excelente
DGCEPCO5	CEP comercial bom
DGPROF1	Profissão péssimo
DGPROF5	Profissão bom
DGPROF6	Profissão muito bom
DGPROF7	Profissão excelente

Tabela 2: Medidas de qualidade do modelo

Amostra	% de acerto	Valor KS %	Área ROC %
Análise	65,2	32	71,1
Validação	65,6	31,7	70,7
Teste	63,2	31,4	70,7

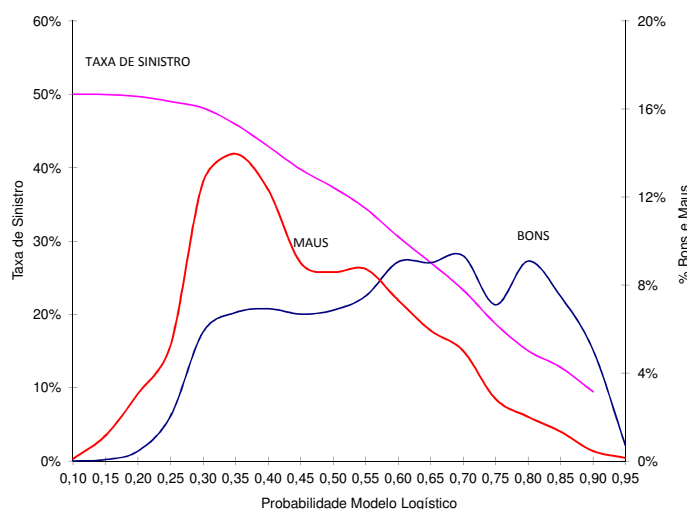


Figura 1: Taxa de Sinistro