**Atividade Final – Multivariada II**

**Profa. MSc Edmila Montezani**

**Aluno: Felipe N. S. Bezerra**

02/09/2018

1. Com base na tabela enviada (Dados\_ativ\_final201802.xlsx), elabore um modelo de regressão logística para avaliar se as variáveis *estado civil, idade* e *sexo* podem determinar a probabilidade de uma pessoa pagar ou não um empréstimo, sendo:

* Estado Civil = 0 (solteiro)
* Estado Civil = 1 (casado)
* Sexo = 0 (masculino)
* Sexo = 1 (feminino)

1. Código usado para a construção do modelo de regressão logística.

R: O código da regressão logística para este exercício pode ser consultado no arquivo “atividade\_dados.R”, em anexo.

1. Quais variáveis são significativas no modelo?

R: Estado civil e sexo mostram-se significativos na determinação da probabilidade de uma pessoa pagar ou não um empréstimo. No modelo incluindo as três variáveis independentes disponíveis, a idade não foi significativa e, após o processo de stepwise, esta variável foi excluída do modelo (provendo melhoria no ajuste do modelo, segundo a redução no critério de informação de akaike de 96,126 para 95,111).

1. Interpretação dos resultados (não esquecer de calcular o ODDS RATIO).

R: A razão de chance de um proponente pagar um empréstimo modifica-se em mais 346,5% caso seja mulher (‘sexo’ = 1) e em -97,6% caso seja casado (‘estadocivil’ = 1). Ou seja, a chance de pagamento é maior para mulheres e/ou solteiros.

A acurácia – taxa de acerto – utilizando observações aleatórias (uma parcela de 40% da amostra) foi de 59,72% e não houve evidência de superdispersão, visto que a razão entre o desvio residual e os graus de liberdade dos resíduos foi menor que 1.

1. A planilha “credito\_ativ\_final21802” apresenta os dados de default (falta de pagamento) de crédito a partir de várias variáveis distintas. Os dados referem-se a 500 pessoas que são clientes de uma financeira. Por meio de uma regressão logística pede-se:
2. Quais variáveis são significativas para se elaborar uma boa previsão de risco de default?

R: Tempo no emprego atual e outras dívidas mostram-se variáveis significativas na determinação da probabilidade de default. No modelo incluindo as quatro variáveis independentes disponíveis, a idade e o nível de educação não foram variáveis significativas e, após o processo de stepwise, estas mesmas variável foram excluída do modelo (provendo melhoria no ajuste do modelo, segundo a redução no critério de informação de akaike de 297,72 para 294,35).

1. Calcule a probabilidade de default de um indivíduo com as seguintes características:

* Idade = 40 anos
* Nível de educação = 3
* Emprego atual = 3 anos
* Endereço atual = 5 anos
* Outras dívidas (em milhares) < $30,00

R: A probabilidade de default de um indivíduo com as características supracitadas é de 99,95119%.

1. Interprete os resultados e anexe/envie o código utilizado para a construção do modelo.

R: A razão de chance de default de um indivíduo reduz-se em 17,36% para cada ano no emprego atual e eleva-se em 34,02% a cada $1.000,00 em outras dívidas. Ou seja, a chance de default tende a ser maior para maiores somatórias de outras dívidas e menor para mais anos de permanência no último emprego.

A acurácia – taxa de acerto – utilizando observações aleatórias (uma parcela de 40% da amostra) foi de 75,5% e não houve evidência de superdispersão, visto que a razão entre o desvio residual e os graus de liberdade dos resíduos foi menor que 1.

O código da regressão logística para este exercício pode ser consultado no arquivo “atividade\_credito.R”, em anexo.