

### **Atividade Prática no RStudio n. 2 – Valor: 5 pontos**

#### **Providências preliminares.**

1. Importe a biblioteca “MASS”.
2. Importe a biblioteca “AER”.
3. Carregue a base de dados “Boston”. Para tanto basta acionar o comando `data("Boston")` após já ter importado a biblioteca “MASS”.

#### **Descrição do caso.**

Você foi contratado por uma imobiliária de Boston para desenvolver modelos econométricos que expliquem o valor mediano das em uma vizinhança (variável “medv”) através do percentual de moradores de baixo status econômico (variável “lstat”). Para tanto, realize as tarefas abaixo e produza um arquivo RMarkdown que gere um relatório que apresente os resultados alcançados.

1. Desenvolva um modelo linear (“mod”) com intercepto em que a variável dependente seja “medv” e a variável independente seja “lstat”. Apresente a tabela com o resumo (*summary*) desse modelo.
2. Apresente um gráfico de dispersão no qual o eixo horizontal é a variável “lstat” e o eixo vertical é a variável “medv”. Nesse mesmo gráfico, plote a linha da regressão linear do modelo “mod” na cor vermelha.
3. Repare, ao analisar o gráfico, que aparentemente existe uma relação não-linear entre a variável “lstat” e “medv” que não está sendo captada pelo modelo. Desenvolva um outro modelo linear (“log\_mod”) com intercepto em que a variável dependente seja “medv” e a variável independente seja o **logaritmo natural** de “lstat”. Apresente a tabela com o resumo (*summary*) desse modelo.
4. Apresente um gráfico de dispersão no qual o eixo horizontal é o **logaritmo natural** da variável “lstat” e o eixo vertical é a variável “medv”. Nesse mesmo gráfico, plote a linha da regressão linear do modelo “log\_mod” na cor vermelha.
5. Repare que a transformação logarítmica melhorou a qualidade de ajuste do modelo (aumento do  $R^2$ ). Mas ainda é possível melhorar mais. Desenvolva um outro modelo linear (“log2\_mod”) com intercepto que possui como variáveis explicativas tanto o logaritmo da variável “lstat” quanto o quadrado do logaritmo da variável “lstat”. Apresente a tabela com o resumo (*summary*) desse modelo. Dica: pesquise sobre a função `I()`.
6. Como o modelo “log2\_mod” está na forma logarítmica e quadrática, não se pode simplesmente utilizar os coeficientes para interpretar o efeito de uma mudança unitária da

variável “lstat” sobre a variável “medv” porque esse efeito depende do nível de “lstat”. Portanto, caso se queira mensurar o efeito estimado da variação de uma variável sobre a outra é necessário testar valores. Dessa forma, utilize o R para calcular qual é a variação em “medv” quando “lstat” varia de 10 para 11 (ou seja, quando o percentual de moradores de baixo nível econômico aumenta de 10% para 11%).

7. Considere o modelo de regressão linear abaixo:

$$medv_i = \beta_0 + \beta_1 \times chas_i + \beta_2 \times old_i + \beta_3 \times (chas_i \cdot old_i) + u_i$$

Utilize o comando “?Boston” para ter uma descrição da variável “chas”. A variável “old”, por sua vez, ainda não está na base de dados Boston e precisa ser incluída por você. A variável “old” deve ser construída como uma variável *dummy* que assume o valor 1 se variável “age” (que já existe na base de dados Boston) for igual ou maior que 95, e zero caso contrário. Rode a regressão apresentada acima e a nomeie como “mod\_co”. Apresente a tabela com o resumo (*summary*) desse modelo. Dicas: o operador “>=” pode ser utilizado para criar um vetor lógico; a função “as.numeric” pode ser utilizada para transformar um vetor lógico em um vetor numérico;

8. Considerando o modelo “mod\_co” desenvolvido na tarefa anterior, responda: qual é o efeito estimado sobre a variável “medv” ao se mudar de uma vizinhança com menos de 95% de prédios antigos para uma vizinhança com pelo menos 95% de prédios antigos?
9. Considere o modelo de regressão linear abaixo:

$$medv_i = \beta_0 + \beta_1 \times indus_i + \beta_2 \times old_i + \beta_3 \times (indus_i \cdot old_i) + u_i$$

A variável “old” já foi definida e incluída na base de dados por você anteriormente. A variável “indus” já está na base de dados (utilize o comando “?Boston” para obter a descrição de variável). Rode a regressão apresentada acima e a nomeie como “mod\_io”. Apresente a tabela com o resumo (*summary*) desse modelo.

10. Considerando o modelo “mod\_io” desenvolvido na tarefa anterior, qual é o efeito esperado sobre “medv” de um aumento de uma unidade na variável “indus”?