

Universidade Federal do Paraná - Departamento de Estatística

CE310 - Modelos de Regressão Linear

Prof. Cesar Augusto Taconeli

Lista de exercícios - Diagnóstico do ajuste

Nota: Os exercícios desta lista foram extraídos do livro **Linear models with R**, de Julian J. Faraway. Todas as bases de dados estão disponíveis no pacote **faraway** do R. Antes de iniciar qualquer análise, consulte a documentação da base, verifique o contexto dos dados e a descrição das variáveis. Adicionalmente, todas as questões devem ser precedidas por uma análise descritiva/exploratória, composta por gráficos e medidas descritivas pertinentes.

Exercício 1- Usando o conjunto de dados **sat**, ajuste um modelo de regressão linear múltipla com o escore SAT total como resposta e as variáveis **expend**, **salary**, **ratio** e **takers** como explicativas. Conduza o diagnóstico do ajuste e responda às questões apresentadas. Utilize os gráficos que julgar relevantes. Apresente possíveis melhorias ou medidas corretivas para o modelo, quando apropriado.

- a) Cheque a suposição de variância constante para os erros;
- b) Cheque a suposição de normalidade;
- c) Cheque a existência de pontos de alavanca;
- d) Cheque se há outliers;
- e) Cheque a existência de pontos influentes;
- f) Cheque a estrutura da relação entre a resposta e as variáveis explicativas.

Exercício 2- Semelhante ao exercício 1, mas desta vez para o conjunto de dados **prostate**, tomando a variável **lpsa** como resposta e as demais variáveis como explicativas.

Exercício 3- Semelhante ao exercício 1, mas desta vez para o conjunto de dados **happy**, tomando a variável **happy** como resposta e as demais variáveis como explicativas.

Exercício 4- Semelhante ao exercício 1, mas desta vez para o conjunto de dados **divusa**, tomando a variável **divorce** como resposta e as demais variáveis como explicativas (exceto **year**). Adicionalmente:

- a) Analise possível autocorrelação temporal nos resíduos;
- b) Obtenha os valores dos VIFs e analise as correlações entre as variáveis. Há evidência de que colinearidade é a justificativa para algumas variáveis explicativas serem não significativas? Justifique.
- c) Investigue se a remoção de variáveis explicativas não significativas do modelo reduz a colinearidade.

Exercício 5- Considere o conjunto de dados **fat** e o ajuste do modelo de regressão linear múltipla da variável resposta **brozek** em função de **age**, **weight**, **height**, **neck**, **chest**, **hip**, **abdom**, **thigh**, **ankle**, **biceps**, **forearm** e **wrist**.

- a) Investigue possível colinearidade nos dados;
- b) As observações 39 e 42 são atípicas. Reajuste o modelo sem essas observações. Comente as diferenças observadas para o modelo ajustado com todos os dados;
- c) Ajuste o modelo com **brozek** como resposta e **age**, **weight** e **height** como preditores. Faça a análise de colinearidade e compare os resultados aos produzidos pelo modelo original;
- d) Obtenha o intervalo de predição 95% para **brozek** considerando os valores medianos para **age**, **weight** e **height**;
- e) Obtenha o intervalo de predição 95% para **brozek** para **age**=40, **weight**=200 e **height**=73. Como esse intervalo se compara ao obtido no item anterior?

- f) Obtenha o intervalo de predição 95% para `brozek` para `age=40`, `weight=130` e `height=73`. Esses valores para os preditores são atípicos? Compare os intervalo obtido aos produzidos nos itens anteriores.