

Universidade Federal do Paraná - Departamento de Estatística

CE310 - Modelos de Regressão Linear

Prof. Cesar Augusto Taconeli

Lista de exercícios - Medidas corretivas

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- Use o conjunto de dados **cheddar** para essa questão.

- Ajuste o modelo de regressão linear múltipla com **taste** como resposta e as outras três variáveis como explicativas. Analise os resíduos parciais e verifique a necessidade de transformação nas variáveis explicativas;
- Use o método de Box-Cox para identificar uma transformação ótima para a resposta. Seria razoável manter a resposta na escala original?
- Ajuste novamente o modelo do item (a), agora com a resposta transformada. Analise novamente os resíduos parciais. Houve alguma mudança nos resultados?

Exercício 2- Pesquisadores do NIST (National Institutes of Standards and Technology) coletaram dados referentes à profundidade de defeitos avaliados em diferentes pontos de um oleoduto no Alasca primeiramente no campo e, num segundo momento, num laboratório. As medidas de profundidade obtidas no laboratório são mais acuradas que aquelas obtidas no campo, mas são mais caras e demandam mais tempo. Deseja-se ajustar uma equação de regressão para corrigir as medidas obtidas no campo. Os dados estão disponíveis no conjunto de dados **pipeline** do pacote **faraway**.

- Ajuste um modelo de regressão para **Lab~Campo**. Cheque a homogeneidade das variâncias;
- Vamos agora usar pesos para acomodar a variância não constante. Divida os dados originais em 12 grupos de tamanho nove (exceto para o oitavo grupo, que terá apenas oito observações), com base nos valores ordenados das medidas tomadas no campo. Dentro de cada grupo, calcule a média da medida de campo (**medcampo**) e a variância das medidas tomadas no laboratório (**varlab**).

Suponha que a variância da resposta seja relacionada à medida de campo da seguinte forma:

$$\text{Var}(\text{Lab}) = \alpha_0 \text{Campo}^{\alpha_1}$$

Ajuste a regressão de $\log(\text{varlab})$ em função de $\log(\text{medcampo})$. Use os resultados desse ajuste para determinar pesos apropriados para estimação por mínimos quadrados ponderados no ajuste de **Lab~Campo**. Você deverá usar como pesos os inversos dos valores ajustados para **varlab**. Analise o sumário da regressão, e compare com o sumário do ajuste obtido no item anterior, em que não se usa ponderação.

- Uma alternativa à ponderação é transformar os dados. Encontre a transformação de Lab e/ou Campo tal que na escala transformada a relação seja linear com variância aproximadamente constante. Você pode restringir sua escolha às transformações raiz quadrada, logarítmica e inversa.

Exercício 3- Usando o conjunto de dados **stackloss**, ajuste o modelo de regressão linear múltipla para **stack.loss** em função das demais variáveis usando os seguintes métodos:

- Mínimos quadrados;

- b) Least absolute deviations;
- c) Huber method;
- d) Bisquare method;
- e) Least trimmed squares.

Compare os resultados. Voltando ao ajuste por mínimos quadrados, faça o diagnóstico do ajuste e detect outliers e pontos influentes. Remova essas observações e use novamente mínimos quadrados. Compare os resultados.