

4ª Lista de Exercícios - SME-260 Análise de Regressão

Profª Cibele Russo - Entregar exercícios com * no dia 30/05/2011

1.* Considere os dados de consumo de combustível em estados norte-americanos, disponíveis no arquivo `combustível.txt` (WIKI), com

- Y : consumo de combustível (galões por pessoa),
 - X_1 o imposto sobre combustível (centavos de dólar/galão),
 - X_2 o percentual da população com carteira de habilitação,
 - X_3 a renda per capita (milhares de dólares) e
 - X_4 a quantidade de estradas federais pavimentadas (em milhares de milhas).
- (a) Ajuste um modelo de regressão linear múltipla, relacionando o consumo de combustível com as 4 covariáveis.
- (b) Avalie a significância da regressão.
- (c) Obtenha o coeficiente de determinação do modelo R^2 .
- (d) Implemente medidas diagnósticas em R e verifique a presença de pontos aberrantes, de alavanca ou influentes. Se houver observações com essas características, identifique-as e verifique possíveis motivos para essa classificação. (Referências: C.13.1 de Montgomery, Peck and Vining (2006) pág 572-573 ou cap 10 de Kutner et al, 2005, pág 384)
- (e) Considerando o método passo atrás (backward), obtenha o melhor modelo (Dica: No R, use `stepAIC` (MASS)).

2. Considere o modelo de regressão linear múltipla,

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

com $\mathbf{Y}_{n \times 1}$ um vetor aleatório, $X_{n \times (p+1)}$ uma matriz conhecida de covariáveis, $\boldsymbol{\beta}_{(p+1) \times 1}$ um vetor de parâmetros desconhecidos e $\boldsymbol{\epsilon}$ um vetor de erros com distribuição normal multivariada $N_n(\mathbf{0}, \sigma^2 I_n)$ e \mathbf{e} o vetor de resíduos do modelo ajustado. Mostre que

- (a) $r(\mathbf{e}, \mathbf{Y}) = \sqrt{1 - R^2}$, com r o coeficiente de correlação amostral. Conclua que não é adequado utilizar o gráfico de $\mathbf{e} \times \mathbf{Y}$ na análise de resíduos.
- (b) Mostre que o coeficiente angular do ajuste de mínimos quadrados de \mathbf{e} em \mathbf{Y} é $(1 - R^2)$.
- (c) Mostre que $r(\mathbf{e}, \hat{\mathbf{Y}}) = 0$.

Outros exercícios sugeridos:

- Draper and Smith (1998) pág 215 ex. A, D; pág 356 ex. B;
- Montgomery, Peck and Vining (2006) pág 200 ex. 6.1; pág 232-235 ex. 7.4, 7.6, 7.16
- Montgomery, Peck and Vining (2006): Ler seção C.7 pág 562-564 (estatística PRESS)
- Montgomery, Peck and Vining (2006): Ler seção C.10 pág 568-569 independência entre resíduos e valores ajustados
- Kutner, Nachtstheim, Neter and Li (2005): pág 252-253 ex. 6.24, pág 415-416 ex. 10.1, 10.3, 10.9, 10.23, 10.27

Referências:

Draper and Smith (1998). Applied Regression Analysis. Wiley.
Montgomery, Peck and Vining (2006). Introduction to Linear Regression Analysis. Wiley.
Kutner, Nachtstheim, Neter and Li (2005). Applied Linear Statistical Models. McGraw Hill