4ª Lista de Exercícios - SME0820 Análise de Regressão - 03/06/2014

Exercício 1. Considere os dados de consumo de combustível em estados norte-americanos, disponíveis no arquivo `combustível.txt`, 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 diagnóstico 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.
(e) Considerando o método passo atrás (backward), obtenha o melhor modelo.

Exercício 2. Considere o modelo de regressão linear múltipla,

$$\underline{\mathsf{Y}} = X\underline{\beta} + \underline{\epsilon},$$

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

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

Exercício 3. Considere o modelo de regressão linear simples sem intercepto e com erros heteroscedásticos e não correlacionados,

$$(1) \qquad \begin{cases} Y_i = \beta X_i + \epsilon_i,\ i = 1, \dots, n, \text{ com} \\ \mathsf{E}(\epsilon_i) = 0,\ \mathsf{Var}(\epsilon_i) = \sigma_i^2,\ \sigma_i^2 = (a + bX_i^2)\sigma^2, \end{cases}$$

com a e b constantes conhecidas.

(a) Interprete o modelo (1).
(b) Obtenha o estimador de mínimos quadrados de β para o caso em que $a = 0$ e $b = 1$. Verifique se esse estimador é não viesado e calcule sua variância.
(c) Repita o item (b) para o caso em que $a = 0$ e $b = \sigma^{-2}$.
(d) Interprete o modelo para o caso em que $a \neq 0$ e $b = 1$.

Exercício 4. Considere o modelo de regressão linear múltipla,

$$\underline{\mathsf{Y}} = X\underline{\beta} + \underline{\epsilon},$$

com $\underline{\mathsf{Y}}_{n\times 1}$ um vetor de variáveis resposta, $X_{n\times(p+1)}$ uma matriz conhecida de covariáveis, $\underline{\beta}_{(p+1)\times 1}$ um vetor de parâmetros desconhecidos e $\underline{\epsilon}$ um vetor de erros com $\mathsf{E}(\underline{\epsilon})$ e $\mathsf{Var}(\underline{\epsilon}) = \sigma^2 I$, V uma matriz diagonal conhecida, positiva definida e diferente da identidade.

(a) Verifique se
$$\frac{\underline{Y}'V^{-1}\underline{Y} - \underline{Y}'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\underline{Y}}{n-p-1}$$
é um estimador não viesado de σ^2. (Dica: escreva a expressão como uma forma quadrática).
(b) Obtenha o resultado particular análogo para o modelo de regressão linear simples.

Exercício 5. Considere o modelo de regressão linear simples sem intercepto e com erros heteroscedásticos e não correlacionados, $Y_i = \beta X_i + \epsilon_i,\ i = 1, \ldots, n$, com $\mathsf{E}(\epsilon_i) = 0$, $\mathsf{Var}(\epsilon_i) = \sigma_i^2$ e $\sigma_i^2 = X_i^2\sigma^2$.
(a) Apresente uma transformação que estabiliza a variância dos erros. Justifique.
(b) Obtenha um estimador não viesado para β e calcule sua variância.
(c) Verifique se $\dfrac{1}{n-1}\left[\displaystyle\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right)^2 - \left(\sum_{i=1}^{n}\frac{Y_i}{X_i}\right)^2 \Big/ n\right]$ é um estimador não viesado de σ^2.

Exercício 6. Considere o modelo de regressão linear simples
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \text{ com } \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \text{ para } i = 1, \ldots, n.$$
(a) Mostre que os elementos da matriz chapéu (*hat matrix*) são dados por
$$h_{ij} = \frac{1}{n} + \frac{(X_i - \overline{X})(X_j - \overline{X})}{S_{XX}} \text{ e } h_{ii} = \frac{1}{n} + \frac{(X_i - \overline{X})^2}{S_{XX}}$$
(b) Discuta o comportamento de h_{ii} quando X_i está distante de $\overline{X}$.

Exercício 7. Explique com suas palavras o que significa análise de diagnóstico. Você deve incluir na discussão análises de resíduos, identificação de pontos aberrantes, influentes e de alavanca. Dê exemplos.

Exercício 8. Explique o que é multicolinearidade. Dê exemplos e apresente possíveis soluções.

Exercício 9. Explique a utilidade e necessidade da seleção de modelos.

Exercício 10. Para cada um dos modelos abaixo, verifique se corresponde a um modelo linear ou se existe uma transformação conveniente que o linearize.
(a) $Y = \beta_0 + \beta_1 X_1 + \beta_2 \log_{10} X_2 + \beta_3 X_1^2 + \epsilon$
(b) $Y = \epsilon(e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2^2})$
(c) $Y = \beta_0 e^{\beta_1 X_1} + \epsilon$
(d) $Y = [1 + e^{\beta_0 + \beta_1 X_1 + \epsilon}]^{-1}$