2ª Lista de Exercícios - SME0812 Modelos Lineares - 1º/2014

1 Considere o modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ com $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$ e $\boldsymbol{X}$ de posto completo $p < n$. Se $\widehat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, prove que

(i) $\text{Cov}(\widehat{\boldsymbol{\epsilon}}, \widehat{\boldsymbol{\beta}}) = \boldsymbol{0}$,

(ii) $\text{Cov}(\boldsymbol{\epsilon}, \widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$,

(iii) $\text{Cov}(\boldsymbol{\epsilon}, \boldsymbol{Y}) = \sigma^2[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$ e

(iv) $\text{Cov}(\boldsymbol{\epsilon}, \widehat{\boldsymbol{Y}}) = \boldsymbol{0}$.

Com base nos itens (iii) e (iv), discuta por que é preferível a análise gráfica de $\boldsymbol{\epsilon}$ *versus* $\widehat{\boldsymbol{Y}}$ em relação à análise gráfica de $\boldsymbol{\epsilon}$ *versus* $\boldsymbol{Y}$.

2 Seja $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ com $\boldsymbol{X}$ de posto completo $p < n$. Prove que

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi^2_p.$$

3 Considere um modelo linear geral $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ e as partições

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} \text{ com } \boldsymbol{\beta}_1 = (\beta_1, \beta_2, \ldots, \beta_p)' \text{ e } \boldsymbol{X} = [\boldsymbol{1}_n \;\; \boldsymbol{X}_1].$$

Considere $\boldsymbol{S} = \boldsymbol{X}_1'\boldsymbol{X}_1 - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{x}}'$ e $\overline{\boldsymbol{x}} = (\overline{x}_2, \overline{x}_3, \ldots, \overline{x}_p)'$, em que $\overline{x}_j$ é a média dos elementos da j-ésima coluna de $\boldsymbol{X}$, $j = 2, \ldots, p$. Prove que

(i) $\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{S}^{-1}\left(\boldsymbol{X}_1'\boldsymbol{Y} - \frac{1}{n}\boldsymbol{X}_1'\boldsymbol{1}_n\boldsymbol{1}_n'\boldsymbol{Y}\right) = \boldsymbol{S}^{-1}(\boldsymbol{X}_1'\boldsymbol{Y} - n\overline{Y}\,\overline{\boldsymbol{x}}')$ e

(ii) $\widehat{\beta}_0 = \overline{Y} - \overline{\boldsymbol{x}}'\boldsymbol{S}^{-1}(\boldsymbol{X}_1'\boldsymbol{Y} - n\overline{Y}\overline{\boldsymbol{x}}) = \overline{Y} - \overline{\boldsymbol{x}}'\widehat{\boldsymbol{\beta}}_1$.

4 No modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, com $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$, determine a distribuição do vetor de resíduos $\widehat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.

5 Para o modelo $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$ independentes, $i = 1, \ldots, 20$, em que $\sum_{i=1}^{20} y_i^2 = 90$ e o sistema de equações normais é

$$\begin{bmatrix} 20 & 10 \\ 10 & 6 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 40 \\ 22 \end{pmatrix}.$$

(a) Calcule $\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}$ e SQRes$= \boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}$.

(b) Construa um intervalo de confiança com coeficiente de confiança 0,95 para $E(Y|x = 3)$.

(c) Construa o correspondente intervalo de predição.

6 Sejam $Y_1 = \theta + \epsilon_1$, $Y_2 = 2\theta - \gamma + \epsilon_2$ e $Y_3 = \theta + 2\gamma + \epsilon_3$, em que $E(\epsilon_i) = 0$, $i = 1, 2, 3$. Determine os estimadores de mínimos quadrados de θ e γ.

7 Prove que o coeficiente de determinação, R^2, em um modelo de regressão linear simples é o quadrado do coeficiente de correlação entre $\boldsymbol{Y}$ e $\widehat{\boldsymbol{Y}}$.

8 No modelo de regressão linear simples com as suposições usuais, determine o estimador de máxima verossimilhança da inclinação β_2 e sob a hipótese $H_0 : \beta_1 = 2$, em que β_1 denota o intercepto.

9 Num problema de estocar sorvetes a baixa temperatura, verifica-se que a perda de sorvete Y está relacionada com o tempo de estocagem através do modelo $Y = \beta t + \epsilon$. Em um experimento foi registrada a perda de sorvete (Y, em cm^3) em função do tempo (t, em semanas), com os seguintes resultados:

t	1	2	3	4	5	6	7	8
y	2,1	2,81	3,04	3,1	6,24	8,01	5,79	8,38

Suponha que $\epsilon_i \sim N(0, \sigma^2)$, independentes, $i = 1, \ldots, 8$.

(a) Estime β e σ^2.

(b) Apresente estimativas intervalares com base em intervalos de 95% de confiança para β e σ^2.

(c) Teste, ao nível de significância 0,05, as hipóteses $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$.

10 Mostre que $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, com $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, e conclua que $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$ é minimizada para $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$.

11 Suponha que, em um modelo linear geral com intercepto, multiplicamos todos os valores das variáveis independentes de modo que $w_{ij} = k_j x_{ij}$, $k_j \neq 0$, para todo i e j, resultando em uma matriz modelo $\boldsymbol{W}$. Prove que $\widehat{\boldsymbol{Y}}$ permanece inalterado.

12 Decidiu-se ajustar o modelo $Y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$, sendo que os erros são variáveis aleatórias independentes com distribuição $N(0, \sigma^2)$. Com base nos dados

x_1	1	1	1	2	2	2	0	0
x_2	1	-1	0	1	0	-1	0	1
y	8,5	3	5,5	12,5	10	7	2	5

(a) Obtenha o modelo ajustado $\widehat{Y} = \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$.

(b) Prove que, nesse caso, os estimadores $\widehat{\beta}_1$ e $\widehat{\beta}_2$ são variáveis aleatórias independentes e determine suas distribuições marginais.

(c) Construa a tabela de análise de variância e teste ao nível de significância de 0,05 a hipótese $H_0 : \beta_1 = \beta_2 = 0$ contra $H_1 :$ pelo menos um $\beta_i \neq 0$, $i = 1, 2$.