MI 406 - Regressão
Segundo semestre de 2014
Lista de Exercícios I
Data da entrega: até o dia 22/09/2014, no começo ou no final da aula
Exercícios selecionados para a entrega: 3, 4, 6 (apresente o teste somente para a questão 2), 9, 10.

1. Resolva TODOS os exercícios deixados em sala.

2. Considere o modelo $Y_i = \beta_0 + \beta_1 x_i + \xi_i, i = 1, 2, ..., n$, em que $\xi_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Defina $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, em que $\widehat{\beta}_j, j = 0, 1$ é o estimador de MQO de $\beta_j, j = 0, 1$ e $\widehat{\xi}_i = Y_i - \widehat{Y}_i$. Prove que:

    a) Nesse caso, os estimadores de MV coincidem com os estimadores de MQO.
    b) $\sum_{i=1}^n \widehat{\xi}_i = 0$.
    c) $\sum_{i=1}^n x_i \widehat{\xi}_i = 0$.
    d) $\sum_{i=1}^n \widehat{Y}_i \widehat{\xi}_i = 0$.

3. Considere o modelo $Y_i = \beta_0 + \beta_1(x_i - \overline{x}) + \xi_i, i = 1, 2, ..., n$, em que $\xi_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Encontre a forma escalar dos estimadores de mínimos quadrados ordinários (MQO) de $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ e as respectivas distribuições marginais. Compare esses estimadores com aqueles obtibos no modelo $Y_i = \alpha_0 + \alpha_1 x_i + \xi_i$. Calcule $\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1)$ e $\mathrm{Cov}(\widehat{\alpha}_0, \widehat{\alpha}_1)$. O que você pode dizer sobre a independência entre $\widehat{\beta}_0$ e $\widehat{\beta}_1$ e entre $\widehat{\alpha}_0$ e $\widehat{\alpha}_1$?

4. Considere o modelo $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \xi_i, i = 1, 2, ..., n$, em que $\xi_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Encontre a forma escalar dos estimadores de MQO de $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ e as respectivas distribuições marginais. Calcule também $\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1)$, $\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_2)$ e $\mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2)$ e comente sobre a independência entre os estimadores 2 a 2.

5. Refaça a questão 4 considerando $x_i = x_i - \overline{x}$.

6. Nas questões de 2 a 5, proponha um teste para testar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, em que θ_0 é um valor conhecido e θ reresenta cada um dos parâmetros do modelo em questão. Por exemplo, no modelo da questão 2), θ corresponde à β_0 ou β_1. Considere σ^2 conhecido. Propor um teste significa: constuir uma estatística do teste que faça sentido em termos do problema, apresentar sua distribuição (exata ou assintótica) sob H_0, apresentar as regiões crítica e de aceitação, bem como o p-valor associado. Considere um nível de significância de α.

7. Seja $\boldsymbol{A}_{(m \times m)}$ uma matriz real e simétrica (sugestão: utilize a decomposição apresentada na última página dos slides sobre álgebra de matrizes). Mostre que:

    a) $tr(\boldsymbol{A}) = \sum_{i=1}^m \lambda_i$, em que $\lambda_i, i = 1, ..., m$ são os autovalores de $\boldsymbol{A}$.
    b) $det(\boldsymbol{A}) = \prod_{i=1}^m \lambda_i$.
    c) Se $\boldsymbol{P}$ for uma matriz ortogonal, os autovalores de $\boldsymbol{PAP'}$ são os mesmos de $\boldsymbol{A}$.

8. Seja $\boldsymbol{X} = (X_1, X_2, X_3)'$ uma vetor aleatório tal que

$$Cov(\boldsymbol{X}) = \begin{bmatrix} 7 & 1 & 3 \\ 1 & 3 & 0 \\ 3 & 0 & 5 \end{bmatrix}$$

Responda os itens:

a) Calcule a variância de $X_1 - 3X_2 + 2X_3$.

b) Calcule $Cov(\boldsymbol{Z})$ em que $\boldsymbol{Z} = (Z_1, Z_2)$ com $Z_1 = X_1 + X_2$ e $Z_2 = X_1 + X_2 + X_3$.

9. Os dados contidos no arquivo Braga1998.xls (fonte: http://www.ime.usp.br/ jmsinger/doku.php?id=start) são provenientes de um estudo na área de Cardiologia. O objetivo é comparar as curvas de variação do consumo de oxigênio no ponto de compensação respiratório (PCR) em função da carga utilizada na esteira ergométrica para para pacientes com diferentes etiologias cardíacas. Responda os itens:

a) Especifique (matricial e escalarmente) um modelo linear que permita comparar as curvas de variação do consumo de oxigênio no ponto de compensação respiratório (PCR) em função da carga utilizada na esteira ergométrica para para pacientes num primeiro momento, sem considerar as etiologias cardíacas. Interprete os parâmetros. Admita que não é possível, no PCR, ter-se pacientes submetidos à uma carga nula.

b) Repita o item a) considerando as diferentes etiologias cardíacas.

c) Especifique (escalar e matricialmente) pelo menos duas hipóteses de interesse, para os itens a) e b), interpretando-as em termos do problema.

d) Faça uma análise descritiva dos dados.

10. Considere o modelo de regressão linear normal homocedástico em sua forma matricial, visto em sala. Defina $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, $\boldsymbol{R} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ e $\widehat{\boldsymbol{\theta}}_i = \boldsymbol{C}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{M}_i$ em que $\boldsymbol{C}_{i(a_i \times p)}$ são matrizes não-aleatórias tais que $r(\boldsymbol{C}_i) = a_i < p$ e $\boldsymbol{M}_{i(a_i \times 1)}$ são vetores não aleatórios, $i = 1, 2...$. Obtenha as distribuições conjuntas de i) $\widehat{\boldsymbol{\beta}}$ ii) $\widehat{\boldsymbol{\theta}}_i$ iii) $(\widehat{\boldsymbol{Y}}, \boldsymbol{R})'$, iv) $(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)$ e v) $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)$.