MI427/ME913 - Análise de dados Hierárquicos
Segundo semestre de 2020
Lista de Exercícios I

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 Cov$(\widehat{\beta}_0, \widehat{\beta}_1)$ e 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. Nas questões de 2 e 3, 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. 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. Considere σ^2 conhecido e depois desconhecido.

5. Seja $\boldsymbol{A}_{(m \times m)}$ uma matriz real e simétrica (sugestão: utilize a decomposição espectral). 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 $\boldsymbol{A}$.

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

$$Cov(\boldsymbol{X}) = \begin{bmatrix} 11 & 2 & -4 \\ 2 & 0,5 & 0 \\ -4 & 0 & 9 \end{bmatrix}$$

Responda os itens:

a) Calcule a variância de $X_1 - 4X_2 + 3X_3$.

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

7. Considere o modelo de regressão linear normal homocedástico usual em sua forma matricial, ou seja, $\boldsymbol{Y}_{(n\times 1)} = \boldsymbol{X}_{(n\times p)}\boldsymbol{\beta}_{(p\times 1)} + \boldsymbol{\xi}_{(n\times 1)}$, em que $\boldsymbol{\xi} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. 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) $(\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\theta}}_1')'$.

8. Se $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\boldsymbol{Y} = \boldsymbol{A}_{(q\times p)}\boldsymbol{X} + \boldsymbol{B}_{(q\times 1)}$, em que $\boldsymbol{A}$, $\boldsymbol{B}$ são matrizes não aleatórias, obtenha (demonstrando) a distribuição de $\boldsymbol{Y}$.

9. Considere que $\boldsymbol{Y}_{(n\times 1)} = \boldsymbol{X}_{(n\times p)}\boldsymbol{\beta}_{(p\times 1)} + \boldsymbol{\xi}_{(n\times 1)}$ e $\boldsymbol{\beta}_{(p\times 1)} = \boldsymbol{W}_{(p\times q)}\boldsymbol{\gamma}_{(q\times 1)} + \boldsymbol{u}_{(p\times 1)}$, em que $\boldsymbol{\xi} \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{u} \sim N_p(\boldsymbol{0}, \boldsymbol{\Psi})$, $\boldsymbol{\xi} \perp \boldsymbol{u}$, $\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{\gamma}$ são quantidades não aleatórias. Responda os itens:

a) Obtenha, demonstrando, a distribuição condicional $\boldsymbol{Y}|\boldsymbol{u}$.

b) Obtenha, demonstrando, a distribuição marginal de $\boldsymbol{Y}$.

c) Obtenha, demonstrando, a distribuição condicional $\boldsymbol{\beta}|\boldsymbol{Y} = \boldsymbol{y}$.

10. Considere o seguinte modelo hierárquicos de dois níveis:

$$\begin{aligned}
Y_{ji} &= \beta_{0j} + \beta_{1j}x_{ji} + \xi_{ji}, \text{(nível 1)}; j = 1, 2, ..., 50; i = 1, 2, ..., n_j \\
\beta_{0j} &= \gamma_{00} + u_{0j}, \text{(nível 2)}, \\
\beta_{1j} &= \gamma_{10} + u_{1j}, \text{(nível 2)}, \\
\xi_{ji} &\overset{iid}{\sim} N(0, \sigma^2), u_{ij} \overset{iid}{\sim} N(0, \psi_i), \xi_{ji} \perp u_{ij}, \forall i, j
\end{aligned}$$

Responda os itens:

a) Obtenha as esperanças marginal e condicional (dado $\boldsymbol{u}_j = (u_{0j}, u_{1j})'$) de Y_{ji}.

b) Obtenha as variâncias marginal e condicional (dado $\boldsymbol{u}_j = (u_{0j}, u_{1j})'$) de Y_{ji}.

c) Obtenha $Cov(Y_{ji}, Y_{ji'})$ e $Corre(Y_{ji}, Y_{ji'})$, $\forall j$ e $\forall i \neq i'$.

d) Obtenha a distribuição marginal de $\boldsymbol{Y}_j = (Y_{j1}, ..., Y_{jn_j})'$.

e) Interprete todos os parâmetros (efeitos fixos, aleatórios e componentes de variância).

11. Considere o conjunto de dados "tutorial" disponível no pacote do R "R2MLwiN" (veja a descrição das variáveis no respectivo help (?tutorial). Considere as variáveis "normexam"(resposta), "school" (explicativa, grupo) e "standlrt" (explicativa). O objetivo é comparar o desempenho (aos 16 anos, normexam) dos alunos entre as escolas (school), considerando a variável standlrt (desempenho aos 11 anos). Para esta questão, responda os itens abaixo considerando apenas as duas primeiras variáveis.

   a) Faça uma análise descritiva completa, apresentado os comentários apropriados. O que os resultados sugerem?

   b) Defina (intepretando-o adequamente) e ajuste um MRLNH apropriado, apresentando todos resultados numéricos (com as devidas interpretações). O que as conclusões indicam? Lembre-se de que a verificação da qualidade de ajuste do modelo é a etapa imediamente posterior ao ajuste. Comente sobre o ajuste, indicando, também com base na descrição do problema e análise descritiva, se você entende que um outro modelo poderia ser mais apropriado do que este.

   c) Com base na resposta do item c), defina, intepretando-o adequamente, pelo menos um modelo que você entende ser mais apropriado.

12. Repita a questão 11), considerando as três variáveis, de forma adequada.