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

1. Resolva TODOS os exercícios deixados em sala.

2. Considere o modelo linear hierárquico (normal/normal, de dois níveis) (e sua respectiva correspondendência em termos de um modelo misto), conforme visto em sala de aula, e a decomposição da matriz de covariâncias utilizada pela função "lme" (veja aqui, slides 42 e 43). Responda os itens (lembre-se de que, neste caso, trabalhamos com a verossimilhança (original ou residual) marginal):

   a) Obtenha as expressões dos estimadores de MV (MVR) para $\boldsymbol{\beta}$ e σ^2.
   b) Apresente a função escore para $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\varrho}', \boldsymbol{\theta}_D)$ (você pode deixar as derivadas indicadas).
   c) Prove, de modo argumentativo, a convergência em distribuição dos estimatores de $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\delta}$, $\boldsymbol{\theta}_D$ e $\boldsymbol{\varrho}$, apresentando as suposições e/ou teoremas que devem ser verificados para que seus argumentos sejam válidos.

3. Proponha uma situação e prove, de modo argumentativo, na qual o resíduo normalizado (apresentado eu aula), para o modelo misto, tem aproximadamente distribuição $N(0, 1)$, apresentando as suposições que devem ser verificadas para que seus argumentos sejam válidos. Suponha que $\boldsymbol{R}_j, \boldsymbol{D}$ e σ^2 são conhecidos.

4. Repita o item anterior considerando $\boldsymbol{R}_j$, $\boldsymbol{D}$ e σ^2 são desconhecidos.

5. Prove, de modo argumentativo, que a distribuição assintótica da estatítica Q (para testar hipóteses do tipo $\boldsymbol{C\beta} = \boldsymbol{M}$), sob H_0 e H_1 são aquelas apresentadas em sala, apresentando as suposições que devem ser verficadas para que seus argumentos sejam válidos.

6. Proponha uma estatística (análoga àquela apresentada para testar hipóteses do tipo $\boldsymbol{C\beta} = \boldsymbol{M}$) para testar hipóteses do tipo $\boldsymbol{C\theta} = \boldsymbol{M}$ em que $\boldsymbol{\theta}$ são os parâmetros de variância e correlação, relativo aos efeitos aleatórios. Obtenha sua distribuição assintótica (de modo argumentativo) sob H_0 e sob H_1.

7. Considere o MLGH visto em aula. Apresente a forma do vetor escore e da Informação de Fisher para $(\boldsymbol{\gamma},^t, \phi)^t$ (suponha que $\boldsymbol{\Psi}$ seja conhecida), em relação ao método de máxima verossimilhança marginal.

8. Repita o item anterior supondo $\boldsymbol{\Psi}$ desconhecida e não estruturada, para $(\boldsymbol{\gamma},^t, \phi)^t$ e $\boldsymbol{\Psi}$.

9. Em relação a Questão 7), apresente as expressões (em função de integrais) dos preditores (esperança da disrtribuição condicional) dos efeitos aleatórios. Considere um modelo com, somente intercepto aleatório.
10. Prove, de modo argumentativo, que os estimadores de máxima verossimilhança residual, para os efeitos fixos e componentes de variância, convergem em distribuição para uma dada distribuição normal multivariada, para o MH de dois níveis (normal/normal), especificando as condições nas quais isso ocorre.
11. Repita o item anterior para o MLGH de dois níveis.
12. Implemente, no programa de sua preferência, a construção do gráfico QQplot com evelopes para RCD do MLGH binominal negativo (semelhante ao que fora feito para o resíduos de Pearson).
13. Repita o item anterior para o MLGH gama.
14. Compare as funções glmer.nb e glmmadmb (ambos via quadratura adaptativa), através de um estudo de simulação, da seguinte forma: Gere (para cada cenário, ou seja, combinação dos fatores de interesse), R = 100 réplicas (respostas), do modelo:

$$\begin{array}{rcl} Y_{ji}|u_{0j} & \overset{ind.}{\sim} & \text{BN}(\mu_{ji}, \phi = 20) \\ \ln(\mu_{ji}) & = & \mu_j + 0,5x_j \\ \mu_j & = & 1 + u_{0j}, \\ u_{0j} & \overset{iid}{\sim} & N(0, \sigma^2 = 1) j = 1, ..., J; i = 1, ..., n_j \end{array}$$

    Fatores: $J \in \{20, 50, 100\}$, $n_j \in 5, 10, 50$. Utilize asestatísticas de comparação usuais, ou seja, média, variância, vício, raiz quadrática do erro quadrático médio e box-plot das estimativas, para comparar as estimativas de $\gamma_{00} = 1, \gamma_{10} = 0,5, \phi$ e σ^2. Além disso, simule uma única vez $x_j \overset{iid}{\sim} U(0,1)$, centralizando, posteriormente, na média amostral, esses valores e os considere para todas as réplicas.

15. Repita o item anterior para o MLGH gama, considerando o modelo:

$$\begin{array}{rcl} Y_{ji}|u_{0j} & \overset{ind.}{\sim} & \text{gama}(\mu_{ji}, \phi = 20) \\ \ln(\mu_{ji}) & = & \mu_j + 0,5x_j \\ \mu_j & = & 1 + u_{0j}, \\ u_{0j} & \overset{iid}{\sim} & N(0, \sigma^2 = 1) j = 1, ..., J; i = 1, ..., n_j \end{array}$$

16. Analise os dados da Questão 7, da Lista II, através de (pelo menos) um MLGH para dados de contagem, da forma mais completa possível. Compare, da forma mais completa possível, o ajuste do modelo aqui escolhidom com aquele que o fora na referida Questão/Lista.

17. Analise os dados da Questão 8, da Lista II, através de (pelo menos) um MLGH para dados de contagem, da forma mais completa possível. Compare, da forma mais completa possível, o ajuste do modelo aqui escolhidom com aquele que o fora na referida Questão/Lista.

18. O problema a ser modelado está relacionado a um estudo sobre privação de sono. O tempo médio de reação por dia para indivíduos em um estudo de privação de sono. No dia 0, os indivíduos tiveram sua quantidade normal de sono. A partir daquela noite, eles estavam restritos a 3 horas de sono por noite. As observações representam o tempo médio de reação (Reaction) em uma série de testes dados a cada dia (Days) para cada sujeito (Subject). Os dados estão disponíveis no pacote do R “lme4” sob o nome “sleepstudy”. Análise, descritiva e inferencialmente, da forma mais completa possível, o tempo de reação, ao longo dos dias, para cada indivíduo. Utilize o modelo hierárquico mais apropriado que você identificar.