

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

1. Resolva TODOS os exercícios deixados em sala.
2. Considere o conjunto de dados de Potthoff and Roy. Escreva os modelos abaixo (sempre considerando retas distintas entre os sexos, para modelar a média da distância em função do ano), apresentando pelo menos uma interpretação para todos os parâmetros (efeitos fixos, componentes de variância e efeitos aleatórios) e calculando os momentos (médias, variâncias, covariâncias e correlações) e distribuições (condicionais e marginais) da resposta.
 - a) Somente interceptos aleatórios, com distribuições variando entre os sexos.
 - b) Somente coeficientes angulares aleatórios, com distribuições variando entre os sexos.
 - c) Interceptos e coeficientes angulares aleatórios, com distribuições variando entre os sexos.
3. Considere uma situação com J hospitais, cada um com n_j , $j = 1, 2, \dots, J$ pacientes, para os quais mediu-se o nível de colesterol (em mg/dL , resposta), bem como o ICM (em Kg/m^2 , covariável) e idade (anos, covariável), desses pacientes. Proponha um modelo hierárquico apropriado, de modo que todos os parâmetros (efeitos fixos, componentes da variância, efeitos aleatórios) tenham interpretação e que as estimativas dos parâmetros de regressão sejam diretamente comparáveis. Apresente pelo menos uma interpretação para todos os parâmetros (efeitos fixos, componentes de variância e efeitos aleatórios), calculando os momentos (médias, variâncias, covariâncias e correlações) e distribuições (condicionais e marginais) da resposta.
4. Considere o MH de dois níveis (na forma matricial), pelos elementos do nível 2, bem como considerando todas as observações concatenadas (link, slides 8 e 9). Responda os itens:
 - a) Obtenha as distribuições (condicionais e marginais) de \mathbf{Y}_j , $j = 1, \dots, J$ e de $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_J)'$, em relação aos efeitos aleatórios (\mathbf{u}_j e \mathbf{u} , respectivamente).
 - b) Obtenha as distribuições (condicional e marginal, em relação a \mathbf{u}_j) do vetor $\mathbf{Y}_j^* = \mathbf{U}_j \mathbf{Y}_j$, $\mathbf{U}_j = \mathbf{I}_{n_j} - \mathbf{Z}_j (\mathbf{Z}'_j \mathbf{Z}_j)^{-1} \mathbf{Z}'_j$.
 - c) Obtenha as distribuições (condicional e marginal, em relação a \mathbf{u}) do vetor $\mathbf{Y}^* = \mathbf{U} \mathbf{Y}$, $\mathbf{U} = \mathbf{I}_n - \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}'$.
 - d) Com base na distribuição de \mathbf{Y}^* (item c)) obtenha as marginais de \mathbf{Y}_j^* . Compare-as com as distribuições obtidas no item b). Suas conclusões eram esperadas? Justifique, adequadamente, seus comentários.

5. Considere o conjunto de dados Achieve ([link](#)) e as variáveis: “gender” (sexo, 1: feminino, 2: masculino, explicativa), “school” (escola, nível 2), “gemath” (escore em Matemática, quanto maior, melhor o desempenho, resposta), “gelang” (escore em Linguagem, quanto maior, melhor o desempenho, explicativa) e id (designa o aluno, nível 1). O objetivo é verificar (através de modelos hierárquicos apropriados) como a variável “gelang” afeta a variável “gemath”, ao longo das escolas, considerando a estrutura hierárquica presente nos dados. Analise, da forma mais completa possível os dados, respondendo as questões de interesse.
6. Repita o item anterior, considerando, adicionalmente, a variável “gender”.
7. Considere o conjunto de dados aqui descritos e disponibilizados: ([link](#)). Resumidamente: os dados são de um estudo sobre ganho de peso (resposta), onde os pesquisadores designaram aleatoriamente 30 ratos em três grupos de tratamento: o tratamento 1 foi um controle (sem aditivo); os tratamentos 2 e 3 consistiram em dois aditivos diferentes (tiouracil e tiroxina respectivamente), na água ingerida pelos ratos. O peso foi medido em um tempo basal (semana 0) e nas semanas 1, 2, 3 e 4. Nota: Devido a um acidente no início do estudo, dados de 3 ratos do grupo tiroxina não está disponíveis. O objetivo é comparar os grupos, em relação ao ganho de peso ao longo do tempo, considerando a estrutura hierárquica presente (medidas repetidas agrupadas nos ratos), através de modelos hierárquicos apropriados. Veja também : o pacote do R ALA ([link](#)). Analise, da forma mais completa possível os dados, respondendo as questões de interesse.
8. Considere o conjunto de dados “Mmmec” disponível no pacote do R “mlmRev”. O interesse reside em estudar o comportamento do número de mortes de homens por melanoma maligno (“deaths”, resposta) em função da medida centrada da dose de UVB atingindo a superfície da Terra em cada condado (“uvb”) em diversos condados europeus (“county”, nível 1) ao longo de diversos países (“nation”, nível 2), através de modelos hierárquicos apropriados. Analise, da forma mais completa possível os dados, respondendo as questões de interesse.