

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

1. Resolva TODOS os exercícios deixados em sala.
2. (Re)Analise todos os conjuntos de dados vistos até agora, do ponto de vista frequentista, de forma adequada, usando o modelo mais adequado possível (obs: você pode se ater aos modelos vistos durante o curso).
3. Analise todos os conjuntos de dados vistos até agora, do ponto de vista bayesiana, de forma adequada, usando o modelo mais adequado possível (obs: você pode se ater aos modelos vistos durante o curso).
4. Considere o seguinte MLGH: $Y_{ji}|u_{0j} \stackrel{ind.}{\sim} \text{gama}(\mu_{ji}, \phi)$ (veja a forma da densidade no livro Paula (2013)), em que $\mathcal{E}(Y_{ji}|b_j) = \mu_{ji}$, $\mathcal{V}(Y_{ji}|u_{0j}) = \mu_{ji}^2/\phi$, $\ln \mu_{ji} = \beta_0 + \beta_1(x_{ji} - \bar{x}) + u_{0j}$, $\bar{x} = \frac{1}{J} \sum_{j=1}^J \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji}$ e $u_{0j} \stackrel{iid}{\sim} N(0, \psi)$. Encontre as expressões analíticas para $\mathcal{E}(Y_{ji})$, $\mathcal{V}(Y_{ji})$, $Cov(Y_{ji}, Y_{j'i'})$ e $Corr(Y_{ji}, Y_{j'i'})$.
5. Repita o item anterior considerando $\ln \mu_{ji} = \beta_0 + \beta_1(x_{ij} - \bar{x}) + u_{0j} + u_{1j}(x_{ij} - \bar{x})$, em que $\mathbf{u}_j = (u_{0j}, u_{1j})' \stackrel{iid}{\sim} N_2(\mathbf{0}, \mathbf{\Psi})$, $\mathbf{\Psi} = \begin{bmatrix} \psi_{00} & \psi_{01} \\ \psi_{01} & \psi_{11} \end{bmatrix}$.
6. Repitam o item anterior para os modelos Poisson e binomial negativo.
7. Assuma que $Y \sim LN(\alpha, \beta)$, $Y = \ln(X)$, $X \sim N(\alpha, \beta)$. Reparametrize a distribuição de Y , em termos da média (μ) e de um parâmetro de precisão (ϕ). Com base nessa parametrização, defina, como o usual, um MLGH de dois níveis, com apenas uma covariável no nível 1, sem covariáveis no nível 2, e intercepto aleatório (veja a questão 4). Encontre as expressões analíticas para $\mathcal{E}(Y_{ji})$, $\mathcal{V}(Y_{ji})$, $Cov(Y_{ji}, Y_{j'i'})$ e $Corr(Y_{ji}, Y_{j'i'})$.
8. Considere o modelo de regressão usual beta binomial ([aqui](#), slides 12,13 e 14). Defina um modelo hierárquico de dois níveis, com intercepto aleatório somente e função de ligação apropriada, calculando, da forma mais simplificada possível, $\mathcal{E}(Y_{ji})$, $\mathcal{V}(Y_{ji})$, $Cov(Y_{ji}, Y_{j'i'})$ e $Corr(Y_{ji}, Y_{j'i'})$.
9. Repita a questão anterior considerando tanto intercepto quanto coeficientes angulares apropriados.
10. (Paula (2013) No arquivo “ratosgee.dat” estão os dados de um experimento em que 30 ratos tiveram uma condição de leucemia induzida. Três drogas quimio-terápicas foram utilizadas no tratamento dos animais. Foram coletadas de cada animal a quantidade de

células brancas (WBC), a quantidade de células vermelhas (RBV) e o número de colônias de células cancerosas (RESP) em quatro períodos diferentes. O objetivo é verificar o comportamento dos grupos ao longo do tempo, em relação aos três grupos, levando-se em consideração as covariáveis. Faça análises, descritiva e inferencial, completas e apropriadas, apresentando os comentários e conclusões, pertinentes.

11. (O arquivo de dados está disponível no R no pacote “prLogistic” sob o nome “Toenail”). Os dados são de um estudo comparando dois tratamentos orais para a infecção do dedo do pé-unha, incluindo informações para 294 pacientes medidas em 7 semanas, compreendendo um total de 1908 medições. A variável binária resultado “onycholysis” indica o grau de separação da lâmina ungueal da unha-cama (nenhuma ou leve contra moderada ou grave). Os pacientes foram avaliados no início do estudo (semana 0) e nas semanas 4, 8, 12, 24, 36, 48 e daí em diante. As variáveis no arquivo de dados são: ID - identificação do paciente, Response - Grau de onycholysis (0 se nenhum ou leve, 1 se moderada ou grave), Treatment - Indicador oral tratamento (1 = Tratamento A, 0 = Tratamento B), Month - Tempo exato das medições, em meses, Visit - (números da visita, 1-7, correspondendo as visitas agendadas a 0, 4, 8, 12, 24, 36 e 48 semanas). Faça análises, descritiva e inferencial, completas e apropriadas, apresentando os comentários e conclusões, pertinentes.
12. (Conjunto de dados disponível no pacote R “bild” sob o nome de “locust”): O conjunto de dados foi apresentado e analisado por MacDonald e Raubenheimer (1995), e posteriormente examinado por outros autores, incluindo Gonçalves e Azzalini (2008). Os dados foram coletados para estudar o efeito da fome sobre o comportamento de locomoção de gafanhotos (*Locusta migratoria*). Especificamente, 24 gafanhotos foram observados em 161 instantes de tempo, em intervalos de trinta segundos; as unidades experimentais foram divididas em dois grupos de tratamento (“feed” e “unfeed”), e dentro de cada um dos dois grupos, os sujeitos eram alternados em relação ao sexo “masculinos” e “femininos”. Para o propósito desta análise, as categorias da variável de resposta são “em movimento” e “não em movimento” (veja também a descrição disponível no R). Considere todas as variáveis, sendo a variável “move”, a resposta. Faça análises, descritiva e inferencial, completas e apropriadas, apresentando os comentários e conclusões, pertinentes.
13. Este conjunto de dados faz parte de um outro conjunto de dados longitudinais coletados pelo Hospital Geral da cidade de St. John’s, no Canadá. Consideramos aqui os dados de contagem longitudinal que contêm os registros completos de 144 indivíduos por quatro anos ($n = 4$) de 1985 a 1988. O número de visitas a um médico por cada indivíduo durante um determinado ano foi registrado (resposta), durante esse período. Além disso, as informações sobre quatro covariáveis, a saber, sexo, número de condições crônicas, nível de escolaridade e idade, foram registrados para cada indivíduo. Faça análises, descritiva e inferencial, completas e apropriadas, apresentando os comentários e conclusões, pertinentes (veja também [aqui](#), dados: planilha health care utilization data.ods).