- Questão 1

  a) Usando propriedades de esperança e variância condicionais, temos que: $\mathcal{E}(Y_{ji}|u_{0j}) = \gamma_{01}x_j + u_{0j}$, $\mathcal{E}(Y_{ji}) = \gamma_{01}x_j$, $\mathcal{V}(Y_{ji}|u_{0j}) = \sigma^2$ e $\mathcal{V}(Y_{ji}) = \sigma^2 + \psi$.
  b) Usando propriedades da distribuição normal (multivariada) e o item a), temos que: $Y_{ji}|u_{0j} \sim N(\gamma_{01}x_j + u_{0j}, \sigma^2)$ e $Y_{ji} \sim N(\gamma_{01}x_j, \sigma^2 + \psi)$.
  c) Pela estrutura do modelo e por propriedades do cálculo de probabilidades, temos que: $Cov((Y_{ji}, Y_{ji'})|u_{0j}) = 0$, $Corre((Y_{ji}, Y_{ji'})|u_{0j}) = 0$, $Cov(Y_{ji}, Y_{ji'}) = \psi$, $Corre(Y_{ji}, Y_{ji'}) = \frac{\psi}{\psi+\sigma^2}$, $i \neq i'$
  d) Eu deveria ter colocado $n_j = n, \forall j$, mas considerei tudo o que fora feito por vocês. Nesse caso, a solução passa por usar algo como descrito em aqui, slides de 1 a 9.

- Questão 2 : Comentários

  1. As covariáveis tem de ser inseridas de forma padronizada, de modo a poder se comparar, diretamente, o grau de influencia de cada uma delas (dentre aquelas que são significativas no modelo) na resposta. Veja aqui, slide 16.
  2. Como não temos informações nesse sentido, podemos propor um modelo de regressão linear hierárquico de dois níveis usual.
  3. Não há motivos para não considerarmos interceptos e coeficientes angulares aleatórios.

- Questão 3 (sem apresentar justificativas, foram comentadas em aula)

  a) Parece positivo, linear (aproximadamente) e equivalente entre os tipos de escolas.
  b) São as interpretações usuais, em função do problema. Por exemplo: (γ_{000}): o aritPost esperado, para alunos de escolas públicas com QI = 11 é igual a 18,41. Assim por diante. Não são as interpretações genéricas.
  c) Parece haver influência positiva (e signficativa) porém, os resultados sugerem que ela é a mesma entre os tipos de escola.
  d) O modelo parece ser razoável pois possibilita verificar se o comportamento dos tipos de escolas, em relação à como o QI influencia o aritPost, é o mesmo, além de contemplar a aparente varibilidade entre escolas, de forma adequada. A estrutura linear de regressão, nos dois níveis, também é razoável. Talvez um modelo (inicial) sem coeficientes angulares aleatórios fosse mais apropriado.
  e) Em termos da especificação do erro condicional, não apresenta problemas, pois as hipóteses de normalidade, heterocedasticidade e indenpendência parecem ser satisfetias (RCM). Quanto aos efeitos aleatórios, parece haver algum problema na especificação da distribuição deles, e/ouuma má especificação da estrutura de regressão do nível 2 (RCP,RMP). Em geral, o modelo apresenta problemas de ajuste, sendo um caminho modificar a estrutura dos efeitos aleatórios.

- Questão 4

  a) Dois níveis: nível 1 - medidas repetidas (dentro dos indivíduos), níve 2 - indivíduos (recém-nascidos). Resposta: concentração de bilirrubina, covariável: tempo (semanas da medição).
  b) CCI: 1. interpretações, percentual da variabilidade devida aos recem-nascidos; 2. correlação das medidas intra-indivíduos.

c) A relação entre CB e semanas é, visivelmente, não linear, crescendo, depois diminuindo. Dentro da classe de MLH (normal/normal) seria apropriado um modelo de regressão segmentado (modelos quadráticos também poderiam ser considerados). Veja uma análise apropriada, dentro dessa classe aqui (slides de 45 a 62) e aqui (slides de 19 a 20), O RCM apresenta leve assimetria com alguns outliers enquanto que os efeitos aleatórios apresentam forte assimetria positiva. O RCP apresenta caudas pesadas e heterocedasticidade; Sugestão: ajustar um modelo com distribuição t-assimétrica para o erro condicional.

d) Quanto à interpretação, é semlhante à Questão 3), ou seja, particularizar a interpretação geral de todos os parâmetros, com base nas estimativas obtidas.