6.1 The dataset and the description below come from OpenBugs examples, following the study

H. Goldstein *et al.* (1993). *A Multilevel Analysis of School Examination Results*, Oxford Review of Education, **19** (4), pp. 425–433.

The authors analyse exam results from inner London schools and student to study study the between-school variation in order to provide ranking of school.
Standardized mean examination scores were available for 1978 pupils from 38 different schools. Pupil-level covariates included gender plus a standardized London Reading Test (LRT) score and a verbal reasoning (VR) test category (1, 2 or 3, where 1 represents the highest ability group) measured when each child was aged 11. Each school was classified by gender intake (all girls, all boys or mixed) and denomination (Church of England, Roman Catholic, State school or other); these were used as categorical school-level covariates. Both the London reading test score and the verbal reasoning test were performed at the beginning of the year.
The `goldstein` data contains the following variables:

- `score`: standardized end-of-year exam score for each pupil,
- `school`: school id,
- `LRT`: London reading test score,
- `VR`: verbal reasoning (VR) test category (`1`, `2` or `3`, where `1` represents the highest ability group and `3` the lowest),
- `gender`: gender of pupil, either female (`0`) or male (`1`),
- `type`: type of school, either all-girl schools, all-boy school or mixed,
- `denom`: denomination of school, either Church of England, Roman Catholic, state or other.

(a) Give the range of the number of pupils per school and use this information to determine if it is feasible to estimate a fixed (group-)effect for `school`.

(b) Write down the equation of the postulated model for `score` that includes `LRT`, `VR`, `sex`, `type` and `denom` as fixed effects and `school` as random effect, with `VR=3`, mixed for `type` and other for `denom` as baseline categories. Fit the model and interpret the significant parameters associated to the fixed effects.

(c) One could consider adding `VR` as random effect as opposed to fixed effect. Which of the two makes the most sense and how do the models conceptually differ?

(d) Using the fitted model (with a random intercept for `school`), obtain the estimated covariance matrix for school 37 and explain how to obtain it given the estimated covariance parameters. Write down the proportion of the total variance due to school.

(e) Produce a normal quantile-quantile plot of the predicted random effects for school and hence comment on the model assumption for the random effect.

(f) The goal of Goldstein *et al.* (1993) was to rank schools. What is the benefit of pooling information from schools in order to estimate their average score? Plot the predicted school random effect as a function of school id, with prediction intervals based on the formula $\widehat{b}_i \pm 1.96\mathrm{se}(\widehat{b}_i)$ (you may need to perform obtain the bounds manually). What is the predicted top five ranking? Why is this information alone not sufficient for ranking the schools?

6.2 **Gender disparities in GSCE scores**: the data are from *The Associated Examining Board* in Guildford, and contains 1905 records. They have been used to examine the relationship between candidates' sex and their examination performances in

Cresswell, M.J. (1990). *Gender Effects in GCSE: Some Initial Analyses*, Associated Examining Board Research Report RAC/517.

The General Certificate of Secondary Education (GSCE) is a national exam administered in the UK that leads to high school degree. The variables include

- `center`: categorical variable identifying the 72 examination centers.
- `sex`: binary indicator, equal to zero for male (0) and unity for female (1).
- `result`: result on the GCSE exam
- `coursework`: grades on related coursework evaluated by candidate's teacher.

The response variable is `result`. Fit the following three models, including `sex` and `coursework` as covariates in each.

- Model 6.2.1, a linear regression model with a compound symmetry covariance on the errors within-center.
- Model 6.2.2, a linear regression model with fixed effect for `center` and independent errors.
- Model 6.2.3, a mixed model with a random intercept in each center and independent error terms.

(a) Briefly explain why it could be legitimate to model the correlation in these data. State what the "group" would represent in this context.

Solution

The different center may represent geographic areas where disparities in income or ressources may affect performance. The group is `center`.

(b) Explain the main benefit of using Model 6.2.3 over Model 6.2.1.

Solution

We can get predictions of group effects for each `center`.

(c) Explain the two main benefits of using Model 6.2.3 over Model 6.2.2.

Solution

The random intercept naturally induces correlation between observations. The random effects are not estimated parameters, but can be used for prediction so we can estimate other coefficients that would not be identifiable (or not estimable because of small sample sizes).

(d) Write down the postulated covariance matrix **within a center** with three candidates for both the response $\boldsymbol{Y}$ and the errors $\boldsymbol{\varepsilon}$ for each of Models 6.2.1 and 6.2.3.

Solution

For Model 6.2.1,

$$\mathsf{Cov}\left(\boldsymbol{y}\right) = \begin{pmatrix} \sigma^2+\tau & \tau & \tau \\ \tau & \sigma^2+\tau & \tau \\ \tau & \tau & \sigma^2+\tau \end{pmatrix}, \qquad \mathsf{Cov}\left(\boldsymbol{\varepsilon}\right) = \mathsf{Cov}\left(\boldsymbol{y}\right),$$

whereas, for Model 6.2.3,

$$\mathsf{Cov}\left(\boldsymbol{y}\right) = \begin{pmatrix} \sigma^2+\sigma^2_b & \sigma^2_b & \sigma^2_b \\ \sigma^2_b & \sigma^2+\sigma^2_b & \sigma^2_b \\ \sigma^2_b & \sigma^2_b & \sigma^2+\sigma^2_b \end{pmatrix}, \qquad \mathsf{Cov}\left(\boldsymbol{\varepsilon}\right) = \sigma^2\mathbf{I}_3.$$

(e) What is the estimated correlation between two individuals in different center according to Model 6.2.3?

Solution
Zero, since we assume they are independent.

(f) What is the estimated correlation between two individuals in the same center according to Model 6.2.3?

Solution
The estimated correlation is $\widehat{\sigma}_b^2/(\widehat{\sigma}_b^2+\widehat{\sigma}^2) = 106.20/(240.72+106.2) = 0.308$.

(g) Would a first-order autoregressive model, AR(1), be adequate for modelling the within-group correlation in this problem? Briefly explain why or why not.

Solution
No, because the AR(1) model implies decaying correlation, yet individuals are exchangeable within groups.

(h) Using Model 6.2.3, predict the GSCE score for a women who obtained 91 on her coursework and takes the examination in center 2.

Solution
The predicted score is $34.2301 + 91 \times 0.5993 - 8.4188 - 7.9898 = 72.36$ points.

(i) Using Model 6.2.3, what will the average score be for men candidates whose average score for coursework is 100 and who take their examination in a newly opened center?

Solution
The estimated average on the GSCE is $34.23 + 59.93 = 94.16$ points.

(j) Is the coefficient for `sex` in Model 6.2.3 statistically significant? Justify your answer and interpret the estimated coefficient.

Solution
Based on the output, the Wald test statistic is -10.96 (equal to the signed square root of the F-statistic, is 120.17). There is a significant difference (p-value negligible) with an estimated difference of 8.42 in favor of women (so women have on average, *ceteris paribus*, scores that are 8.42 points higher than men).