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`.

Solution

There are between one and 198 pupils per school. Seven schools have fewer than five pupils, so estimating a fixed school-effect is not possible for those.

(b) Write down the equation of the postulated (theoretical) 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.
Don't forget to specify the distribution of the errors and random effects.

Solution

The equation of the model

$$\begin{aligned}\texttt{score}_{ij} &= \beta_0 + b_i + \beta_1\texttt{LRT}_{ij} + \beta_2\texttt{VR}_{ij1} + \beta_3\texttt{VR}_{ij2} + \beta_4\texttt{sex}_{ij} + \beta_5\texttt{type}_{ij\texttt{allboys}} \\ &\quad + \beta_6\texttt{type}_{ij\texttt{allgirls}} + \beta_7\texttt{denom}_{ij\texttt{angli}} + \beta_8\texttt{denom}_{ij\texttt{catho}} + \beta_9\texttt{denom}_{ij\texttt{state}} + \varepsilon_{ij}\end{aligned}$$

where i is the school and j the id of the student within the school. We assume that $b_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0,\sigma_s^2)$ and that $\varepsilon_{ij} \stackrel{\text{iid}}{\sim} \mathcal{N}(0,\sigma^2)$, in addition to independence between the random intercepts (per school) and errors.

(c) Fit the model and interpret the significant parameters associated to the fixed effects.

Solution

Based on multivariate Wald tests (type III), all variables are statistically significant, but `type` (note: depending on the test, `denom` may or not be statistically significant). The parameter corresponds

- The intercept ($\widehat{\beta}_0 = -0.568$) gives the estimated score of women from mixed schools with other denomination who scored average on the London reading test and have low verbal reasoning scores (`VR=3`).
- The coefficient estimate $\widehat{\beta}_1 = 0.03$ corresponds to an increaseof 0.03 of the standardized score for every unit increase of the (standardized) LRT score, *ceteris paribus*.

- Everything else constant, girls standardized scores are 0.154 higher than those of boys on the end-of-year exam.
- *Ceteris paribus*, the difference in standardized scores between pupils from Church of England/Roman Catholic/State schools relative to other denomination are $\widehat{\beta}_7 = -0.27/\widehat{\beta}_8 = 0.186/\widehat{\beta}_9 = -0.17$, respectively.

(d) 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?

Solution

Fixed effects are more logical here because we are interested in the mean effect for `VR` and we don't necessarily want to introduce correlation between people from different school. However, this is a judgment call; we have only three categories with 1918 observations and enough in each modality; the effect is also statistically significant. The differences between the two models lie in different parametrizations (the mean of the random effects is zero, so the constraint on the coefficients is on the sum rather than on `VR=3` — while not exactly equal, the difference in mean prediction for `VR=3` versus `VR=1` and `VR=2` are almost the same. The random effect for `VR` also adds to the covariance matrix correlation between students with similar `VR` scores and shrinks the estimates.

(e) 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.

Solution

The covariance parameter estimates are $\widehat{\sigma}^2_s = 0.039$, and $\widehat{\sigma}^2 = 0.5715$ for respectively the variances of the random intercepts for school and errors. School 37 consists of four students so, by independence of the errors and the random effects, the covariance matrix is

$$\mathsf{Var}\left(\boldsymbol{y} \mid \texttt{school} = 37\right) = \begin{pmatrix} \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s \\ \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s \\ \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s \\ \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s & \widehat{\sigma}^2_s \end{pmatrix} + \begin{pmatrix} \widehat{\sigma}^2 & 0 & 0 & 0 \\ 0 & \widehat{\sigma}^2 & 0 & 0 \\ 0 & 0 & \widehat{\sigma}^2 & 0 \\ 0 & 0 & 0 & \widehat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} 0.611 & 0.0395 & 0.0395 & 0.0395 \\ 0.0395 & 0.611 & 0.0395 & 0.0395 \\ 0.0395 & 0.0395 & 0.611 & 0.0395 \\ 0.0395 & 0.0395 & 0.0395 & 0.611 \end{pmatrix}$$

The estimated percentage of the total variance due to school is 0.0646; it corresponds to the correlation between pupils in any given school — it can also be obtained from the covariance model.

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

Solution

There are 38 schools, so we expect on average 2 points to fall outside the simulated pointwise confidence interval. The quantile-quantile plot on the right-panel of Figure 1 shows no evidence against the normality of the random intercepts for school. Other standard diagnostics of residuals suggest no evidence against linearity, heteroscedasticity and normality of errors.

(g) The goal of Goldstein *et al.* (1993) was to rank schools. What is the benefit of pooling information from schools using random effects in order to estimate their average score?

Solution

Pooling allows us to borrow information for schools with very few pupil for various parameters and to estimate the variance based on the schools with enough pupils.

(h) Plot the predicted school 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 use the formula to calculate the bounds).

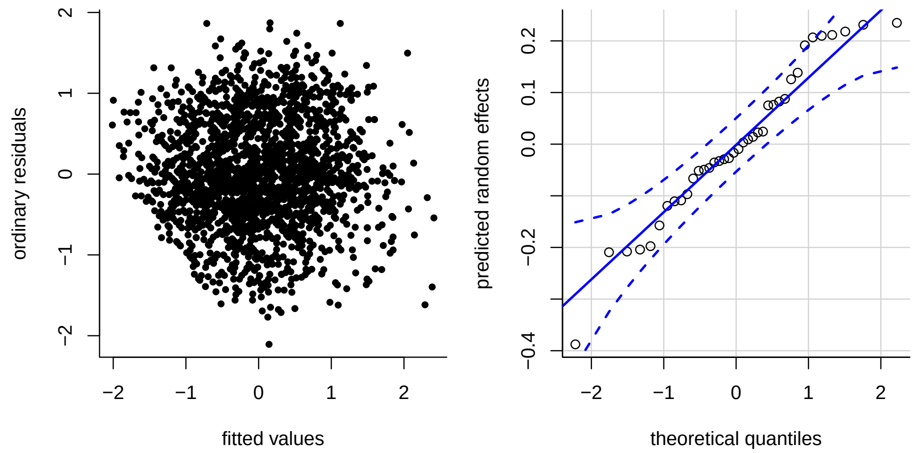

Figure 1: Diagnostic plot of fitted value against ordinary residuals (left) and normal quantile-quantile plot of predicted school-specific random intercepts with simulated 95% pointwise confidence intervals (right).

Solution

There is a fair dose of uncertainty in the potential true ranking (for every of these schools), since there are more than 35 pupils to base estimation from — this fact is reflected in the smaller predictions standard errors for the random intercept.

(i) What is the predicted top five school ranking?

Solution

The top-five ranking for the random effects are schools 2, 13, 18, 9, 30; see Figure 2. However, this effect merely indicates which schools outperform the average of their category. We should minimally include the `type` and `denom` of the school to the ranking to decide which one is best (other characteristics are student-specific).
The correct way to predict would thus be to consider a lambda individual and keep only the explanatories that are fixed within school and fix all the other explanatories to a common values. This way, the prediction would include both random effects (via the predicted random intercept) and the fixed effects of `type` and `denom`. One caveat here is that all-girls school cannot have boys (and vice-versa).
The school ranking from best score (from best to fifth rank) is 9, 13, 10, 18 and 19.

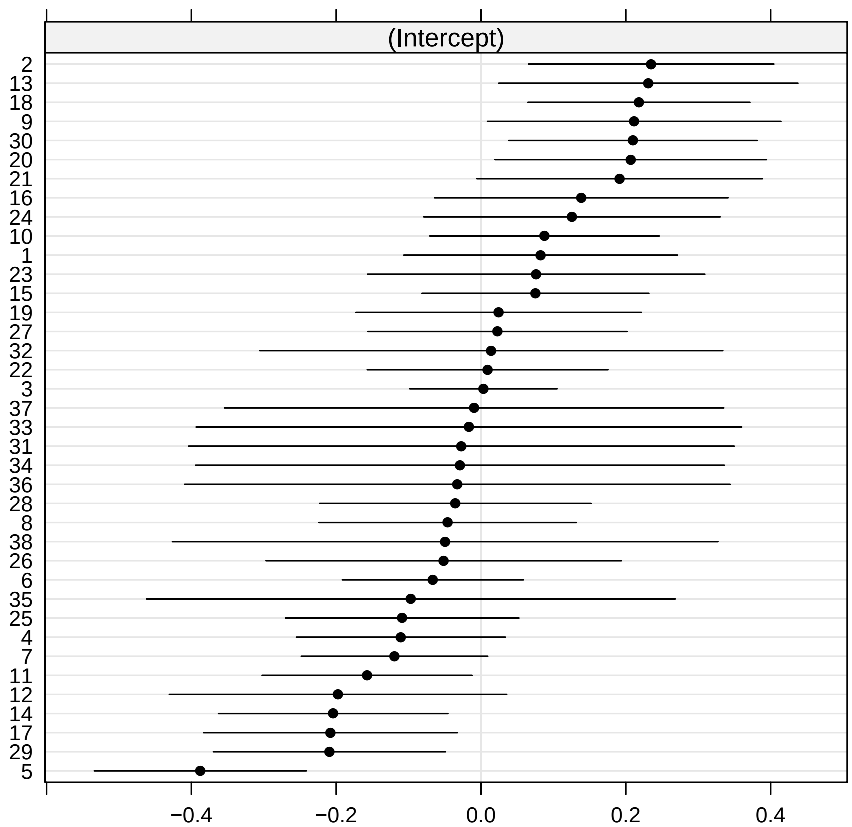

Figure 2: Plot of conditional random intercept (per school) with 95% Wald-based predictions intervals

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 $\widehat{\sigma}^2_b/(\widehat{\sigma}^2_b+\widehat{\sigma}^2) = 106.20/(240.72+106.2) = 0.308$.

(g) Would a first-order autoregressive model, $\mathsf{AR}(1)$, 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 boys (so girls have on average, *ceteris paribus*, scores that are 8.42 points lower than men).