Multilevel Analysis of School Examination Data

Benjamin Chan

The Chinese University of Hong Kong

November 20, 2018

Outline

Reference

- Data Source:
  Centre for Multilevel Modelling at University of Bristol
  `http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html`

- Data Set: **EXAM**
  The data set consists of 4,059 students from 65 schools in Inner London. The objective is to investigate the relationship between the exam score and the student-level and/or school-level variables.

- GitHub Repository:
  `https://github.com/BenjaminChanChunHo/Advanced-Modeling-and-Data-Analysis`

Motivation for Multilevel Models

- Multilevel Models are appropriate for hierarchically arranged data with nested sources of variability.
- In the data **EXAM**, the sources of variability include
  1. Different student abilities (subject-specific random errors)
  2. Differences between schools (random cluster effects)
- Cross sectional data:
  Students at the lower level cluster at the higher level, i.e. schools.
- To address the similarity between students in the same school, multilevel models are used.

Data Set Description

- The attribute name, attribute type and description are given on the website as follows:

Name	Data Type	Description
School ID	Integer	Unique to School
Student ID	Integer	Non-unique across School
Exam Score	Continuous	Standardized Response
LR* Test Score	Continuous	Standardized Predictor
Student Gender	Binary	0 = Boy; 1 = Girl
School Gender	Trinary	1 = Mix; 2 = Boy; 3 = Girl
School Intake Score	Continuous	Unique to School
Student VR* Band	Trinary	1 = Bot; 2 = Mid; 3 = Top*
Student Intake Band	Trinary	1 = Bot; 2 = Mid; 3 = Top*

Table 1: Attribute Information

* London Reading; Verbal Reasoning; Bot 25%, Mid 50%, Top 25%

Notation

- Let y_{ij} be exam score for school $j = 1, \dots, 65$ at level 2 and student within school $i = 1, \dots, n_j$ at level 1. Here $n = \sum_{j=1}^{65} n_j = 4059$.

Variable Name	Range	Notation	Original Variable
School ID	$\{1, 2, \dots, 65\}$	j	Same
Exam Score	-3.66 to 3.67	y	Same
LR Test Score	-2.93 to 3.02	x_1	Same
Student Gender	$\{0, 1\}$	x_2	Same
Student VR Mid	$\{0, 1\}$	x_3	VR Score $= 2$
Student VR Top	$\{0, 1\}$	x_4	VR Score $= 3$
Student Band Mid	$\{0, 1\}$	x_5	Intake Band $= 2$
Student Band Top	$\{0, 1\}$	x_6	Intake Band $= 3$
School Intake Score	-0.76 to 0.64	w_1	Same
School Boys	$\{0, 1\}$	w_2	School Gender $= 2$
School Girls	$\{0, 1\}$	w_3	School Gender $= 3$

Table 2: Notation

Univariate Plots at Student Level

- Assume that y and x_1 follow normal distribution. Specifically, $y \sim N(0, 1)$ and $x_1 \sim N(0, 1)$ are supported by the histograms:

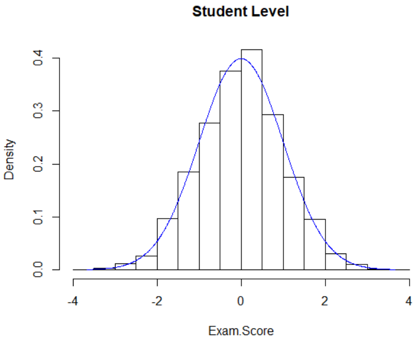

Figure 1: Histogram of y

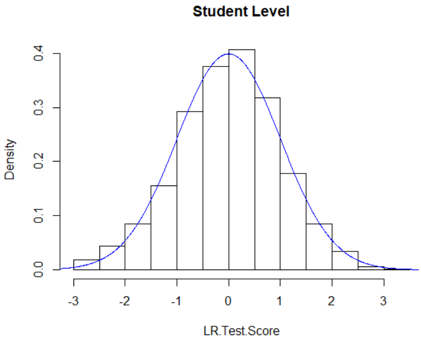

Figure 2: Histogram of x_1

Univariate Plots at School Level

- Heterogeneity among schools is supported by the histograms:

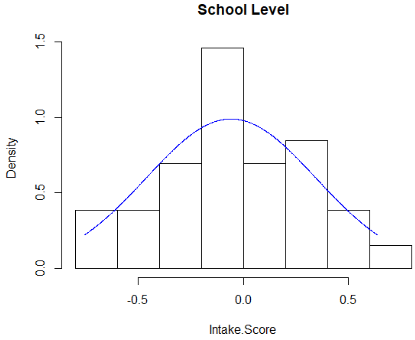

Figure 3: Histogram of w_1

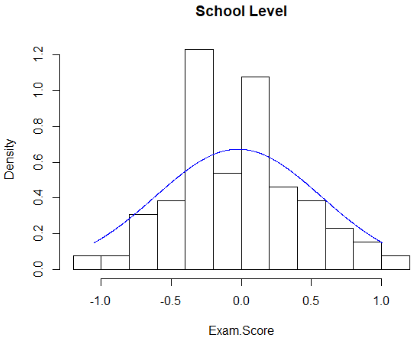

Figure 4: Histogram of $\bar{y}_j$

where $\bar{y}_j$ is the sample mean of y under School ID j.

Bivariate Relationship at Student Level

- Exam Score y and LR Test Score x_1 are positively correlated:

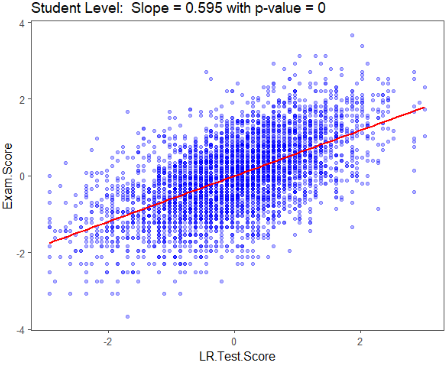

Figure 5: Scatter Plot of x_1 vs y

- The higher the Student VR Band is, the higher the Mean Exam Score is

Student VR Band	Mean Exam Score
1	-0.412
2	-0.062
3	0.349

Table 3: Mean Exam Score under each Student VR Band

- The lower the Student Intake Band is, the higher the Mean Exam Score is

Student Intake Band	Mean Exam Score
1	0.736
2	-0.142
3	-0.990

Table 4: Mean Exam Score under each Student Intake Band

Bivariate Relationship at School Level

- Sample mean of exam scores for each school $\bar{y}_j$ and School Intake Score w_1 are positively correlated:

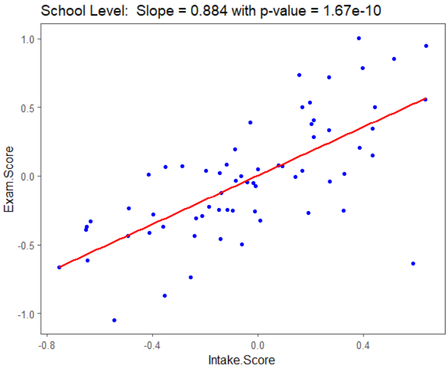

Figure 6: Scatter Plot of w_1 vs $\bar{y}_j$

Bayesian Estimation

- **Full Model** is defined to be a model using <u>all</u> variables in Table 2 while **Submodel** is defined to be a model using <u>some</u> variables.
- Markov Chain Monte Carlo (MCMC) algorithms are used to simulate a large number of observations from the joint posterior distribution.
- The Bayesian estimates of unknown parameters can be obtained from the corresponding sample means of simulated observations.
- The software WinBUGS is used via the R package R2WinBUGS to do the Bayesian inferences.
- For simplicity, the conjugate prior distributions (e.g. normal and gamma distributions) are used and hyperparameters are specified.

- Two sequences are generated with different initial values. The burn-in iterations required for achieving convergence of the MCMC algorithms are set to be 8,000. In fact, it can be checked by running mean plots of parameters.
- The number of simulated observations collected after the burn-in iterations is set to be 6,000. The posterior means and standard error estimates of parameters are reported.
- DIC and parsimony principle are used to compare models.

2.1 Full Model with Random and Fixed Intercept

- Let the full model with random and fixed intercept be

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + u_{0j} + \varepsilon_{ij} \\ &= (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \cdots + \beta_6 x_{6ij} + \varepsilon_{ij}, \end{aligned}$$

  where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$.

- Note that $\boldsymbol{x}_{ij}$ is 7×1 vector of covariates (including constant 1) and u_{0j} is level 2 random intercepts (school effects). Assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

- Using school level predictors $\boldsymbol{w}_j^T = (1, w_{1j}, w_{2j}, w_{3j})$,

$$\mu_{0j} = \boldsymbol{w}_j^T \boldsymbol{\alpha}_0 = \alpha_{00} + \alpha_{01} w_{1j} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j}.$$

Bayesian Setting

- The conjugate prior distributions of σ_0^2, σ_y^2, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_0$ are

$$\sigma_0^{-2} \sim \text{Gamma}[a_0, b_0], \quad \sigma_y^{-2} \sim \text{Gamma}[a_y, b_y],$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \text{ and } \boldsymbol{\alpha}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0),$$

  where a_0, b_0, a_y, b_y, and elements in $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are hyperparameters, and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_0$ are positive definite matrices.

- Since there is no available prior knowledge of the parameters, the hyperparameters are specified for convenience:

$$a_0 = 6, \;\; b_0 = 10, \;\; a_y = 6, \;\; b_y = 10,$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}, \;\; \boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \boldsymbol{I}, \;\; \boldsymbol{\mu}_0 = \mathbf{0} \text{ and } \boldsymbol{\Sigma}_0 = \boldsymbol{I}.$$

- There are two sets of initial values of parameters (denoted by *).
- The first set of initial values is:

$$\boldsymbol{\beta}^* = \mathbf{0}, \;\; \boldsymbol{\alpha}_0^* = \mathbf{0}, \;\; \sigma_0^{2*} = 1, \;\; \sigma_y^{2*} = 1$$

- The second set of initial values is

$$\beta_0^* = 0, \;\; \beta_1^* = \beta_3^* = 1, \;\; \beta_2^* = \beta_4^* = \beta_5^* = \beta_6^* = -1,$$

$$\alpha_{00}^* = 0, \;\; \alpha_{01}^* = \alpha_{03}^* = 1, \;\; \alpha_{02}^* = -1 \;\; \sigma_0^{2*} = 2, \;\; \sigma_y^{2*} = 2.$$

- Treatment is similar for other models and hence details are omitted thereafter. Please refer to GitHub for model specification and running mean plots.

Results

- Full Model with Random and Fixed Intercept:
  DIC = 9033.81

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	0.42	0.72	-0.80	1.68	Y	
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.23	N	*
β_3	Student.VR.Mid	0.00	0.25	-0.49	0.48	Y	
β_4	Student.VR.Top	0.08	0.39	-0.72	0.84	Y	
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.86	-0.65	N	*
α_{00}	u_Int	-0.26	0.67	-1.42	0.77	Y	
α_{01}	School.Intake.Score	0.21	0.41	-0.58	1.03	Y	
α_{02}	School.Boys	0.19	0.22	-0.24	0.61	Y	
α_{03}	School.Girls	0.14	0.17	-0.20	0.48	Y	
σ_0^2	var_u0	0.37	0.06	0.26	0.51	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

2.2 Full Model with Random but no Fixed Intercept

- The fixed intercepts β_0 and α_{00} are not significant in the sense that each of the Bayesian credible intervals contains 0.

- Let the full model with random but no fixed intercept be

$$y_{ij} = u_{0j} + \beta_1 x_{1ij} + \cdots + \beta_6 x_{6ij} + \varepsilon_{ij}$$
$$\mu_{0j} = \alpha_{01} w_{1j} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j}$$

  where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Full Model with Random but no Fixed Intercept:
  DIC = 9033.82

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.23	N	*
β_3	Student.VR.Mid	0.13	0.13	-0.13	0.38	Y	
β_4	Student.VR.Top	0.28	0.21	-0.13	0.74	Y	
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.86	-0.65	N	*
α_{01}	School.Intake.Score	0.04	0.29	-0.55	0.59	Y	
α_{02}	School.Boys	0.23	0.21	-0.19	0.64	Y	
α_{03}	School.Girls	0.16	0.17	-0.18	0.49	Y	
σ_0^2	var_u0	0.37	0.06	0.26	0.51	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

2.3 Full Model with Random Intercept and LRT Effect

- On top of the random intercept, consider random slope associated with the continuous student-level variable LR Test Score x_1.

- Let the full model with random intercept and LRT effect be

$$
\begin{aligned}
y_{ij} &= u_{0j} + (\beta_1 + u_{1j})x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_6 x_{6ij} + \varepsilon_{ij} \\
\mu_{0j} &= \alpha_{01} w_{1j} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j} \\
\mu_{1j} &= \alpha_{11} w_{1j} + \alpha_{12} w_{2j} + \alpha_{13} w_{3j}
\end{aligned}
$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$, $u_{1j} \sim N(\mu_{1j}, \sigma_1^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Full Model with Random Intercept and LRT Effect:
  DIC = 8990.33

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_1	LR.Test.Score	0.38	0.09	0.22	0.57	N	*
β_2	Student.Gender	0.17	0.03	0.10	0.23	N	*
β_3	Student.VR.Mid	0.14	0.13	-0.12	0.40	Y	
β_4	Student.VR.Top	0.21	0.22	-0.17	0.67	Y	
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.34	N	*
β_6	Student.Band.Top	-0.77	0.05	-0.87	-0.66	N	*
α_{01}	School.Intake.Score	0.14	0.30	-0.46	0.71	Y	
α_{02}	School.Boys	0.22	0.21	-0.20	0.64	Y	
α_{03}	School.Girls	0.16	0.17	-0.18	0.49	Y	
α_{11}	School.Intake.Score	0.12	0.21	-0.30	0.53	Y	
α_{12}	School.Boys	-0.01	0.20	-0.40	0.38	Y	
α_{13}	School.Girls	-0.03	0.16	-0.35	0.28	Y	
σ_0^2	var_u0	0.37	0.06	0.26	0.51	N	*
σ_1^2	var_u1	0.32	0.05	0.23	0.44	N	*
σ_y^2	var_y	0.52	0.01	0.50	0.55	N	*

- Consider removing the random intercept and leaving the random LRT effect.

- Let the full model without intercept, with random LRT effect be

$$
\begin{aligned}
y_{ij} &= (\beta_1 + u_{1j})x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_6 x_{6ij} + \varepsilon_{ij} \\
\mu_{1j} &= \alpha_{11} w_{1j} + \alpha_{12} w_{2j} + \alpha_{13} w_{3j}
\end{aligned}
$$

  where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{1j} \sim N(\mu_{1j}, \sigma_1^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Full Model without Intercept, with Random LRT Effect: DIC = 9391.79

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_1	LR.Test.Score	0.39	0.10	0.21	0.57	N	*
β_2	Student.Gender	0.17	0.02	0.12	0.22	N	*
β_3	Student.VR.Mid	0.17	0.03	0.12	0.23	N	*
β_4	Student.VR.Top	0.30	0.03	0.23	0.36	N	*
β_5	Student.Band.Mid	-0.37	0.03	-0.43	-0.32	N	*
β_6	Student.Band.Top	-0.71	0.05	-0.81	-0.62	N	*
α_{11}	School.Intake.Score	0.14	0.21	-0.27	0.56	Y	
α_{12}	School.Boys	0.00	0.21	-0.41	0.40	Y	
α_{13}	School.Girls	-0.02	0.16	-0.34	0.30	Y	
σ_1^2	var_u1	0.33	0.06	0.24	0.46	N	*
σ_y^2	var_y	0.59	0.01	0.56	0.61	N	*

- The DIC rises dramatically, indicating that random intercept may be needed.
- Let the submodel with random and fixed intercept, without school effect predictor be

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \cdots + \beta_6 x_{6ij} + \varepsilon_{ij},$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(0, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Intercepts, without School Effect Predictor: DIC = 9033.65

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	0.15	0.16	-0.18	0.45	Y	
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.23	N	*
β_3	Student.VR.Mid	0.08	0.19	-0.28	0.44	Y	
β_4	Student.VR.Top	0.24	0.21	-0.16	0.65	Y	
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.86	-0.65	N	*
σ_0^2	var_u0	0.36	0.06	0.26	0.50	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

- Parameter estimates of Student VR Band have been shown to be insignificant in many models. Consider removing it as a predictor.

- Let the submodel with random and fixed intercept, without student VR band be

$$
\begin{aligned}
y_{ij} &= (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \varepsilon_{ij} \\
\mu_{0j} &= \alpha_{00} + \alpha_{01} w_{1j} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j}
\end{aligned}
$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Intercepts, without Student VR Band: DIC = 9033.75

		Estimate		95% CI		Significance	
	Variable	Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	0.54	0.55	-0.43	1.69	Y	
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.22	N	*
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.86	-0.65	N	*
α_{00}	u_Int	-0.36	0.55	-1.51	0.64	Y	
α_{01}	School.Intake.Score	0.29	0.22	-0.14	0.71	Y	
α_{02}	School.Boys	0.19	0.21	-0.23	0.61	Y	
α_{03}	School.Girls	0.15	0.17	-0.18	0.48	Y	
σ_0^2	var_u0	0.36	0.06	0.26	0.50	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

- For experimental purpose, consider removing Student Intake Band as a predictor.

- Let the submodel with random and fixed intercept, without student VR and intake band be

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$$
$$\mu_{0j} = \alpha_{00} + \alpha_{01} w_{1j} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j}$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Intercepts, without Student VR and Intake Band: DIC = 9248.36

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	0.02	0.48	-0.87	0.89	Y	
β_1	LR.Test.Score	0.56	0.01	0.53	0.58	N	*
β_2	Student.Gender	0.17	0.03	0.10	0.24	N	*
α_{00}	u_Int	-0.18	0.49	-1.07	0.73	Y	
α_{01}	School.Intake.Score	0.31	0.22	-0.12	0.74	Y	
α_{02}	School.Boys	0.19	0.22	-0.24	0.62	Y	
α_{03}	School.Girls	0.14	0.17	-0.19	0.48	Y	
σ_0^2	var_u0	0.36	0.06	0.26	0.50	N	*
σ_y^2	var_y	0.57	0.01	0.54	0.59	N	*

3.4 Submodel with Intercepts, without Student Gender and VR

- The DIC rises dramatically, indicating that Student Intake Band may be needed. For experimental purpose, consider removing Student Gender and VR Band as predictors.

- Let the submodel with random and fixed intercept, without student gender and VR band be

$$
\begin{aligned}
y_{ij} &= (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \varepsilon_{ij} \\
\mu_{0j} &= \alpha_{00} + \alpha_{01} w_{1j} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j}
\end{aligned}
$$

  where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Intercepts, without Student Gender and VR Band: DIC = 9055.27

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	1.07	0.45	-0.16	1.76	Y	
β_1	LR.Test.Score	0.39	0.02	0.36	0.42	N	*
β_5	Student.Band.Mid	-0.41	0.03	-0.48	-0.35	N	*
β_6	Student.Band.Top	-0.76	0.05	-0.86	-0.65	N	*
α_{00}	u_Int	-0.80	0.46	-1.51	0.42	Y	
α_{01}	School.Intake.Score	0.29	0.22	-0.13	0.72	Y	
α_{02}	School.Boys	0.10	0.22	-0.33	0.52	Y	
α_{03}	School.Girls	0.22	0.17	-0.11	0.55	Y	
σ_0^2	var_u0	0.36	0.06	0.26	0.50	N	*
σ_y^2	var_y	0.54	0.01	0.52	0.56	N	*

- For experimental purpose, consider removing Student VR Band and School Gender as predictors.

- Let the submodel with random, fixed intercept, without student VR band, school gender be

$$
\begin{aligned}
y_{ij} &= (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \varepsilon_{ij} \\
\mu_{0j} &= \alpha_{00} + \alpha_{01} w_{1j}
\end{aligned}
$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Intercepts, without Student VR, School Gender: DIC = 9033.53

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	-0.09	0.52	-1.20	0.67	Y	
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.23	N	*
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.86	-0.65	N	*
α_{00}	u_Int	0.35	0.52	-0.43	1.45	Y	
α_{01}	School.Intake.Score	0.28	0.22	-0.15	0.70	Y	
σ_0^2	var_u0	0.36	0.06	0.26	0.49	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

3.6 Submodel with Intercepts, without Student VR, School Intake

- For experimental purpose, consider removing Student VR Band and School Intake Score as predictors.
- Let the submodel with random, fixed intercept, without student VR, school intake score be

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \varepsilon_{ij}$$
$$\mu_{0j} = \alpha_{00} + \alpha_{02} w_{2j} + \alpha_{03} w_{3j}$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(\mu_{0j}, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Intercepts, without Student VR, School Intake: DIC = 9033.55

	Variable	Estimate		95% CI		Significance	
		Mean	SE	2.5%	97.5%	Contain 0	Significant
β_0	y_Int	0.84	0.36	0.05	1.45	N	*
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.23	N	*
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.86	-0.65	N	*
α_{00}	u_Int	-0.66	0.36	-1.30	0.15	Y	
α_{02}	School.Boys	0.16	0.21	-0.26	0.58	Y	
α_{03}	School.Girls	0.14	0.17	-0.19	0.48	Y	
σ_0^2	var_u0	0.36	0.06	0.26	0.50	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

- Parameter estimates of school effect predictors have been shown to be insignificant in many models. Consider removing all of them while keeping all student level predictors.

- Let the submodel with random intercept and without school effect predictor be

$$y_{ij} = u_{0j} + \beta_1 x_{1ij} + \cdots + \beta_6 x_{6ij} + \varepsilon_{ij}$$

where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(0, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Submodel with Random Intercept, without School Level Predictor: DIC = 9033.34

		Estimate		95% CI		Significance	
	Variable	Mean	SE	2.5%	97.5%	Contain 0	Significant
β_1	LR.Test.Score	0.39	0.02	0.35	0.42	N	*
β_2	Student.Gender	0.16	0.03	0.10	0.23	N	*
β_3	Student.VR.Mid	0.21	0.11	0.00	0.42	N	*
β_4	Student.VR.Top	0.37	0.15	0.08	0.66	N	*
β_5	Student.Band.Mid	-0.41	0.03	-0.47	-0.35	N	*
β_6	Student.Band.Top	-0.75	0.05	-0.85	-0.64	N	*
σ_0^2	var_u0	0.36	0.06	0.26	0.50	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

4. Recommended Model

- The effect of Student VR Band is rather weak so consider removing it as a predictor.

- Let the model with random intercept, without Student VR band and school effect predictor be

$$y_{ij} = u_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \varepsilon_{ij}$$

  where $j = 1, \ldots, 65$ and $i = 1, \ldots, n_j$, and assume that $u_{0j} \sim N(0, \sigma_0^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_y^2)$.

Results

- Model with Random Intercept, without Student VR Band and School Effect Predictor:
  DIC = 9034.77

- It is recommended because DIC value is close to the best value obtained before. Moreover, it is a parsimonious model that aims to explain phenomena using fewer parameters.

		Estimate		95% CI		Significance	
	Variable	Mean	SE	2.5%	97.5%	Contain 0	Significant
β_1	LR.Test.Score	0.40	0.02	0.36	0.43	N	*
β_2	Student.Gender	0.18	0.03	0.12	0.25	N	*
β_5	Student.Band.Mid	-0.38	0.03	-0.44	-0.33	N	*
β_6	Student.Band.Top	-0.71	0.05	-0.82	-0.61	N	*
σ_0^2	var_u0	0.40	0.07	0.29	0.54	N	*
σ_y^2	var_y	0.54	0.01	0.51	0.56	N	*

Running Mean Plots

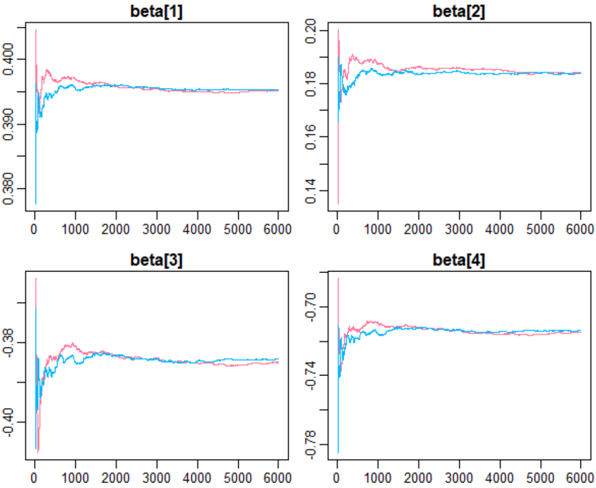

Figure 7: Running Mean Plots of β_1, β_2, β_5, β_6

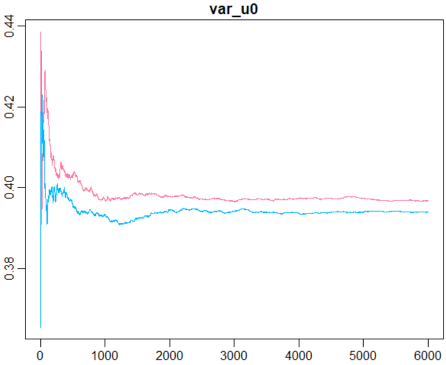

Figure 8: Running Mean Plots of σ_0^2

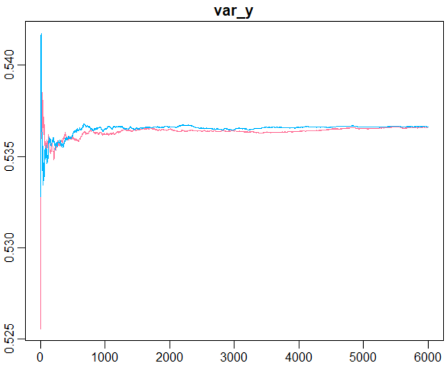

Figure 9: Running Mean Plots of σ_y^2

Interpretation of Recommended Model

- $\sigma_0^2 = 0.40$ (0.07) – Multilevel model is useful since there are random school/cluster effects.

- $\sigma_y^2 = 0.54$ (0.01) – There are student-specific random errors.

- $\beta_1 = 0.40$ (0.02) – The higher the LR Test Score is, the higher the Exam Score is. It is consistent with Figure 5.

- $\beta_2 = 0.18$ (0.03) – Girls perform better than boys do in general. It is consistent with the data that the mean exam score of girls is 0.09 while that of boys is -0.14.

- $\beta_5 = -0.38$ (0.03) and $\beta_6 = -0.71$ (0.05) – The lower the Student Intake Band is, the higher the mean exam score is. It is consistent with Table 4.

- Student Verbal Reasoning Score Band is not important to Exam Score, in the presence of other variables. Possible explanation is that other student-level variables (say LR Test Score) have incorporated the information.

- School Intake Score and School Gender are not important to random school effects, in the presence of other variables. In case of further study, other school-level variables may be needed, e.g. school teaching quality measure.