Cross – Classified Models
Part 2 – Model Fitting

Professor George Leckie
Centre for Multilevel Modelling
University of Bristol

Lecture outline

1. In lecture 1 we introduced cross-classified models and the practical example

2. In this lecture we fit models to the Fife dataset and also the following:

3. Notation and classification diagrams

4. The effect of ignoring cross-classification

5. Variance Partition Coefficients and Intra-Class Correlation Coefficients

6. Including predictor variables in cross-classified models

7. In lecture 3 we will look at extensions to different application areas and more levels of crossing.

Recap: Fife Education Dataset

- Dataset from Fife in Scotland that was used by Rasbash and Goldstein (1994) when they first introduced extensions of the IGLS algorithm for cross-classified models.

- Dataset contains 3,435 students from 19 secondary schools and 148 primary schools

- The response is a total attainment score based on national examinations taken at the end of compulsory schooling (age 16) which ranges from 1 to 10

- Dataset has several student level predictors (mother and father education, verbal reasoning intake score, choice of school, social class, gender) that can be used to explain variation in attainment.

Cross-classified model

- Model 1 is a cross-classified **variance-components** model

$$\mathbf{attain}_i = \beta_0 + u^{(3)}_{\mathbf{sid}(i)} + u^{(2)}_{\mathbf{pid}(i)} + e_i$$

$$u^{(3)}_{\mathbf{sid}(i)} \sim \mathrm{N}\left(0, \sigma^2_{u(3)}\right)$$

$$u^{(2)}_{\mathbf{pid}(i)} \sim \mathrm{N}\left(0, \sigma^2_{u(2)}\right)$$

$$e_i \sim \mathrm{N}(0, \sigma^2_e)$$

- We have written the model in **classification notation** as **standard hierarchical notation** (ijk) breaks down in non-hierarchical models
  - ijk implies unit i is nested in cluster j in supercluster k
  - Primary and Secondary Schools are both conceptually at level 2

Classification notation

- In classification notation, the index i uniquely identifies the lowest level units
- The (2) and (3) superscripts distinguish the different higher classifications
- The $\mathbf{sid}(i)$ and $\mathbf{pid}(i)$ subscripts are **classification functions** which return the secondary school attended and primary school attended by student i
- So if student 1 attends secondary school 5 having attended primary school 17 we would have

$$\mathbf{sid}(1) = 5, \qquad \mathbf{pid}(1) = 17$$

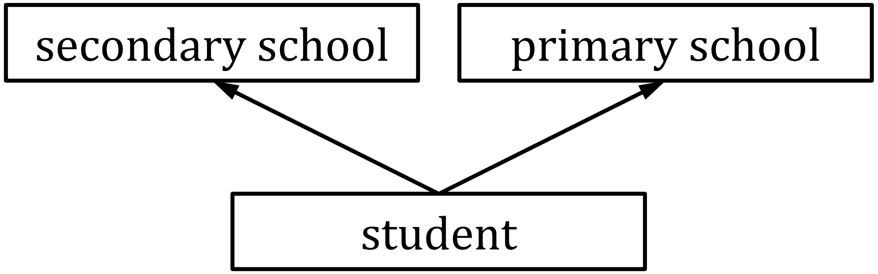
and the model equation for student 1 would be written as

$$\mathbf{attain}_1 = \beta_0 + u_5^{(3)} + u_{17}^{(2)} + e_1$$

Classification diagrams

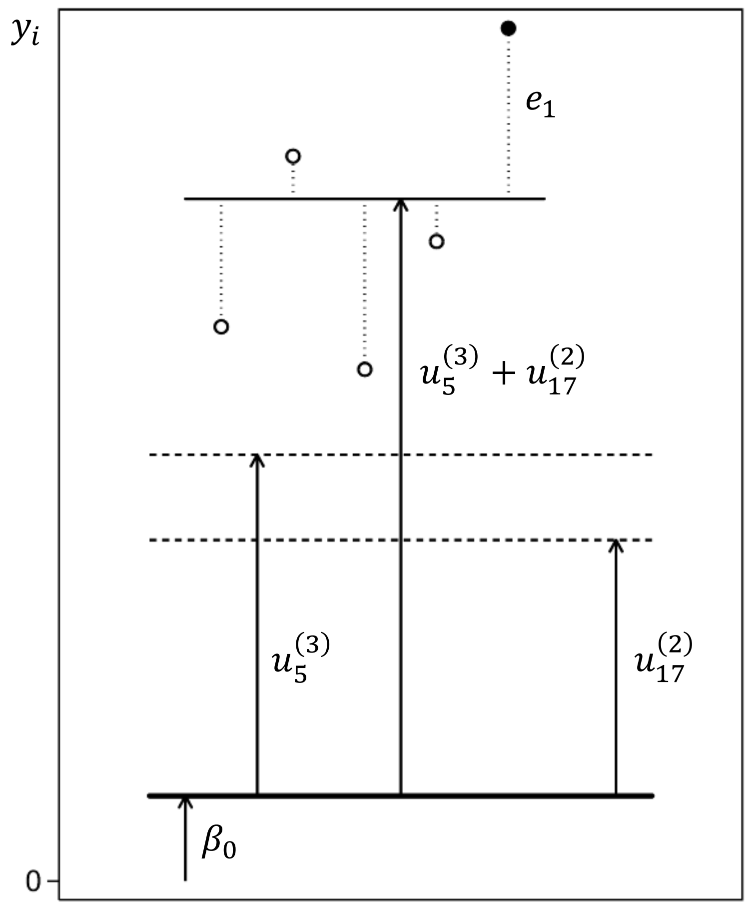
- Classification notation provides no information on the hierarchical, crossed or multiple membership structures present in the data

- Models should therefore be presented with **classification diagrams** which provide simple summaries of data structures (c.f., units diagrams)



- Classification diagrams have one node for each classification in the model
  - Single arrows indicate nested structures
  - Unconnected nodes indicate crossed structures

Illustration of the cross-classified model



- Consider again the model equation for student 1

$$\textbf{attain}_1 = \beta_0 + u_5^{(3)} + u_{17}^{(2)} + e_1$$

- Note how the secondary and primary school effects are assumed additive

- It is possible to add a random interaction effect classification, but we will not explore this here

Implications of number of units at each classification

- There are 19 secondary schools with an average of 181 students per school
  - We will not reliably estimate the between-secondary variance $\sigma^2_{u(3)}$
  - We will reliably estimate the individual secondary school effects $u^{(3)}_{\mathbf{sid}(i)}$
- There are 148 primary schools with an average of 23 students per school
  - We will reliable estimate the between-primary variance $\sigma^2_{u(2)}$
  - We will estimate the individual primary school effects $u^{(2)}_{\mathbf{pri}(i)}$ less reliably than we do the secondary school effects

Cross-classified model

- Model 1 is a single-level model with no covariates
- Model 2 is the cross-classified variance-components model

		Model 1		Model 2	
	Parameter	**Estimate**	**Std. Err.**	**Estimate**	**Std. Err.**
β_0	Intercept	5.679	0.052	5.504	0.190
$\sigma^2_{u(3)}$	Secondary variance	–	–	0.414	0.210
$\sigma^2_{u(2)}$	Primary variance	–	–	1.153	0.215
σ^2_e	Student variance	9.363	0.227	8.120	0.202
	DIC	17431.3		17047.7	

- Model 2 fits the data significantly better than Model 1 (DIC drops by 383.6)
- Approximately 16% of the variation is due to schooling groups.

Can we ignore the primary school effects?

- Model 3 is the two-level students-within-schools variance-components model

		Model 2		Model 3	
	Parameter	**Estimate**	**Std. Err.**	**Estimate**	**Std. Err.**
β_0	Intercept	5.504	0.190	5.608	0.166
$\sigma^2_{u(3)}$	Secondary variance	0.414	0.210	0.489	0.210
$\sigma^2_{u(2)}$	Primary variance	1.153	0.215	–	–
σ^2_e	Student variance	8.120	0.202	8.989	0.219
	DIC	17047.7		17310.0	

- Ignoring the primary school effects increases the DIC by 262.3
  - The primary school effects are significant, *even after adjusting for secondary schools*
  - Students who attended the same primary are significantly more alike than students from different primaries *and this is not simply because children from the same primary typically attend the same secondary school*

Can we ignore the secondary school effects?

- Model 4 is the two-level students-within-neighbourhoods model

		Model 2		Model 4	
	Parameter	Estimate	Std. Err.	Estimate	Std. Err.
β_0	Intercept	5.504	0.190	5.622	0.113
$\sigma^2_{u(3)}$	Secondary variance	0.414	0.210	–	–
$\sigma^2_{u(2)}$	Primary variance	1.153	0.215	1.248	0.215
σ^2_e	Student variance	8.120	0.202	8.203	0.202
	DIC	17047.7		17080.5	

- Ignoring the school effects increases the DIC diagnostic by 32.8
  - The secondary effects are significant, *even after adjusting for the primary previously attended*
  - Students from the same secondary are significantly more alike than students from different secondaries *and this is not simply because children from the same school are more likely to have studied in the same primaries*

VPCs and ICCs

- Reconsider the Model 2 results. The secondary school-level VPC (and ICC) is

$$\text{VPC}^{(3)} \equiv \text{ICC}^{(3)} = \frac{\sigma^2_{u(3)}}{\sigma^2_{u(3)} + \sigma^2_{u(2)} + \sigma^2_e} = \frac{0.414}{0.414 + 1.153 + 8.120} = 0.043$$

- The primary school-level VPC (and ICC) is

$$\text{VPC}^{(2)} \equiv \text{ICC}^{(2)} = \frac{\sigma^2_{u(2)}}{\sigma^2_{u(3)} + \sigma^2_{u(2)} + \sigma^2_e} = \frac{1.153}{0.414 + 1.153 + 8.120} = 0.119$$

- The secondary and primary school combined VPC (and ICC) is

$$\text{VPC}^{(2,3)} \equiv \text{ICC}^{(2,3)} = \frac{\sigma^2_{u(3)} + \sigma^2_{u(2)}}{\sigma^2_{u(3)} + \sigma^2_{u(2)} + \sigma^2_e} = \frac{0.414 + 1.153}{0.414 + 1.153 + 8.120} = 0.162$$

- There are stronger educational disparities across the 148 primary schools than there are across the 19 secondary schools

Including level-1 covariates

- As you will see in the practical there are several predictor variables and we will here include a prior test (VRQ), social class, father's and mother's education and the choice of school (where 1 is 1st etc.)

$$\mathbf{attain}_i = \beta_0 + \beta_1\mathbf{vrq}_i + \beta_2\mathbf{sc}_i + \beta_3\mathbf{fed}_i + \beta_4\mathbf{med}_i + \beta_5\mathbf{choice}_i + u^{(3)}_{\mathbf{sid}(i)} + u^{(2)}_{\mathbf{pid}(i)} + e_i$$

$$u^{(3)}_{\mathbf{sid}(i)} \sim \mathrm{N}\left(0, \sigma^2_{u(3)}\right)$$

$$u^{(2)}_{\mathbf{pid}(i)} \sim \mathrm{N}\left(0, \sigma^2_{u(2)}\right)$$

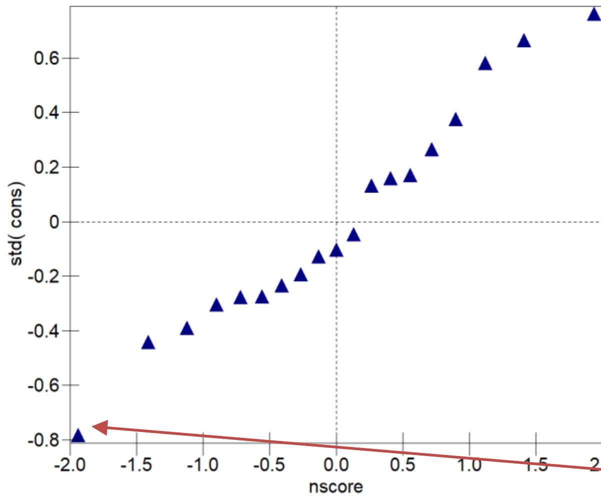
$$e_i \sim \mathrm{N}(0, \sigma^2_e)$$

Including level-1 covariates (cont'd)

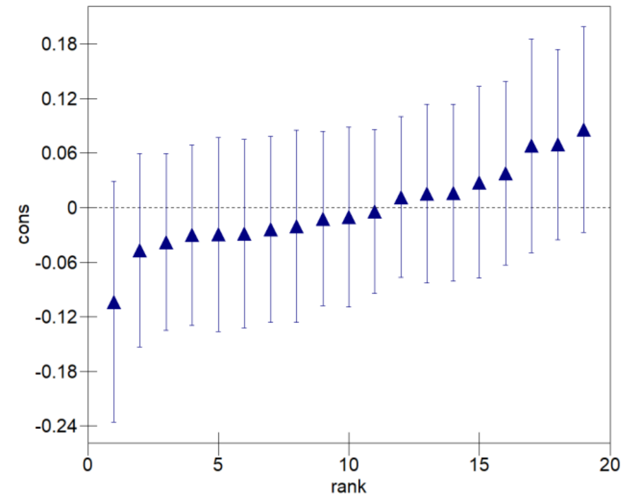
		Model 2		Model 5	
	Parameter	Estimate	Std. Err.	Estimate	Std. Err.
β_0	Intercept	5.504	0.190	-9.726	0.294
β_1	VRQ score	–	–	0.155	0.003
β_2	Social Class	–	–	0.027	0.003
β_3	Father's Education	–	–	0.215	0.092
β_4	Mother's Education	–	–	0.219	0.086
β_5	Choice	–	–	-0.118	0.054
$\sigma^2_{u(3)}$	Secondary variance	0.414	0.210	0.014	0.017
$\sigma^2_{u(2)}$	Primary variance	1.153	0.215	0.206	0.052
σ^2_{e}	Student variance	8.120	0.202	4.171	0.105
	DIC	17047.7		14728.1	

- The covariates explain 97% of the secondary school variance in raw attainment, 82% of the primary variance and 49% of the student variance
- This suggests that the wide educational and social inequalities between secondary and primary schools can be largely explained by the predictors

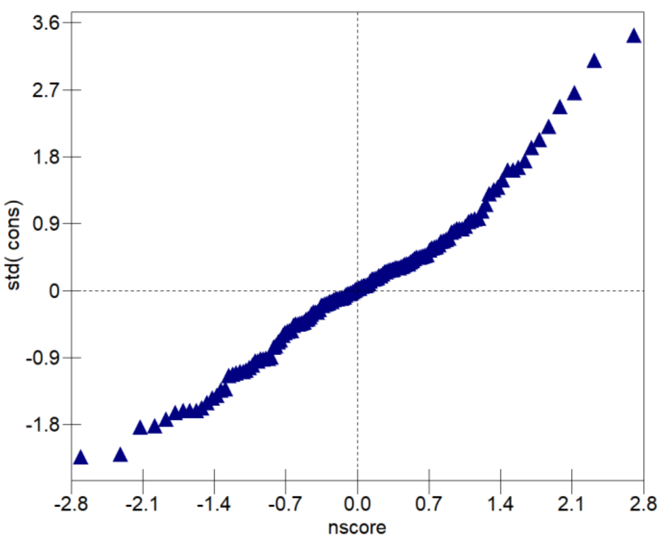
Examining random effects: Q-Q plots and caterpillar plots



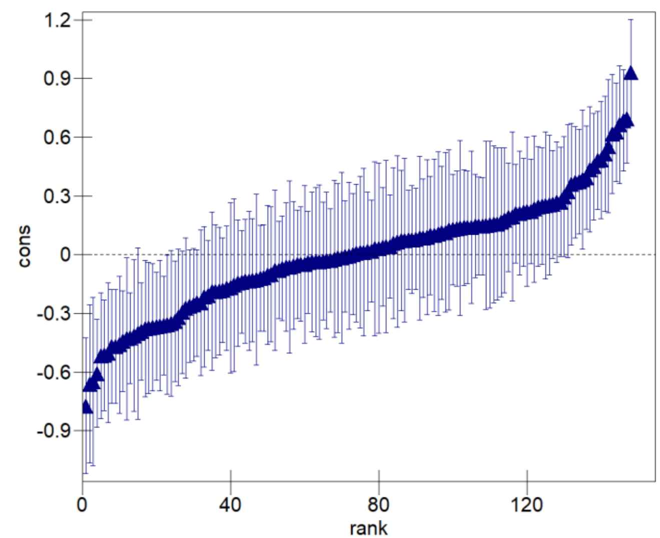
Secondary Schools



Note outlier school 19



Primary Schools



Dummying out secondary school 19

- In the last model we can see that most of the secondary school variance is explained and in fact 1 school (school 19) is an outlier. We could therefore try and replace the secondary random effects with 1 fixed effect dummy for that school:

$$\mathbf{attain}_i = \beta_0 + \beta_1\mathbf{vrq}_i + \beta_2\mathbf{sc}_i + \beta_3\mathbf{fed}_i + \beta_4\mathbf{med}_i + \beta_5\mathbf{choice}_i$$
$$+\beta_6\mathbf{school19}_i + u^{(2)}_{\mathbf{pid}(i)} + e_i$$

$$u^{(2)}_{\mathbf{pid}(i)} \sim \mathrm{N}\left(0, \sigma^2_{u(2)}\right)$$

$$e_i \sim \mathrm{N}(0, \sigma^2_e)$$

- Note here that removing the secondary school effects means this is no longer a cross-classified model! The model is a two-level model.

Dummying out school 19 (cont'd)

	Parameter	Model 5		Model 6	
		Estimate	Std. Err.	Estimate	Std. Err.
β_0	Intercept	-9.726	0.294	-9.646	0.290
β_1	VRQ score	0.155	0.003	0.155	0.003
β_2	Social Class	0.027	0.003	0.027	0.003
β_3	Father's Education	0.215	0.092	0.214	0.092
β_4	Mother's Education	0.219	0.086	0.225	0.086
β_5	Choice	-0.118	0.054	-0.124	0.056
β_6	School 19	–	–	-0.632	0.244
$\sigma^2_{u(3)}$	Secondary variance	0.014	0.017	–	–
$\sigma^2_{u(2)}$	Primary variance	0.206	0.052	0.209	0.053
σ^2_e	Student variance	4.171	0.105	4.168	0.103
	DIC	14728.1		14724.7	

- Here we see a significant negative effect for school 19 and a reduction in DIC so a marginal better (as more parsimonious) model.

Summary

- In this lecture we have fitted several cross-classified models to the Fife dataset
- We have introduced new notation and diagrams to represent cross-classified structures
- We have shown how to extend the ICC and VPC residual clustering statistics to cross-classified models
- We have fitted predictor variables to the model and shown how they explain different sources of variation.
- We have shown that in this example due to an outlying secondary school, if we dummy out this school we can collapse our model to a two-level nested model.
- In the third lecture we will look at further applications of cross-classified models that illustrate new extensions.