PSY9185 - Multilevel models

Cross-classified data

Espen Moen Eilertsen

2021-04-29

Hierarchical data

The multilevel models discussed so far have dealt with hierarchical data structures where units are classified by some factor into higher level clusters, which may again be classified by some other factor into higher level clusters

- Students nested in classes which are again nested in schools

The classification factors (`student`, `class`, `school`) are **nested** in the sense that a lower-level cluster can only belong to one higher level cluster

- A `student` can only belong to one `class` which can only belong to one `school`

Cross-classified data

When units are classified by multiple grouping factors that cannot be arranged into hierarchies, the data is instead **cross-classified**

- Students cross-classified by primary and secondary school
- Patients cross-classified by doctors and nurses
- Reaction times cross-classified by experimental condition and subject
- Children cross-classified by mothers and fathers
- Test scores cross-classified by students and items

In designed experiments the factors `condition` and `subject` can often be *fully* crossed as every subject receive all conditions, whereas the factors `Student` and `teacher` can often be *partially* crossed as not every student have every teacher

Observational designs will often have partially crossed data, whereas experimental designs will often have fully crossed data

Cross-classified data

Grouping factor **A** is *nested* within grouping factor **B** if each *level* of **A** occurs within only one *level* of **B** - if not, they are *crossed*

If all *levels* of grouping factor **A** occurs within all *levels* of grouping factor **B**, the factors are *fully* crossed, if not, they are *partially* crossed

Nested		Fully crossed		Partially crossed	
A	**B**	**A**	**B**	**A**	**B**
1	1	1	1	1	1
2	1	2	1	2	1
3	1	3	1	3	1
4	2	1	2	1	2
5	2	2	2	2	2
6	2	3	2	1	3

A split-plot design

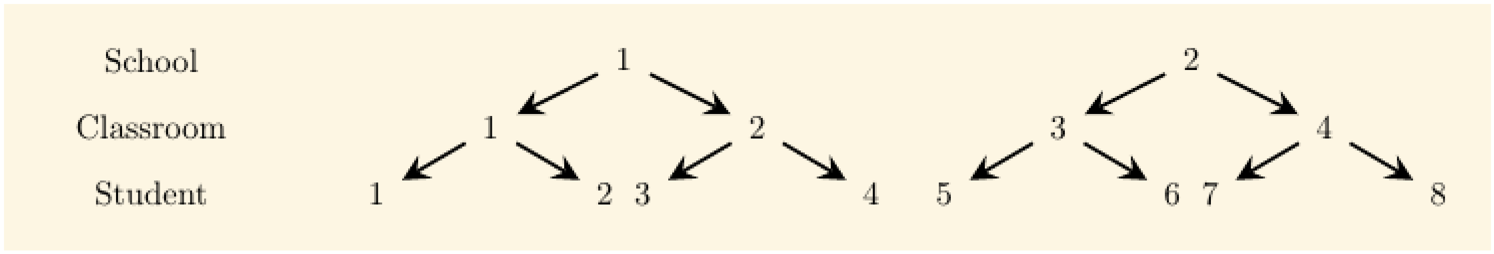
All **subjects** (1-4) received both levels (A, B) of experimental factor **F1** (within-subjects factor). Half of the subjects received level A of experimental factor **F2**, the other half received level B (between-subjects factor)

F1	F2	subject
A	A	1
B	A	1
A	A	2
B	A	2
A	B	3
B	B	3
A	B	4
B	B	4

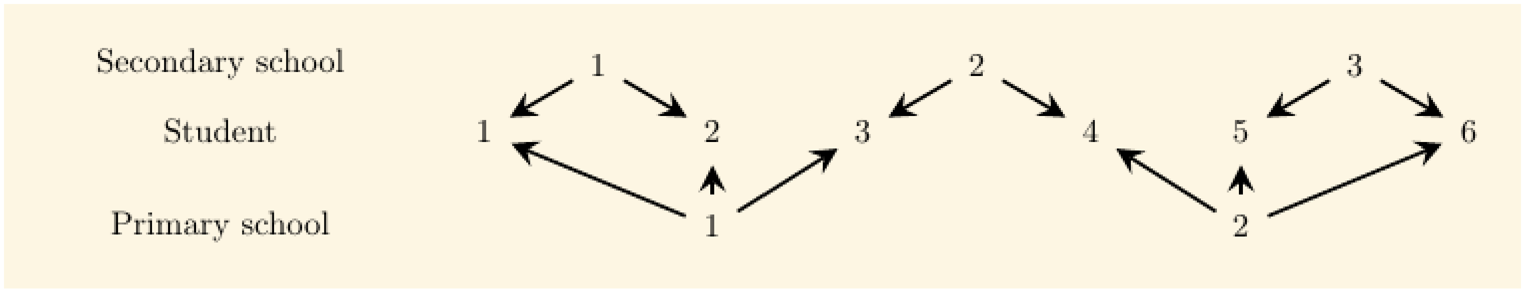
```
    F2
F1   A B
   A 2 2
   B 2 2

    subject
F1   1 2 3 4
   A 1 1 1 1
   B 1 1 1 1

    subject
F2   1 2 3 4
   A 2 2 0 0
   B 0 0 2 2
```

Cross-classified data

Nested data structure



Cross-classified data structure



- Students from the same primary school can go to different secondary schools and students from the same secondary school can come from different primary schools

Exercises 1

Open the file `day3_cross_classified_template.R`. We will work with data from the MLwiN program that consist of test scores from students at age 16. Each row represents a student

- `attain` - educational attainment score
- `pid` - primary school identifier
- `sid` - secondary school identifier

1. How many students, primary schools and secondary schools are in the data? (hint: `unique()`, `length()`)
2. Cross-tabulate the grouping factors primary and secondary school. What does the table show us? (hint: `table(F1, F2)`)

Multilevel models for cross-classified data

If the classification factors may contribute to the outcome that is under study, multilevel models can be used to model those effects - both primary and secondary school may influence educational achievement

Multilevel model for achievements y_{ijk} for student i from secondary school j and primary school k

$$y_{ijk} = \beta + \eta_{1j} + \eta_{2k} + \epsilon_{ijk}$$

- β is a fixed intercept
- η_{1j} is a random intercept for classification factor secondary school j
- η_{2k} is a random intercept for classification factor primary school k
- ϵ_{ijk} is the residual error for each student

Multilevel models for cross-classified data

Similar to other multilevel models, individual covariates can be added, possibly with random coefficients varying over primary and/or secondary school. The random coefficients can be explained by primary and/or secondary school variables

Random intercepts and slopes for secondary school and random intercepts for primary school

$$y_{ijk} = \beta_1 + \beta_2 x_{ijk} + \eta_{1j} + \eta_{2j} x_{ijk} + \eta_{3k} + \epsilon_{ijk}$$

Random intercepts for primary and secondary school with covariate for primary school

$$y_{ijk} = \beta_1 + \beta_2 x_k + \eta_{1j} + \eta_{2k} + \epsilon_{ijk}$$

lme4 with crossed random effects

Many estimation methods/software packages are restricted to models with nested random effects

This is not the case with **lme4**. Crossed random effects are specified the same way as nested random effects

```
lmer(y ~ 1 + (1|factor1) + (1|factor2))
```

lme4 doesn't need to know whether factors are crossed or nested - that is a property of the data

Warning

Don't code data like this

score	student	school
0.3	1	1
-0.5	1	1
0.1	2	1
1.4	2	1
0.9	1	2
0.6	1	2
-0.2	2	2
0.2	2	2

```
lmer(score ~ 1 + (1|student) + (1|school))
```

student and school are crossed

if what you mean is this

score	student	school
0.3	1	1
-0.5	1	1
0.1	2	1
1.4	2	1
0.9	3	2
0.6	3	2
-0.2	4	2
0.2	4	2

```
lmer(score ~ 1 + (1|student) + (1|school))
```

student is nested in school

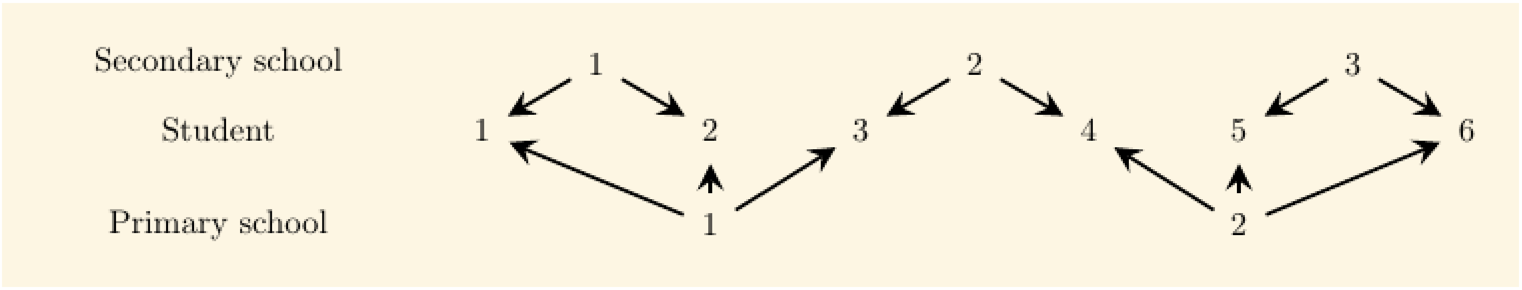
Exercises 2

Continue working with the file `day3_cross_classified_template.R` with test scores from students at age 16

- `attain` - educational attainment score
- `pid` - primary school identifier
- `sid` - secondary school identifier

1. Fit a model for educational attainment with a fixed intercept and random intercepts for both primary and secondary school
2. Does primary or secondary school appear to be most important for attainment?

Dependence among responses



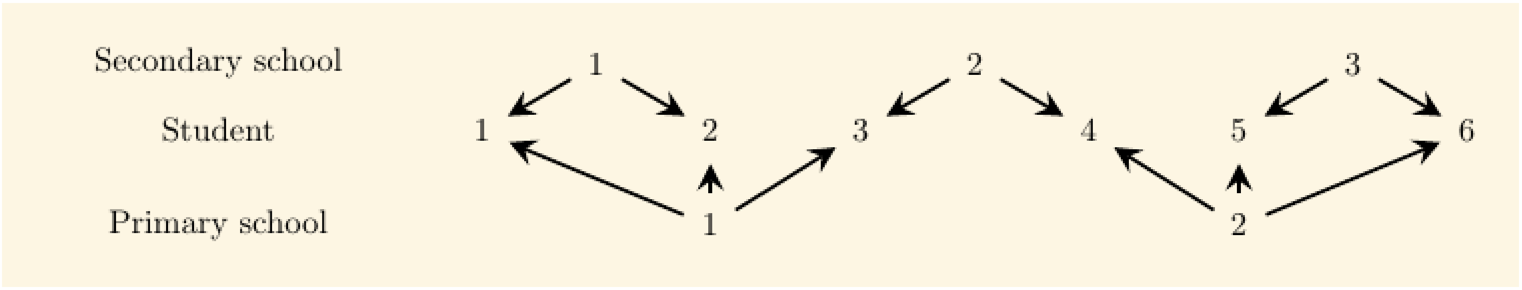
General formulation $y_{ijk} = \beta + \eta_{1j} + \eta_{2k} + \epsilon_{ijk}$

student 1 $y_{111} = \beta + \eta_{11} + \eta_{21} + \epsilon_{111}$

student 2 $y_{211} = \beta + \eta_{11} + \eta_{21} + \epsilon_{211}$

student 3 $y_{321} = \beta + \eta_{12} + \eta_{21} + \epsilon_{321}$

Dependence among responses



Denote the variance of the random intercepts for secondary school with ψ_1, the variance for the random intercepts for primary school with ψ_2, and the residual variance with θ. The total variance for any student is

$$\mathrm{Var}(y_{ijk}) = \psi_1 + \psi_2 + \theta$$

Covariance between student 1 and 2

$$\mathrm{Cov}(y_{111}, y_{211}) = \psi_1 + \psi_2$$

Covariance between student 1 and 3

$$\mathrm{Cov}(y_{111}, y_{321}) = \psi_2$$

Intraclass correlations

Intraclass correlations measures degree of dependence between units because they belong to the same group

The correlation between students attending the same primary but different secondary school is

$$\frac{\psi_2}{\psi_1 + \psi_2 + \theta}$$

Students attending the same secondary but different primary school

$$\frac{\psi_1}{\psi_1 + \psi_2 + \theta}$$

Students attending the same primary *and* secondary school

$$\frac{\psi_1 + \psi_2}{\psi_1 + \psi_2 + \theta}$$

Exercises 3

Continue working with the file `day3_cross_classified_template.R` with test scores from students at age 16

- `attain` - educational attainment score
- `pid` - primary school identifier
- `sid` - secondary school identifier

Use the model from exercise 2

1. Compute the intraclass correlation for students from the same primary but different secondary school
2. Compute the intraclass correlation for students from the same secondary school but different primary school
3. Compute the intraclass correlation for students from the same primary *and* secondary school

References

Bates, D. M. (2010). *lme4: Mixed-effects modeling with R*. Springer New York.

Goldstein, H. (2011). *Multilevel statistical models*. Vol. 922. John Wiley & Sons.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.