Applied Statistical Modelling - Lecture 7: multilevel models

Prof.ssa Chiara Seghieri

Laboratorio di Management e Sanità, Istituto di Management, L'EMbeDS
Scuola Superiore Sant'Anna, Pisa
c.seghieri@santannapisa.it
c.tortu@santannapisa.it

What are multilevel data and multilevel analysis?

What are multilevel data?

- Multilevel data are data where observations are clustered in units
- Observations within the same unit may be more similar than observations in separate units, on average
    - What effect does this have on estimation and statistical inference?

Examples of multilevel data with contextual clustering

- Observations of students, clustered within schools
- Observations of siblings, clustered within families
- Observations of individuals, clustered within countries, states, or neighborhoods
- Patients within hospitals
- Repeated test scores, clustered within students (multilevel data with intra-person clustering)

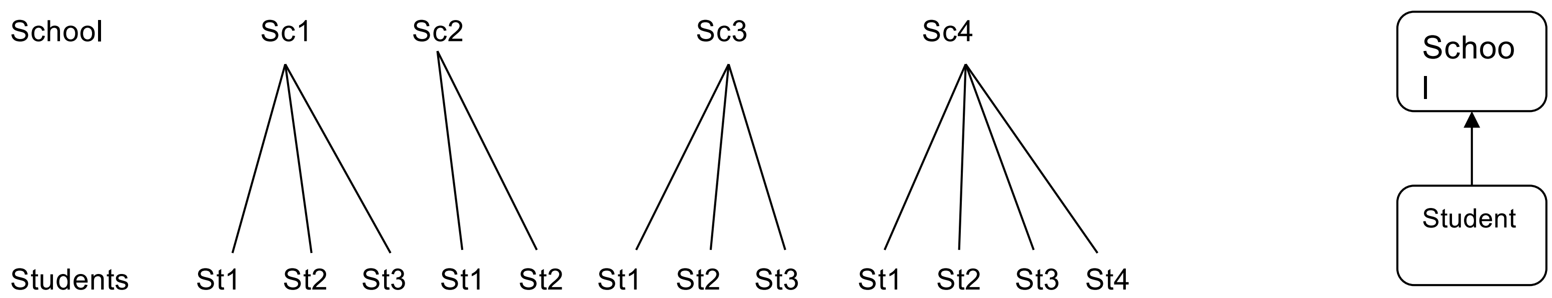
Nested Data

- Data nested within a group tend to be more alike than data from individuals selected at random.

- Nature of group dynamics will tend to exert an effect on individuals.

Two-level hierarchical structures

Students within schools



Students within a school are more alike than a random sample of students. This is the 'clustering' effect of schools.

Nested Data

- Intraclass correlation (ICC) provides a measure of the clustering and dependence of the data

0 (very independent) to 1.0 (very dependent)

Details discussed later

Why Multilevel Modeling vs. Traditional Approaches?

Traditional Approaches – 1-Level

1. Individual level analysis (ignore group)

2. Group level analysis (aggregate data and ignore individuals)

Problems with Traditional Approaches

1. Individual level analysis (ignore group)

   Violation of <u>independence</u> of data assumption leading to misestimated standard errors (standard errors are smaller than they should be).

Problems with Traditional Approaches

1. Group level analysis

   (aggregate data and ignore individuals)

   Aggregation bias = the meaning of a variable at Level-1 (e.g., individual level SES) may not be the same as the meaning at Level-2 (e.g., school level SES)

Multilevel Approach

- 2 or more levels can be considered simultaneously
- Can analyze within- and between-group variability

Multilevel regression models

- Also called
  - Hierarchical Linear Models
  - Mixed Models
  - Multilevel Models
  - Growth Models
  - Slopes-as-Outcomes Models

Multilevel Regression Models

- A form of regression models
- Used to answer questions about the relationship of context to individual outcomes
- Used to estimate both within-unit and between-unit relationships (and cross-level interactions)
  - e.g., within- vs. between-school relationships between SES and achievement

Data frame for student within school example

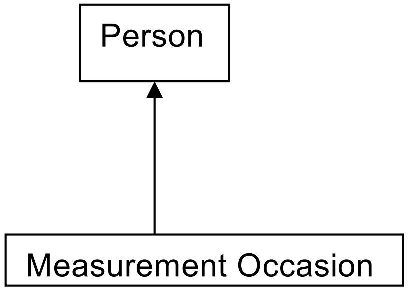
Classifications or levels		**Respons e**	**Explanatory variables**		
Studen t *i*	*School* *j*	*Student Exam* $score_{ij}$	*Student previous Examination* $score_{ij}$	*Student* $gender_{ij}$	*School* $type_j$
1	1	75	56	M	State
2	1	71	45	M	State
3	1	91	72	F	State
1	2	68	49	F	Private
2	2	37	36	M	Private
3	2	67	56	M	Private
1	3	82	76	F	State

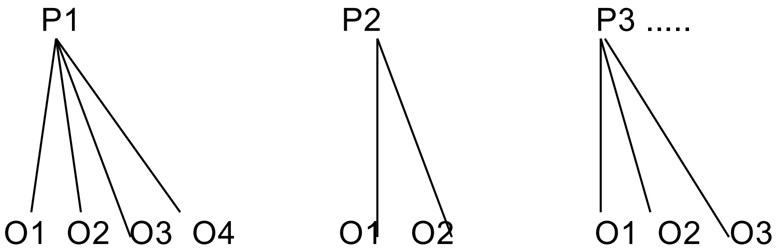
1 Does the gender gap vary across schools?
2 What is the between-school variation in student's progress?

3 Is School X (that is a specific school) different from other schools in the sample in its effect?

4 Are students in public schools less variable in their progress?

Classification, unit diagrams and data frames for **repeated measures structures**.





Person	*H-Occ1*	*H-Occ2*	*H-Occ3*	*Age-Occ1*	*Age-Occ2*	*Age-Occ3*	*Gender*
1	75	85	95	5	6	7	F
2	82	91	*	7	8	*	M
3	88	93	96	5	6	7	F

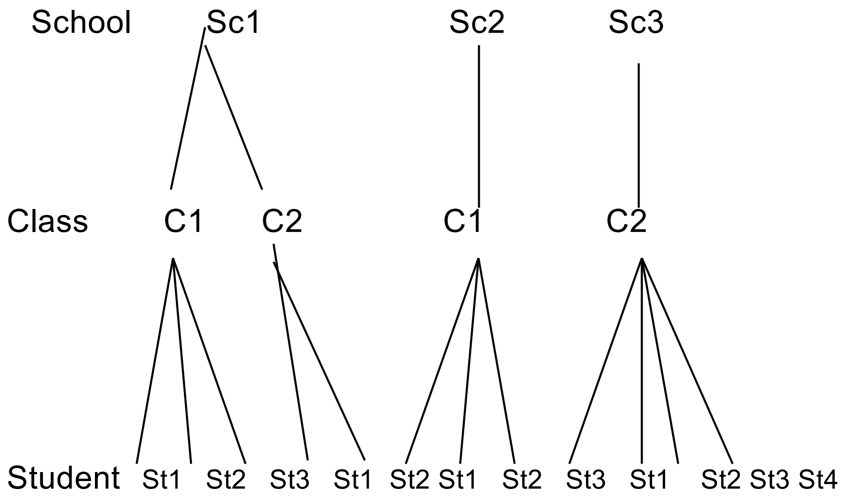
Wide form 1 row per individual

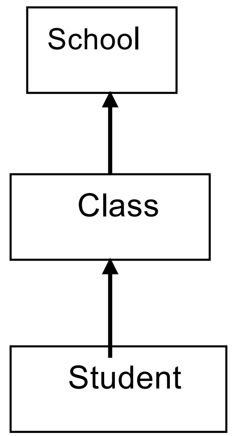
Long form 1 row per occasion(required by *MLwiN*)

Classifications or levels		**Response**	**Explanatory variables**	
Occasion *I*	*Person* *J*	$Height_{ij}$	Age_{ij}	$Gender_j$
1	1	75	5	F
2	1	85	6	F
3	1	95	7	F
1	2	82	7	M
2	2	91	8	M
1	3	88	5	F
2	3	93	6	F
3	3	96	7	F

Three level structures

Students:classes:schools





MLM allow a different number of students in each class and a different number of classes in each school.

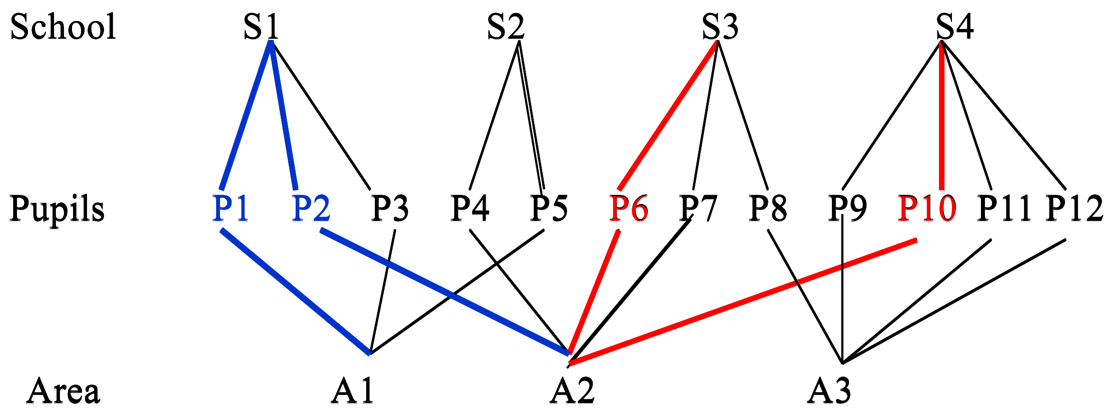
Non-hierarchical structures

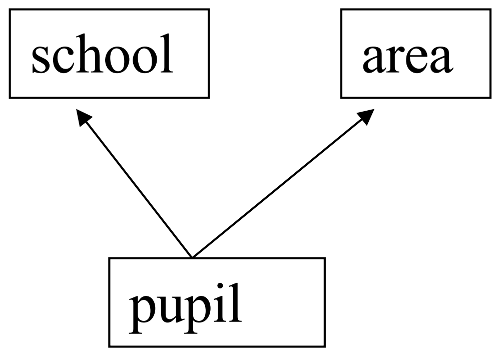
So far all our examples have been exact nesting with lower level units nested in one and only one higher-level unit.

That is we have been dealing with strict hierarchies. But social reality can be more complicated than that.

- Cross-classified structures
- Multiple membership structures

Cross-classified Model





In this structure schools are not nested within areas. For example

Pupils 2 and 3 attend school 1 but come from different areas

Pupils 6 and 10 come from the same area but attend different schools

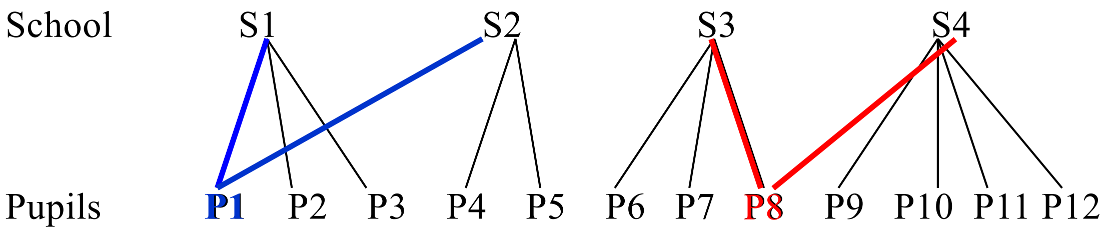
Schools are not nested within areas and areas are not nested within schools. School and area are are cross-classified

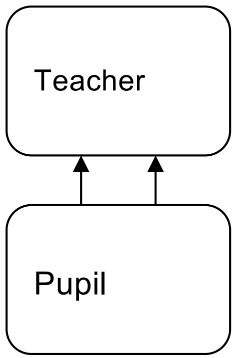
Multiple membership models

Where first level units are seen as nested within more than one unit from a higher level classification :.

Health outcomes where patients are treated by a number of nurses, patients are multiple members of nurses

Students move schools, so some pupils are multiple members of schools.





When do we need multilevel regression?

- In the intermediate case, where knowing the school gives us some, but not complete information about *Y*.
- e.g., test scores vary both within and between schools
- e.g., individuals vary within and between neighborhoods
- e.g., mood varies both within individuals (over time) and between individuals

What we know so far

- Two observations within the same unit may be more similar than two observations chosen at random
- If the regression model does not explain all of the between-unit differences (and it is unlikely that they will), we will have correlated errors within units
- This is a violation of the independence of residuals assumption in OLS
- At a minimum, this results in incorrect standard errors (too small)

How do we allow dependence in the regression model?

- We want a model that explicitly allows the level of the outcome variable to vary across level-two units
- For example, we want to let the mean reading score differ across schools
- So let’s write a model that allows this

Some notation

- *i* indexes level-one units (people within schools, observations within persons)
- *j* indexes level-two units (e.g., schools, if we have students nested within schools)
- We will use *r* to denote a level-one residual, and *u* to denote a level-two residual

Our first multilevel model

- Instead of : $$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}$$

- Let's write:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + r_{ij}$$

What is u_j?

- A residual term
- Specific to unit *j*
- Common to all observations in unit *j*
- Interpretation: the difference between the overall intercept and the intercept in unit *j*

What is r_{ij}?

- A residual term
- Specific to observation *i* in unit *j*
- Has a mean of 0, so any part of ε_{ij} that is common to all observations within *j* has been removed
- So the r_{ij}'s may be independent
- Not guaranteed to be independent

Features of this model

- Note that: $\varepsilon_{ij} = u_j + r_{ij}$
- We also have:

$$
\begin{aligned}
\mathrm{Var}(\varepsilon_{ij}) &= \mathrm{Var}(u_j + r_{ij}) \\
&= \mathrm{Var}(u_j) + \mathrm{Var}(r_{ij}) + 2*\mathrm{Cov}(u_j, r_{ij}) \\
&= \mathrm{Var}(u_j) + \mathrm{Var}(r_{ij})
\end{aligned}
$$

- We will come back to variance decomposition later

Features of this model

- The <u>level</u> of Y_{ij} – after adjusting for X_{ij} – may vary across the units
- We have made no assumptions yet about the distribution of the u_j's or the r_{ij}'s.
- The relationship between X and Y does not depend on j (β_1 does not depend on j)

So how do we estimate this model?

- We want an estimate of β_1 , the relationship between X_{ij} and Y_{ij}.
- Two approaches:
  - Fixed Effects estimator
  - Random Effects estimator

Another way to write this model

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1 X_{ij} + u_j + r_{ij} \\
&= \left(\beta_0 + u_j\right) + \beta_1 X_{ij} + r_{ij} \\
&= \beta_{0j} + \beta_1 X_{ij} + r_{ij}
\end{aligned}
$$

where

$$
\beta_{0j} = \beta_0 + u_j
$$

The fixed effects estimator

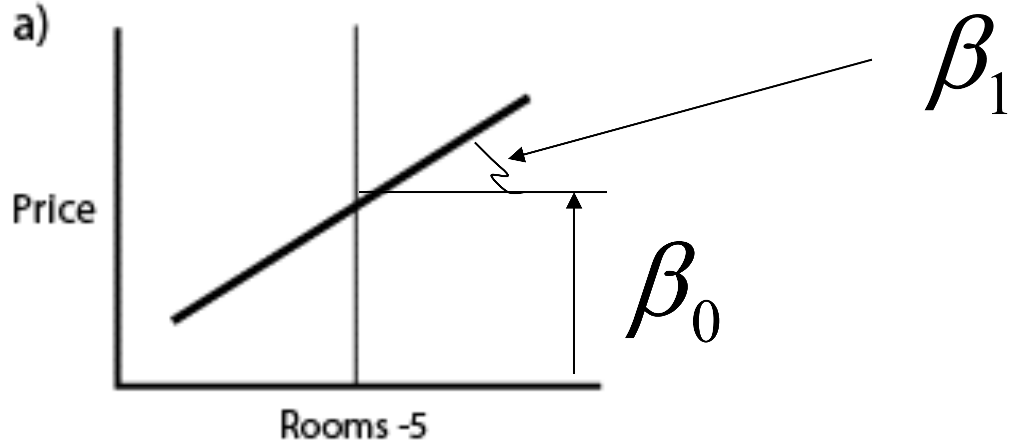
$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + r_{ij}$$

- We have ‘absorbed’ the level-two error terms (the u_j’s) into the intercept
- Now each aggregate unit has its own intercept; so between-unit variation is accounted for in the intercepts
- This solves the dependence problem with the r_{ij}’s (they may still not be independent, but not because of unexplained variation between-level-two units)

- Two level model
- houses at level 1 nested within districts at level 2

- continuous response: price of a house
- predictor: size = number of rooms and this variable has been centred around average size (5)

Simple regression model



y **is the outcome, price of a house**

x_1 **is the predictor, number of rooms, which we shall deviate around its mean**

Simple regression model (cont)

$$y_i = \beta_0 + \beta_1 x_{1i} + (e_i)$$

y_i **is the price of house i**

x_1 **is the individual predictor variable**

β_0 **is the intercept;** β_1 **is the fixed slope term:**

e_i **is the residual/random term, one for every house**

Summarizing the random term: ASSUME IID

Mean of the random term is zero

Constant variability (Homoscedasticy)

No patterning of the residuals (*i.e*, they are independent)

$$e \sim N(0, \sigma_e^2)$$

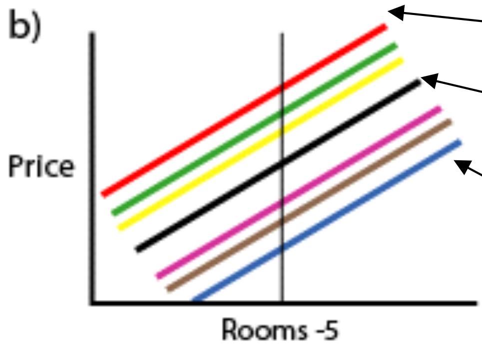
General Structure

- Response = fixed + random

- Specific case: the single level simple regression model

Response	***Systematic Part***				***Random Part***
House = Price	Price of average-sized house	+	Cost of extra room	+	house residual variation
	Intercept		***Slope***		***Residual***

Random intercepts model



Premium $+u_{0j}$

Citywide line $\hat{Y} = \beta_0 + \beta_1 x_1$

Discount $-u_{0j}$

Differential shift for each district j: index the intercept

Micro-model

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{ij}$$

Macro-model: index parameter as a response

$$\beta_{0j} = \beta_0 + u_{0j}$$

Price of average district j = citywide price + differential for district j

Substitute macro into micro…….

Random intercepts COMBINED model

Substituting the macro model into the micro model yields

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + e_{ij}$$

Grouping the random parameters in brackets

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + (u_{0j} + e_{ij})$$

- **Fixed part** $\beta_0 + \beta_1$
- **Random part (Level 2)** $u_{0j} \sim N(0, \sigma^2_{u0})$
- **Random part (Level 1)** $e_{0ij} \sim N(0, \sigma^2_{e0})$
- **District and house differentials are independent** $Cov[u_{0j}, e_{0ij}] = 0$

The regression coefficient β1 is common to all groups!

What is u_j?

- A residual term
- Specific to unit *j*
- Common to all observations in unit *j*
- Subscript *j*, no subscript *i*
- Interpretation: the difference between the overall intercept and the intercept in unit *j*

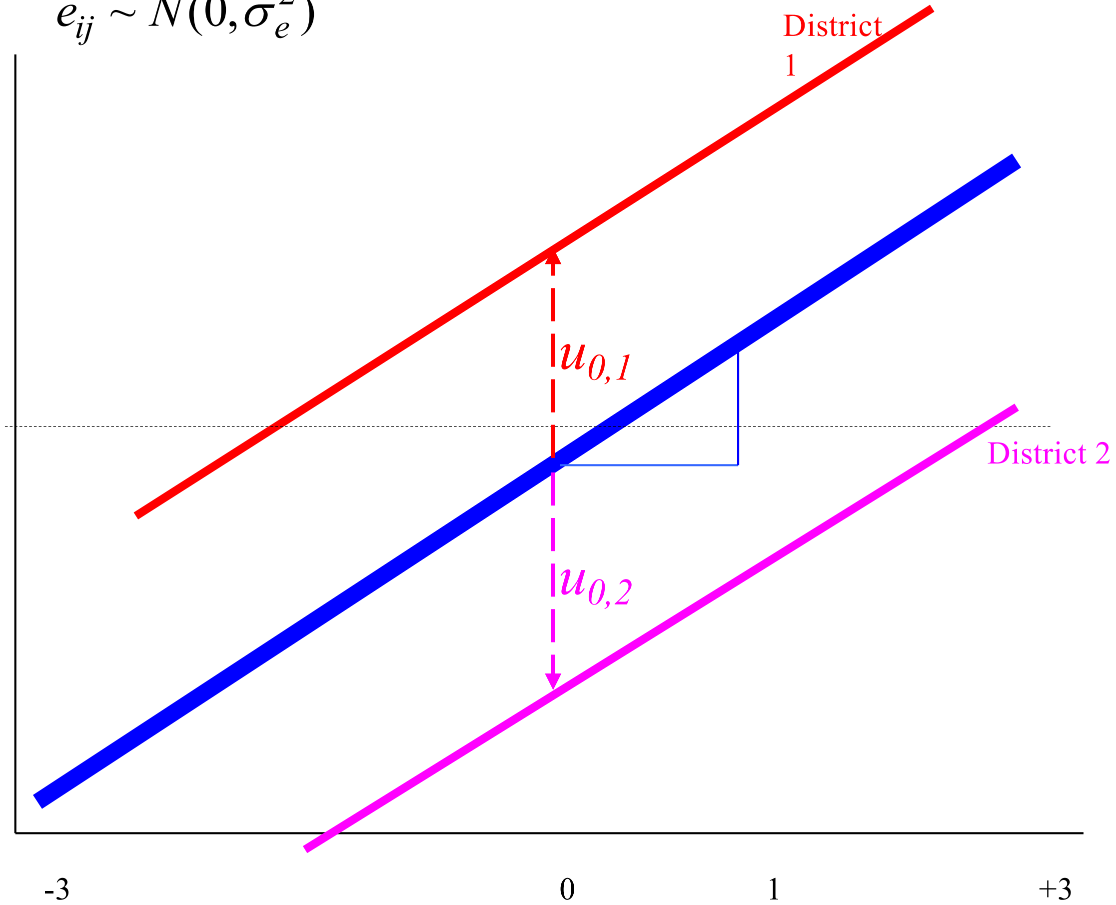
Random intercept models(parallel lines)

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$



What is e_{ij}?

- A residual term
- Specific to observation *i* in unit *j*

The meaning of the random terms

- **Level 2 : between districts**

$$u_{0j} \sim N(0, \sigma^2_{u0})$$

$$\sigma^2_{u0}$$

- **Between district variance conditional on size**

- **Level 1 : within districts between houses**

$$e_{0ij} \sim N(0, \sigma^2_{e0})$$

$$\sigma^2_{e0}$$

- **Within district, between-house variation variance conditional on size**

Variance decomposition

- $\text{Var}(Y_{ij}) = \text{Var}(u_j) + \text{Var}(e_{ij})$

  $= \sigma_u^2 + \sigma_e^2$

- Intraclass Correlation (ρ): the proportion of the total variance in Y_{ij} that is between level-2 units

$$\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$$

Random intercepts and slopes model

Micro-model

$$y_{ij} = \beta_{0j}x_{0ij} + \beta_{1j}x_{1ij} + e_{0ij}x_{0ij}$$

Note: Index the intercept and the slope associated with a constant, and number of rooms, respectively

Macro-model (Random Intercepts)

$$\beta_{0j} = \beta_0 + u_{0j}$$

Macro-model (Random Slopes)

$$\beta_{1j} = \beta_1 + u_{1j}$$

Slope for district j = citywide slope + differential slope for district j

Substitute macro models into micro model…………

Random intercepts and slopes model

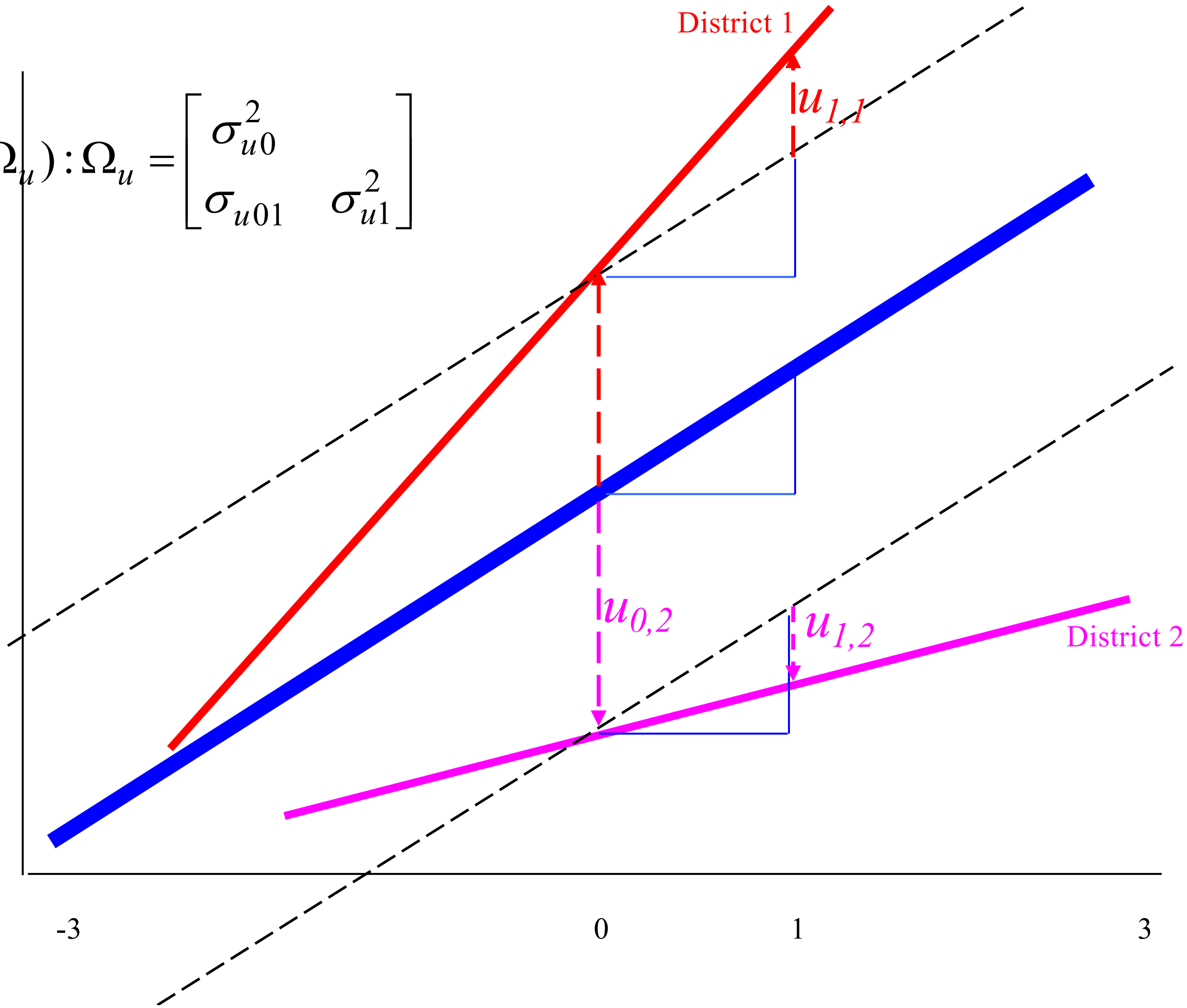
$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0,\Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix}$$

$$e_{ij} \sim N(0,\sigma_e^2)$$



Estimation Methods

- Maximum Likelihood (ML),
- Restricted Maximum Likelihood (RML)
- Iterative Generalized Least Squares (IGLS) algorithms
- Empirical Bayes (EB) estimate

```
Performing gradient-based optimization:

Iteration 0:   log likelihood = -76445.159
Iteration 1:   log likelihood = -76445.159  (backed up)

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =     21,570
Group variable: cosp                            Number of groups  =         35

                                                Obs per group:
                                                              min =        109
  mean of school level intercepts
                                                    avg =      616.3
                          Variance at the school      max =      1,808
                          level (level 2)
                                                Wald chi2(0)      =          .
Log likelihood = -76445.159                     Prob > chi2       =          .

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   9.030971   .5703157    15.84   0.000     7.913173    10.14877
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                  var(_cons) |   11.21163    2.71967      6.969283     18.0364
-----------------------------+------------------------------------------------
               var(Residual) |   69.62139   .6709417      68.31871    70.94891
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 3363.17       Prob >= chibar2 = 0.0000
```

Ho: Random-effects = 0

Variance Decomposition

Variance decomposition

- $\mathrm{Var}(Y_{ij}) = \mathrm{Var}(u_j) + \mathrm{Var}(r_{ij})$

  $= \tau_{00} + \sigma^2$

- <u>Intraclass Correlation</u> (ρ): the proportion of the total variance in Y_{ij} that is between level-2 units

$$\rho = \tau_{00} / (\tau_{00} + \sigma^2)$$

```
. estat icc

Intraclass correlation

------------------------------------------------------------------------------
                       Level |        ICC   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                      school |   .1387012   .0290024       .090962    .2058223
------------------------------------------------------------------------------
```

Intraclass correlation:13.8%

If the interclass correlation (IC) approaches 0 then the grouping by school are of no use (you may as well run a simple regression).

Resources

- Books
  - *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed.* Raudenbush & Bryk, 2002.
  - *Introducing Multilevel Modeling*. Kreft & DeLeeum, 1998.
- Journals
  - Educational and Psychological Measurement
  - Journal of Educational and Behavioral Sciences
  - Multilevel Modeling Newsletter