Multilevel analysis

Introduction to multilevel analysis and the basic two-level regression model

Emmeke Aarts (E.Aarts@uu.nl)
Methodology and Statistics
Faculty of Social and Behavioral Sciences

1

Who

Emmeke Aarts

Course coordinator

Lecturer

Beth Grandfield

Computer labs

Exam

Feedback and grading

2

Course outline - schedule

- Week 1: Mon **Lecture 1**
  Fri **DAT + Start up lab**
- Week 2: Mon **Long computer lab + Q&A**
- Week 3: Mon **Lecture 2** (hand in assignment 1 before start of Lecture 2)
  Fri **DAT + Start up lab**
- Week 4: Mon **Long computer lab + Q&A**
- Week 5: Mon **Lecture 3** (hand in assignment 2 before start of Lecture 3)
  Fri **DAT + Start up lab**
- Week 6: Mon **Long computer lab + Q&A**
- Week 8: **Exam**

3

Course outline - topics

Week 1 & 2
- When/why multilevel analysis
- The multilevel regression model
- The three-level MLM
- MLM assumptions

Green: Lecture

Blue: Discussion additional topics (DAT)

Week 3 & 4
- Longitudinal model
- Contextual effects

Week 5 & 6
- Analyzing dichotomous and ordinal data
- Summary part 1

4

Acknowledgements

Multilevel module content is based on course materials of past and present Utrecht colleagues, including: Cora Maas, Joop Hox, Leoniek Wijngaards-de Meij, Peter van der Heijden and Mirjam Moerbeek.

Permission to use and/or modify their course material is gratefully acknowledged.

Today

1) Introduction: why use multilevel analysis?
2) Building the multilevel regression model
3) Analysis approach
4) Example

Introduction

Multilevel data structures and their implications in analysis

Multilevel regression model

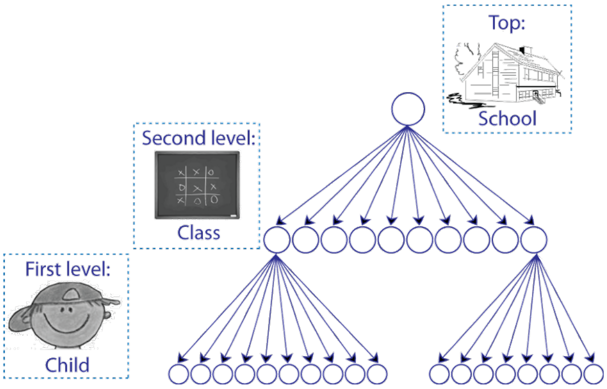
Known in literature under a variety of names

- Hierarchical linear model (HLM)
  - Takes hierarchical data structure into account
- Multilevel model
  - Takes multiple levels of nesting into account
- Random coefficient model
  - Allows effects of predictors to vary across clusters
- Variance component model
  - Partitions the variance into components at individual and cluster level
- Mixed Linear Model
  - Model contains fixed and random effects

Nested data

Example: education / family

level 3 (schools)
level 2 classes / families
level 1 pupils / members

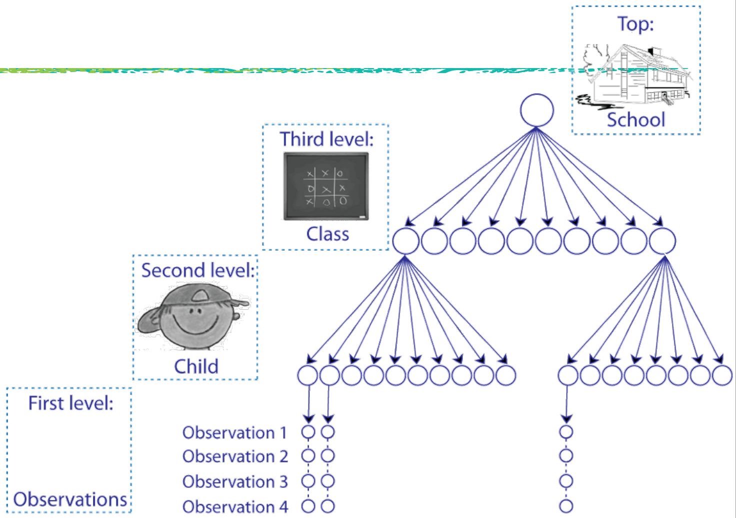


9

Nested data

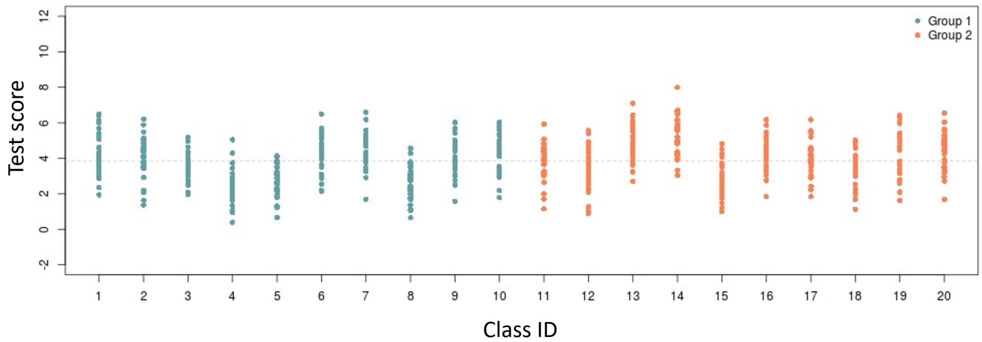
Example: longitudinal

level 3 classes
level 2 pupils
level 1 occasions



10

Nested data - example



When looking at this graph, do you think the collected observations can be seen as independent?

11

Nested data

- Observations in the same cluster are generally not independent
  - they tend to be more similar than observations from different clusters
  - selection, shared history, mutual influence, contextual group effects
- The degree of similarity is indicated by the *intraclass correlation* ρ
- Standard statistical tests are not at all robust against violation of the independence assumption

That is why we need special multilevel techniques!

12

Nested data – more challenges

Not only use proper sample size (i.e., corrected for dependency), but:
- Predict test scores using variables at all levels (student and class level)
- Relation between test score and sex can differ over the clusters (classes)

13

Traditional Approaches

- Disaggregate all variables to the lowest level
  - Do standard analyses (ANOVA, multiple regression)
- Aggregate all variables to the highest level
  - Do standard analyses (ANOVA, multiple regression)
- ANCOVA with clusters as factor (i.e. use dummy variables)

- Some improvements:
  - Explanatory variables as deviations from their cluster mean
  - Have both deviation score and disaggregated cluster mean as predictor
    (separates individual and group effects)

14

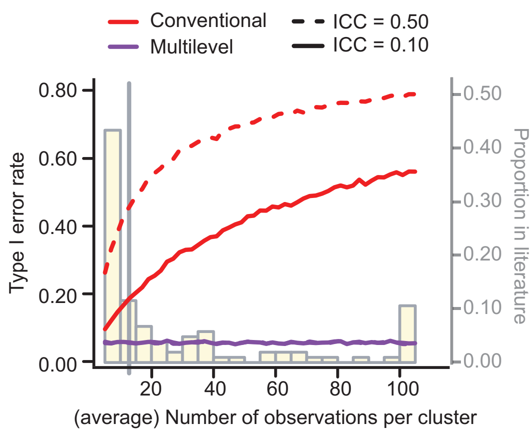
Traditional Approaches

- Disaggregate all variables to the lowest level
  - Do standard analyses (ANOVA, multiple regression)
- Aggregate all variables to the highest level
  - Do standard analyses (ANOVA, multiple regression)
- ANCOVA with clusters as factor (i.e. use dummy variables)

- Some improvements:
  - Explanatory variables as deviations from their cluster mean
  - Have both deviation score and disaggregated cluster mean as predictor
    (separates individual and group effects)

15

Not accommodating dependency inflates Type I error rate

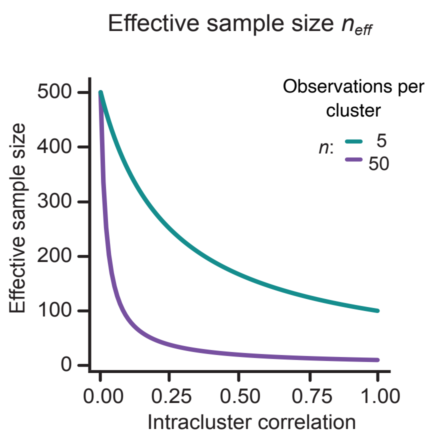


- Outcomes within the same cluster are correlated
- Standard analysis does not take correlation into account
- As a result standard error of effect underestimated
- Results in inflated type I error rate

E. Aarts et al. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature neuroscience*, 17(4), 491-496.

16

Dependency decreases the effective sample size



Given ρ (ICC) and cluster size n_j:

- Effective sample size

$$n_{eff} = n_{total}/[1 + (n_j - 1)\rho\,]$$

- Design effect

$$\text{DE} = 1 + (n_j - 1)\rho$$

E. Aarts et al. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature neuroscience*, 17(4), 491-496.

17

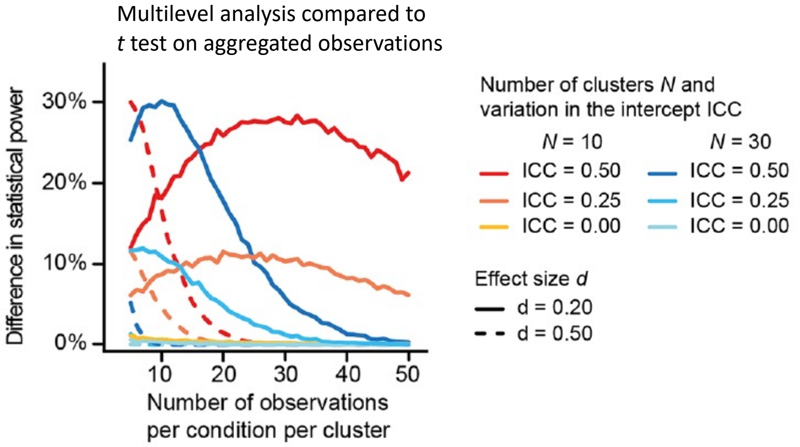
Traditional Approaches

- Disaggregate all variables to the lowest level
  - Do standard analyses (ANOVA, multiple regression)
- Aggregate all variables to the highest level
  - Do standard analyses (ANOVA, multiple regression)
- ANCOVA with clusters as factor (i.e. use dummy variables)

- Some improvements:
  - Explanatory variables as deviations from their cluster mean
  - Have both deviation score and disaggregated cluster mean as predictor
    (separates individual and group effects)

18

Aggregating decreases statistical power



E. Aarts et al. (2015). Multilevel analysis quantifies variation in the experimental effect while optimizing power and preventing false positives. BMC neuroscience, 16(1), 94

Traditional Approaches

- Disaggregate all variables to the lowest level
  - Do standard analyses (ANOVA, multiple regression)
- Aggregate all variables to the highest level
  - Do standard analyses (ANOVA, multiple regression)
- ANCOVA with clusters as factor (i.e. use dummy variables)

- Some improvements:
  - Explanatory variables as deviations from their cluster mean
  - Have both deviation score and disaggregated cluster mean as predictor
    (separates individual and group effects)

AN(C)OVA with cluster as factor

So disaggregating and aggregating are not a good idea.
What if we use the cluster as a predictor in our model?

$$y_{ij} = \beta_0 + u_j + e_{ij}$$

Fixed effects model, with j dummy variables, one for each cluster: one way ANOVA

$$y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + e_{ij}$$

Or one-way ANCOVA (also called variance component model)

AN(C)OVA with cluster as factor

Advantages of ANCOVA approach:

- Simple to compute
- No assumption about distribution of u_j's
- Can cope with large and extreme between cluster differences

Disadvantages of ANCOVA approach

- Large number of parameters when J is large
- Each u_j is poorly estimated if n_j, the number of level one units within each unit j, is small
- Does not make sense if J is a sample from a population and we want to make inferences about that population.
- Suboptimal use of data: some research questions cannot be answered

ANCOVA vs. multilevel analysis

Advantages of ML over ANCOVA approach:

- Efficient estimation (only a few parameters)
- No problem if n_j's are small
- Handles explanatory variables at different levels (cannot be done in ANCOVA approach)
- Quantify variation between clusters

Disadvantages of ML over ANCOVA approach

- Assumptions about distributions of errors

More problems with traditional approaches

- Multiple Regression assumes
  - independent observations
  - independent error terms
  - equal variances of errors for all observations (assumption of homoscedastic errors)
  - normal distribution for errors

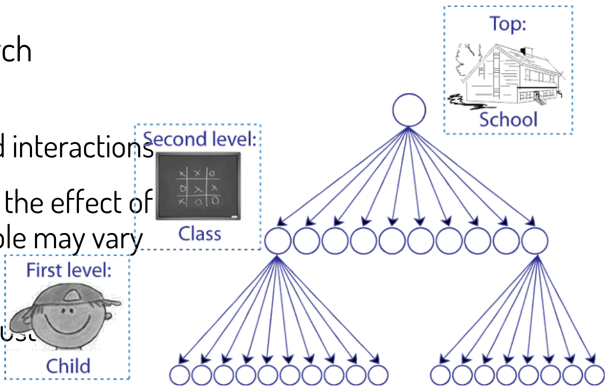
- With hierarchical data
  - observations are not independent
  - errors are not independent
  - different observations may have errors with different variances (heteroscedastic errors)

Why multilevel analysis?

1. Very flexible:
   - *Nested data*
   - *Longitudinal data*
   - *Logistic regression*, Poisson regression
   - Meta-analysis
   - Multivariate
   - *Contextual effects*
   - and so on

Why multilevel analysis?

2. To control for dependency
3. Enables to answer new research questions:
   - Variables at different levels and interactions
   - Variation between clusters: i.e., the effect of an individual explanatory variable may vary over clusters
   - Explaining variation between clusters
   - Contextual effects



Coffee break

Building the multilevel model

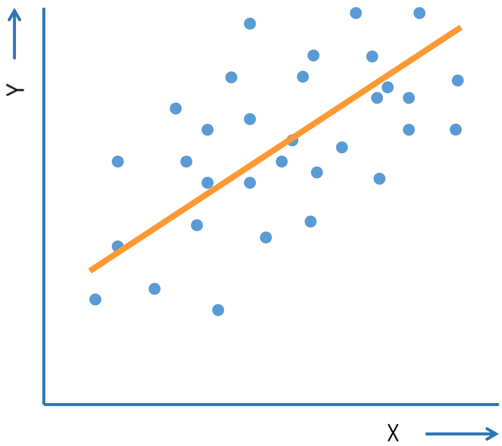
Formulating models at each level of the multilevel data structure

Rehearsal: traditional regression

Ordinary regression with one explanatory variable X:

$y_i = \beta_0 + \beta_1 X_i + e_i$

- y_i outcome of person *i*
- X_i explanatory variable
- β_0 intercept
- β_1 regression slope
- e_i residual error term



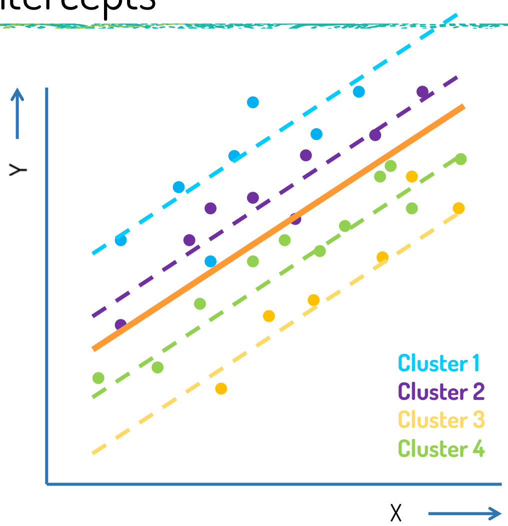
29

Multilevel regression – random intercepts

Multilevel regression with one explanatory variable X:

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

- subscript *j* denotes cluster membership
- y_{ij} outcome of person *i* in cluster *j*
- X_{ij} explanatory variable
- β_{0j} cluster dependent intercept
- β_1 regression slope
- e_{ij} residual error term, are assumed to have mean zero and variance σ_e^2



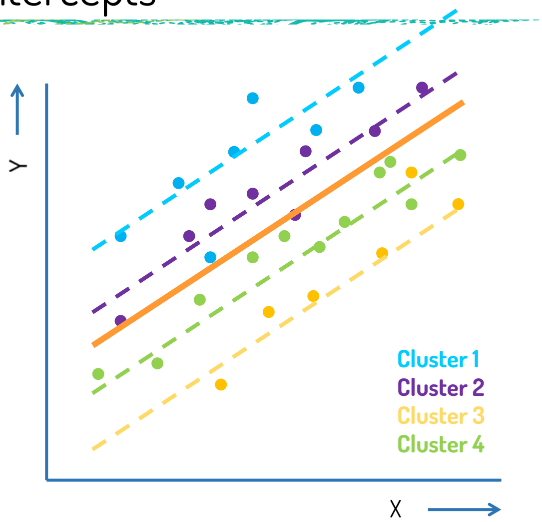
30

Multilevel regression – random intercepts

Multilevel regression with one explanatory variable X:

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

- subscript j denotes cluster membership
- y_{ij} outcome of person i in cluster j
- X_{ij} explanatory variable
- β_{0j} cluster dependent intercept
- β_1 regression slope
- e_{ij} residual error term, are assumed to have mean zero and variance σ_e^2



Emmeke Aarts Multilevel analysis - lecture 1 31

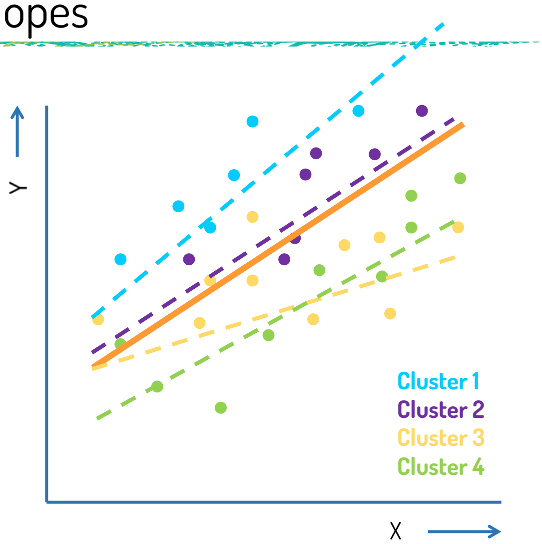
31

Multilevel regression – random slopes

Multilevel regression with one explanatory variable X:

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

- subscript j denotes cluster membership
- y_{ij} outcome of person i in cluster j
- X_{ij} explanatory variable
- β_{0j} cluster dependent intercept
- β_{1j} cluster dependent regression slope
- e_{ij} residual error term, are assumed to have mean zero and variance σ_e^2



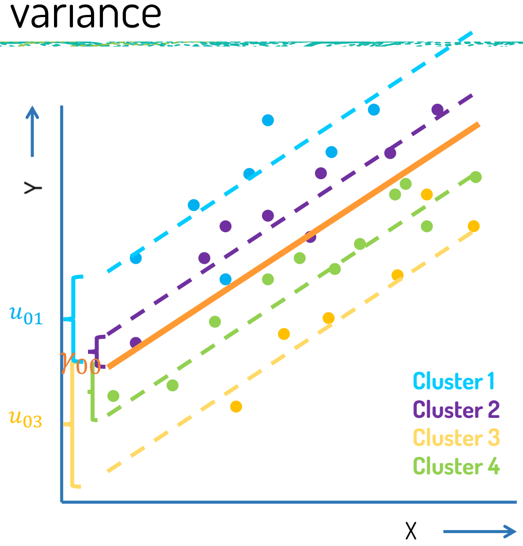
Emmeke Aarts Multilevel analysis - lecture 1 32

32

Multilevel regression – intercept variance

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$



Emmeke Aarts Multilevel analysis - lecture 1 33

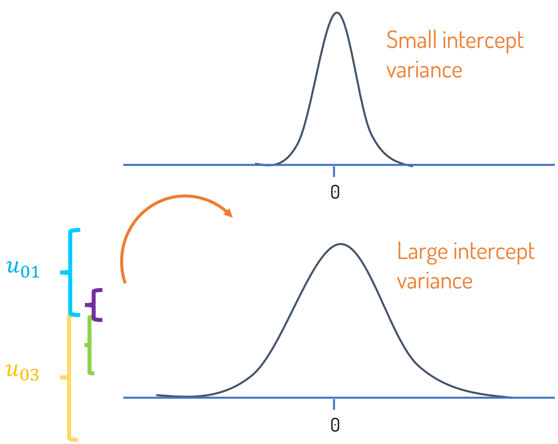
33

Multilevel regression – intercept variance

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$u_{0j} \sim N(0, \sigma^2_{u_0})$$

$\sigma^2_{u_0}$ = intercept variance



Emmeke Aarts Multilevel analysis - lecture 1 34

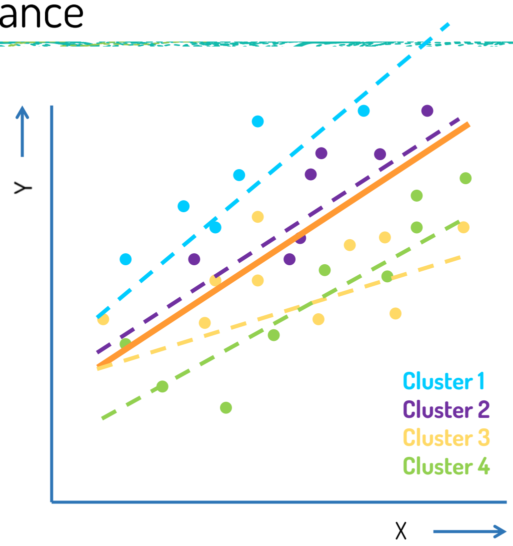
34

Multilevel regression – slope variance

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$u_{1j} \sim N(0, \sigma^2_{u_1})$$

$\sigma^2_{u_1}$ = slope variance



35

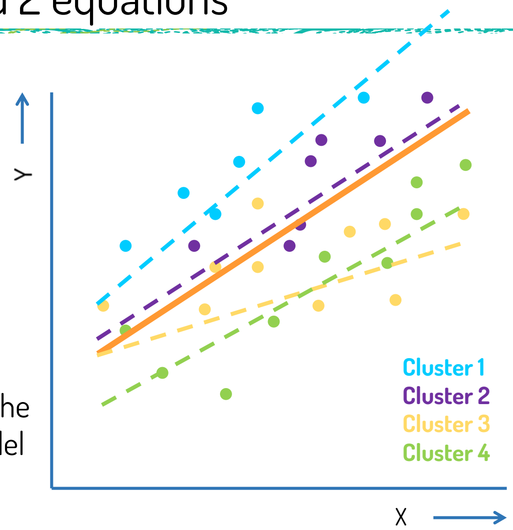
Multilevel regression – level 1 and 2 equations

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$ Level 1 equation (individual level)

$$\left.\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j}\\ \beta_{1j} &= \gamma_{10} + u_{1j}\end{aligned}\right\}$$ Level 2 equations (cluster level)

The intercept and slope coefficients vary across the clusters, hence the term *random coefficient* model

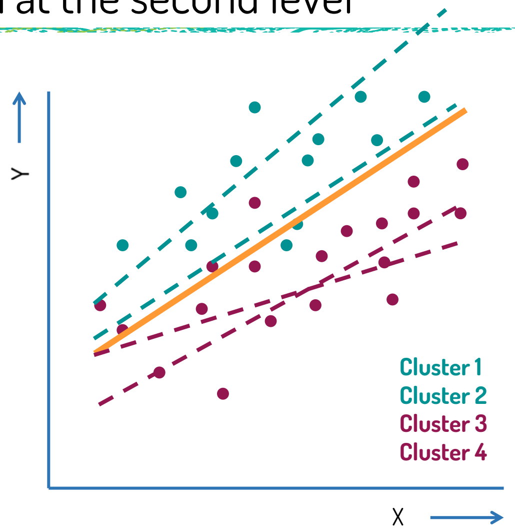


36

Multilevel regression – prediction at the second level

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$



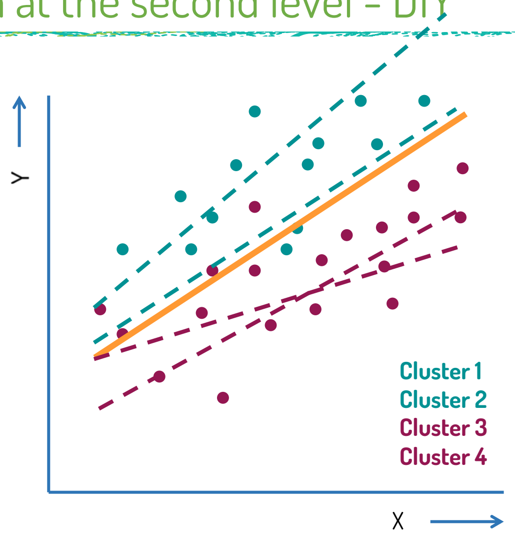
37

Multilevel regression – prediction at the second level – DIY

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

- What (model parameter) decreases if we add $\gamma_{01}Z_j$ to the model?
- What (model parameter) decreases if we add $\gamma_{11}Z_j$ to the model?



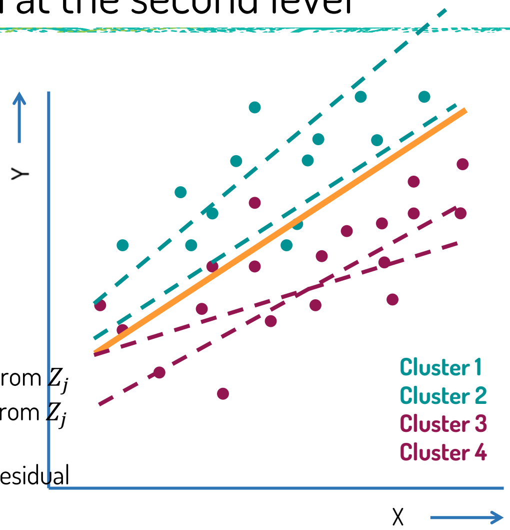
38

Multilevel regression – prediction at the second level

Multilevel regression with one explanatory variable X:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

- γ_{00} and γ_{01} are the intercept and slope to predict β_{0j} from Z_j
- γ_{10} and γ_{11} are the intercept and slope to predict β_{1j} from Z_j
- The regression coefficients γ are **fixed** across clusters
- Between cluster variation left in β's is captured by the residual error terms u_{0j} and u_{1j}



39

Multilevel regression – single equation model

At the lowest (individual) level we have

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

and at the second (cluster) level.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

Combining (substitution and rearranging terms) gives ("mixed" equation)

$$y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

40

Multilevel regression – single equation model

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

This equation has two distinct parts

- $[\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}]$ contains all the fixed coefficients, it is called the **fixed part** of the model
- $[u_{0j} + u_{1j}X_{ij} + e_{ij}]$ contains all the random error terms, it is called the **random part** of the model
- the *cross-level* interaction Z_jX_{ij} results from modeling the regression slope β_{1j} of individual level variable X_{ij} with the group level variable Z_j
- the error term u_{1j} is connected to X_{ij}. Thus the residuals are larger for large values of X_{ij}, implying *heteroscedasticity*

Multilevel regression – interpretation

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

- the fixed part is an ordinary regression model
- complicated error term: $[u_{0j} + u_{1j}X_{ij} + e_{ij}]$

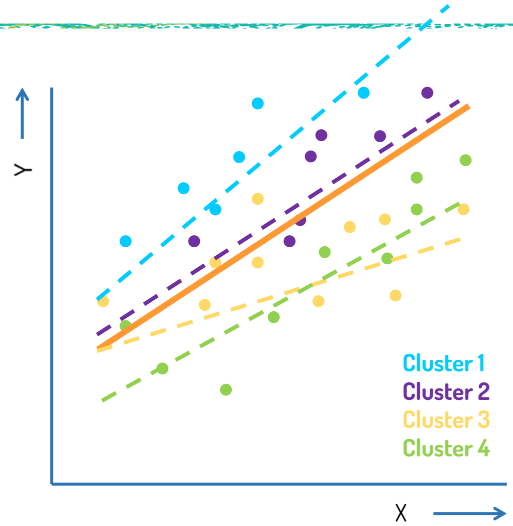
- Several error (co-)variances
  - σ_e^2 variance of the lowest level errors e_{ij}
  - σ_{u0}^2 variance of the highest level errors u_{0j}
  - σ_{u1}^2 variance of the highest level errors u_{1j}
  - σ_{u01} covariance of u_{0j} and u_{1j}

Multilevel regression - DIY

Draw three graphs that depict different scenarios of σ_{u01}, the covariance of u_{0j} and u_{1j}:

1. Positive covariance σ_{u01}
2. Zero covariance σ_{u01}
3. Negative covariance σ_{u01}

If you have time left, think of a real life scenario that would result in each of these three



43

Estimation

- Maximum Likelihood (ML) estimation
- Full maximum likelihood vs Restricted Maximum likelihood (default in R!)
- Measure of model fit:
  - Model *deviance* (-2*loglikelihood)
  - AIC (deviance + 2*k*)
  - Not BIC, penalty based on sample size is ambiguous
- Estimation is iterative, check for convergence

44

Analysis approach

Bottom up: from simple to complicated models

Steps in Model Exploration

1. *Intercept-only* model
   calculate *intraclass correlation coefficient*
2. *Fixed model*, 1st level predictor variables
   test slopes for significance
   calculate proportion explained variance for intercept and residual
3. *Fixed model*, 1st and 2nd level predictor variables
   test slopes for significance
   calculate proportion explained variance for intercept
4. *Random coefficient model*
   test if any 1st level slope has a significant variance component
5. *Full Multilevel Regression Model*
   add predictors for random slopes, test for significance,
   calculate proportion explained variance for slopes

The intercept only model

- Intercept only model (null model, baseline model)
- Contains only intercept and corresponding error terms
  - At the lowest (individual) level we have
    $y_{ij} = \beta_{0j} + e_{ij}$
  - and at the second level
    $\beta_{0j} = \gamma_{00} + u_{0j}$
  - hence
    $y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$

The intercept only model – Intraclass correlation

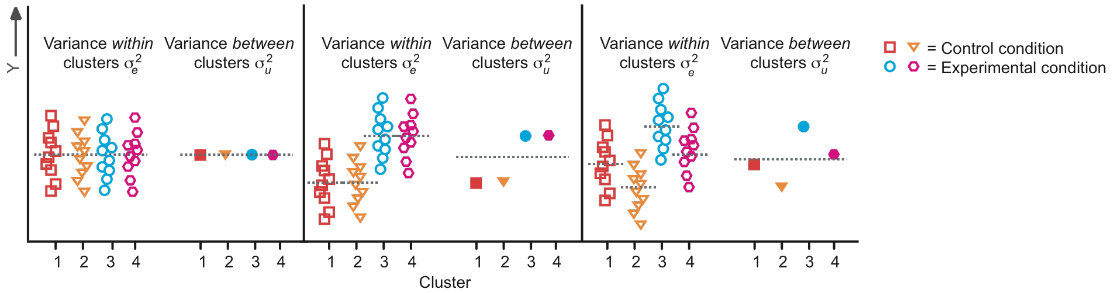
- Intercept only model (null model, baseline model)
    $y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$

- Used to decompose the total variance and compute the *intraclass correlation* ρ
  - $\rho = \sigma_{u0}^2/(\sigma_e^2 + \sigma_{u0}^2)$
  - ρ = cluster level variance / total variance
  - Interpretation:
  1. Expected correlation between two randomly sampled individuals in *same* clsuter
  2. Percentage of variance at the cluster level

The intercept only model - DIY

The graph below displays various scenarios of the intraclass correlation (ICC; ρ) computed as $\rho = \sigma_{u0}^2/(\sigma_e^2 + \sigma_{u0}^2)$

1. What do you think the value of ICC is (approximately) in each of the three graphs?
2. What is the difference between the middle and right panel?



49

The Fixed Model – level 1

Only fixed effects for (individual level) explanatory variables

Slopes are assumed not to vary across groups

- At the lowest (individual) level we have

  $y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$
- and at the second level

  $\beta_{0j} = \gamma_{00} + u_{0j}$ and

  $\beta_{1j} = \gamma_{10}$
- hence

  $y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + e_{ij}$

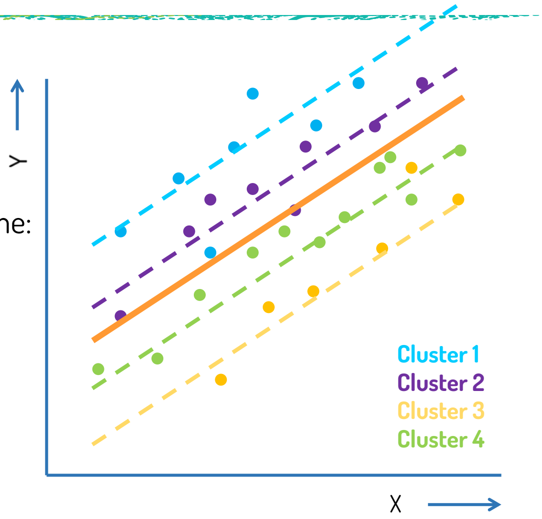
50

The Fixed Model – level 1

Only fixed effects for explanatory variables

Intercepts vary across clusters, slopes are the same:

$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$= (\gamma_{00} + u_{0j}) + \gamma_{10} X_{ij} + e_{ij}$$



The Fixed Model – level 1 and 2

Fixed effects for individual and cluster level explanatory variables

Slopes are assumed not to vary across groups

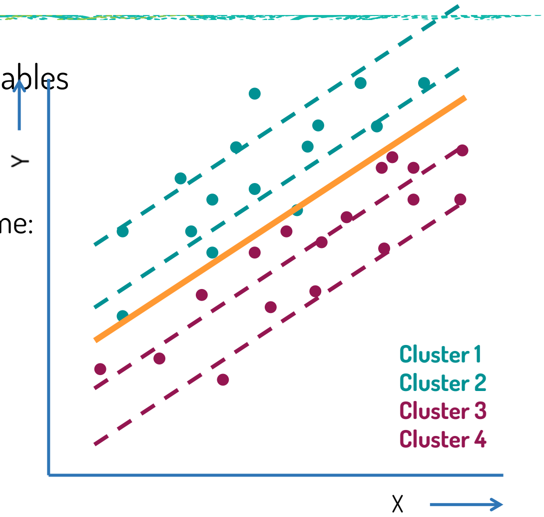
- At the lowest (individual) level we have
  $$y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$$
- and at the second level
  $\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$ and
  $\beta_{1j} = \gamma_{10}$
- hence
  $$y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + u_{0j} + e_{ij}$$
- As we do not include (cross-level) interactions yet, only the intercept is predicted by the cluster level variable

The Fixed Model – level 1 and 2

Fixed effects for level 1 and 2 explanatory variables

Intercepts vary across clusters, slopes are the same:

$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$= (\gamma_{00} + \gamma_{01} Z_j + u_{0j}) + \gamma_{10} X_{ij} + e_{ij}$$



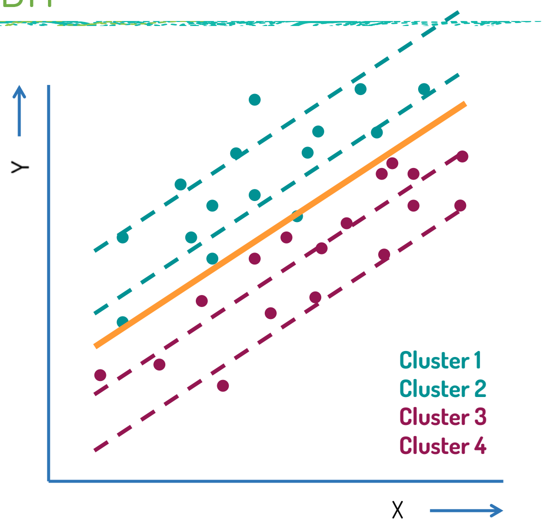
53

The Fixed Model – level 1 and 2 - DIY

In this model, which of the variance components will decrease, compared to the previous model? Why?

Intercepts vary across clusters, slopes are the same:

$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$= (\gamma_{00} + \gamma_{01} Z_j + u_{0j}) + \gamma_{10} X_{ij} + e_{ij}$$



54

The Random Coefficient Model

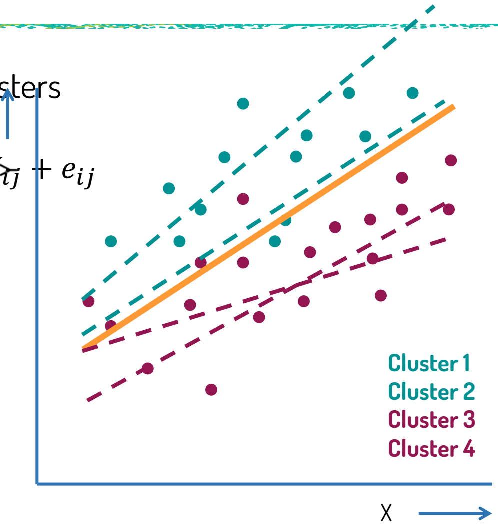
Assumes intercept **and slopes** vary across groups

- At the lowest level
  $y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$
- and at the second level
  $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$ and
  $\beta_{1j} = \gamma_{10} + u_{1j}$
- hence
  $y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + u_{0j} + u_{1j}X_{ij} + e_{ij}$

The Random Coefficient Model

Assumes intercept **and slopes** vary across clusters

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$
$$= (\gamma_{00} + u_{0j} + \gamma_{01}Z_j) + (\gamma_{10} + u_{1j})X_{ij} + e_{ij}$$



Full Multilevel Regression Model

Explanatory variables at all levels

Higher level variables predict variation of lowest level intercept and slopes

- At the lowest (individual) level we have
  $y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$
- and at the second level
  $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$ and
  $\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$
- hence
  $y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij}$

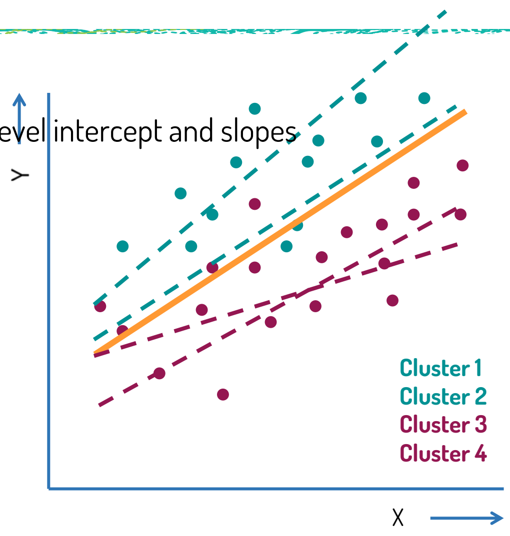
- Predicting the intercept implies a direct effect
- **Predicting slopes implies cross-level interactions**

Full Multilevel Regression Model

Explanatory variables at all levels

Higher level variables predict variation of lowest level intercept and slopes

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$$
$$= (\gamma_{00} + u_{0j} + \gamma_{01}Z_j) + (\gamma_{10} + u_{1j} + \gamma_{11}Z_j)X_{ij} + e_{ij}$$



Coffee break

Explained variance – pseudo R^2

- Intercept only model is the benchmark model for calculating pseudo R^2
- Variance is decomposed in two parts: within / residual variance and between / cluster variance
- We obtain the explained variance for both parts.
- For model with only predictors at level 1 (fixed effects):
  - Explained variance level 1: $R^2_{L1} = \frac{\sigma^2_{e|baseline} - \sigma^2_{e|model1}}{\sigma^2_{e|baseline}}$
  - Explained variance level 2: $R^2_{L2} = \frac{\sigma^2_{u_0|baseline} - \sigma^2_{u_0|model1}}{\sigma^2_{u_0|baseline}}$

Explained variance – pseudo R^2

- Intercept only model is the benchmark model for calculating pseudo R^2
- Variance is decomposed in two parts: within / residual variance and between / cluster variance
- We obtain the explained variance for both parts.
- When adding predictors at level 2 to the model (fixed effects):
  - ~~Explained variance level 1: $R_{L1}^2 = \frac{\sigma_{e|baseline}^2 - \sigma_{e|model2}^2}{\sigma_{e|baseline}^2}$~~
  - Explained variance level 2: $R_{L2}^2 = \frac{\sigma_{u_0|baseline}^2 - \sigma_{u_0|model2}^2}{\sigma_{u_0|baseline}^2}$
  - For subsequent models we thus look at the **added** explained variance

61

Explained variance – pseudo R^2

- Can also compute total explained variance:
  - Suppose 70% of the variance is at the individual level, and 30% at the cluster level
  - We explain in total 45% of the individual variance
  - We explain in total 60% of the cluster level variance
- Total explained variance:
  - (45% of 70%) + (60% of 30%) = 50%

- When adding slope variance, we do not explain any variance
- When adding cross level interactions (i.e., full model), we only look at the explained variance of ... ?

62

Example 1

Popularity in Schools

Predicting school popularity

- Simulated data set
- 100 classes, 2000 pupils
- Continuous outcome: popularity rating
- Explanatory variables pupil level
  - sex (0 = boy, 1 = girl),
  - extraversion (1-10)
- Explanatory variables class level:
  - teacher experience (in years, 2-25)

0. Model ignoring multilevel structure

```
lm(popular ~ 1, data = popular)
```

$$popularity_i = \gamma_{00} + e_i$$

What does γ_{00} represent in this model?

Model:	M$_0$: intercept only – no ML	M$_1$: intercept only
Fixed part	Coefficients (SE)	Coefficients (SE)
Intercept γ_{00}	5.08 (0.03)	
Gender γ_{10}		
Extr γ_{20}		
T exp γ_{01}		
Extr*texp γ_{21}		
Random part		
σ_e^2	1.91	
$\sigma_{u_0}^2$		

1. Intercept-only Model

```
lmer(popular ~ 1 + (1|class), REML = FALSE,
     data = popular)
```

$$popularity_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

Should we perform multilevel analysis?

$$\rho = \frac{\sigma_{u0}^2}{\sigma_e^2 + \sigma_{u0}^2} = \frac{0.69}{1.22 + 0.69} = 0.36$$

About one third of the variance is at the class level (this is unusually high)

Model:	M$_0$: intercept only – no ML	M$_1$: intercept only
Fixed part	Coefficients (SE)	Coefficients (SE)
Intercept γ_{00}	5.08 (0.03)	5.08 (0.09)
Gender γ_{10}		
Extr γ_{20}		
T exp γ_{01}		
Extr*texp γ_{21}		
Random part		
σ_e^2	1.91	1.22
$\sigma_{u_0}^2$		0.69

2. Fixed Model: 1st level predictors

```
lmer(popular ~ gender + extrav + (1|class),
     REML = FALSE, data = popular)
```

$$popularity_{ij} = \gamma_{00} + \gamma_{10}sex_{ij} + \gamma_{20}extr_{ij} + u_{0j} + e_{ij}$$

Model:	M_1: intercept only		M_2: level 1 predictors	
Fixed part	Coefficients (SE)		Coefficients (SE)	
Intercept γ_{00}	5.08	(0.09)	2.14	(0.12)
Gender γ_{10}			1.25	(0.04)
Extr γ_{20}			0.44	(0.02)
T exp γ_{01}				
Extr*texp γ_{21}				
Random part				
σ_e^2	1.22		0.59	
$\sigma_{u_0}^2$	0.69		0.62	
Explained var				
Level 1			0.52	
Level 2			0.11	

3. Fixed Model: 1st and 2nd level predictors

```
lmer(popular ~ gender + extrav + texp +
     (1|class),
     REML = FALSE, data = popular)
```

$$popularity_{ij} = \gamma_{00} + \gamma_{10}sex_{ij} + \gamma_{20}extr_{ij} + \gamma_{01}texp_j + u_{0j} + e_{ij}$$

Model:	M_2: level 1 predictors		M_3: level 1 and 2 predictors	
Fixed part	Coefficients (SE)		Coefficients (SE)	
Intercept γ_{00}	2.14	(0.12)	0.81	(0.17)
Gender γ_{10}	1.25	(0.04)	1.25	(0.04)
Extr γ_{20}	0.44	(0.02)	0.45	(0.02)
T exp γ_{01}			0.09	(0.01)
Extr*texp γ_{21}				
Random part				
σ_e^2	0.59		0.59	
$\sigma_{u_0}^2$	0.62		0.29	
Explained var				
Level 1	0.52		0.52	
Level 2	0.11		0.58	

4. Random Coefficient Model

```
lmer(popular ~ 1 + extrav + gender + texp + (extrav|class),
     REML = FALSE, data = popular)
```

$popularity_{ij} =$
$\gamma_{00} + \gamma_{10}sex_{ij} + \gamma_{20}extr_{ij} + \gamma_{01}texp_j + u_{0j} +$
$u_{1j}sex_{ij} + u_{2j}extr_{ij} + e_{ij}$

What is σ_{u02}?

Significant slope variation for *extraversion*, not for *sex*

Model:	**M$_4$: with random slope**	
Fixed part	Coefficients (SE)	
Intercept γ_{00}	0.76	(0.20)
Gender γ_{10}	1.25	(0.04)
Extr γ_{20}	0.45	(0.03)
T exp γ_{01}	0.09	(0.01)
Extr*texp γ_{21}		
Random part		
σ_e^2	0.55	
$\sigma_{u_0}^2$	1.32	
$\sigma_{u_2}^2$	0.03	
$\sigma_{u_{02}}^2$	-0.19	

5. Full Multilevel Regression Model

```
lmer(popular ~ 1 + extrav + gender + texp +
     extrav*texp + (extrav|class),
     REML = FALSE, data = popular)
```

$popularity_{ij} =$
$\gamma_{00} + \gamma_{10}sex_{ij} + \gamma_{20}extr_{ij} + \gamma_{01}texp_j +$
$\gamma_{21}extr_{ij} * texp_j + u_{0j} + u_{2j}extr_{ij} + e_{ij}$

Smaller slope variation for *extraversion*.

85% explained. No longer significant!

Model:	**M$_4$: with random slope**		**M$_5$: cross-level interaction**	
Fixed part	Coefficients (SE)		Coefficients (SE)	
Intercept γ_{00}	0.74	(0.20)	-1.21	(0.27)
Gender γ_{10}	1.25	(0.03)	1.24	(0.04)
Extr γ_{20}	0.45	(0.02)	0.80	(0.04)
T exp γ_{01}	0.09	(0.01)	0.23	(0.02)
Extr*texp γ_{21}			-0.03	(0.00)
Random part				
σ_e^2	0.55		0.55	
$\sigma_{u_0}^2$	1.30		0.45	
$\sigma_{u_2}^2$	0.03		0.00	
$\sigma_{u_{02}}^2$	-0.19		-0.03	

Summary

- We use multilevel regression analysis to accommodate nested data
- In multilevel analysis, both the intercept and the slope (of level 1 predictors) can be random, that is, vary over clusters
- Using multilevel analysis ensures optimal use of our data: not only accommodate dependency but enables us to answer new research questions
- The amount dependency in the data is quantified by the ICC, obtained from the intercept only model
- The multilevel model is build up from simple to more complex
- The amount of explained variance is obtained both at the individual level (residual variance) and at the cluster level (intercept / slope variance)