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

all same

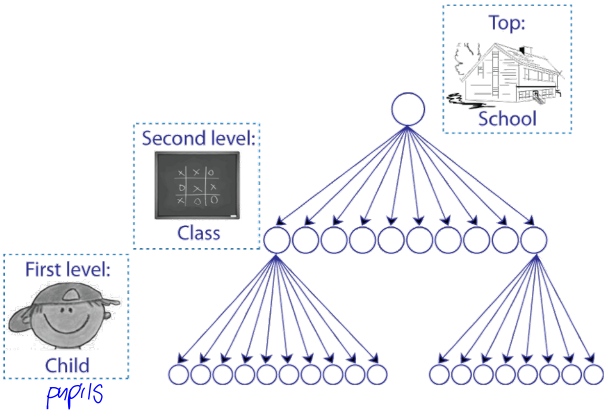
- Hierarchical linear model (HLM)
  - Takes hierarchical data structure into account
- Multilevel model
  - Takes multiple levels of nesting into account
- Random coefficient model
  - Allows effects of predictors to vary across clusters
- Variance component model
  - Partitions the variance into components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

Nested data

~ hierachical structure within the data naturally.
Within the data, there're clusters in which observations tend to be more alike.

Example: education family

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

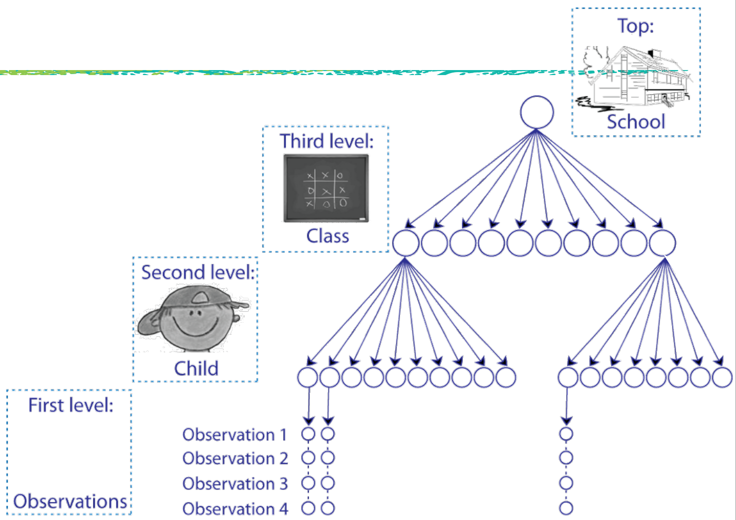


9

Nested data

Example: longitudinal

level 3 classes
level 2 pupils
level 1 occasions



a person
becomes a cluster in which multiple observations
are nested.

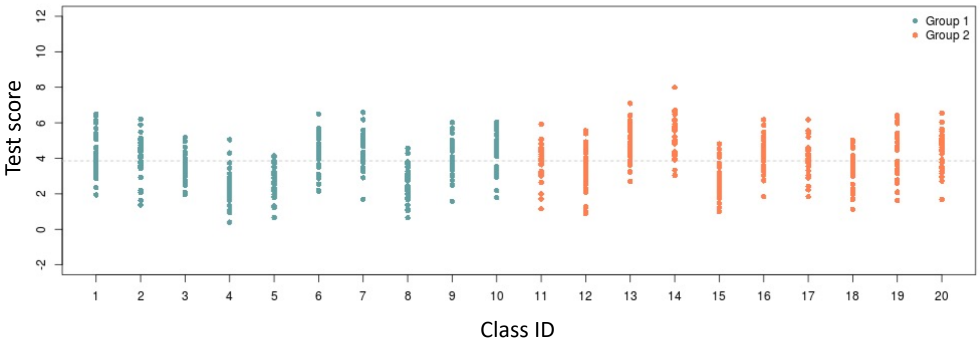
10

If class ID gives no information on Test score, then all the class mean would be the same. Ofc. there could be variation within the class, but the means would be appx the same on the y-axis.

04/02/2022

we can see that the class means do actually differ quite substantially over classes. So if I know which class ppl come from, I can kinda predict their scores. Looking at this plot, I don't think the observations are independent to each other cuz classID gives info. on what the results would be.

Nested data - example



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

No, I can get info by knowing class ID. If they were to be independent, the score levels all should be more or less the same.

Nested data

- Observations in the same cluster are generally not independent
  - they tend to be more similar than observations from different clusters
  - why? → selection, shared history, mutual influence, contextual group effects
- The degree of similarity is indicated by the "*intraclass correlation* ρ" ICC ~ ranges from 0 to 1. 0 being no similarities no dependencies & 1 being compete dependency.
- Standard statistical tests are not at all robust against violation of the independence assumption

That is why we need special multilevel techniques!

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) ~ using variables at their correct level.
- Relation between test score and sex can differ over the clusters (classes) ~ we can try to explain why this is happening. Why the effect of being a girl is larger in one class than the other.

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)

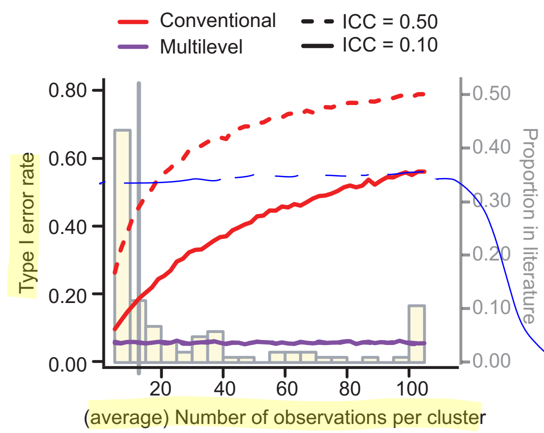
Traditional Approaches

- 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

Ex)
ICC=0.10 & about 100 obs per cluster,
then your Type-I error rate rises to about 50%.
So if you have a sig. result, you actually have no clue
whether you can trust it or not.

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

This shows you if you have dependency
& you want to trust your result, then
you need to accommodate your dependency using ML analysis.

16

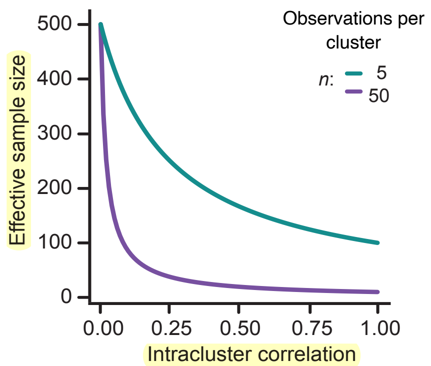
How much info. we bring into the analysis

ICC

Dependency decreases the effective sample size

If there's dependency, becomes smaller than the actual amount of observations you have at the lowest level.

Effective sample size n_{eff}



Given ρ (ICC) and cluster size n_j:

- Effective sample size

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

- Design effect

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

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

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)

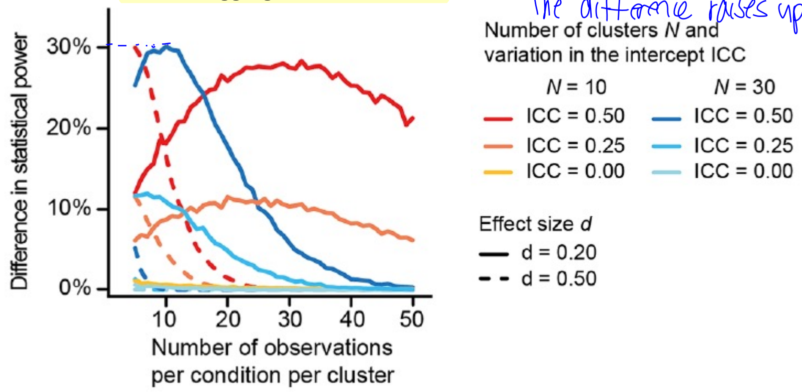
- Some improvements:
  - Explanatory variables as deviations from their cluster mean
  - Have both deviation score and disaggregated cluster mean as predictor
    (separates individual and group effects)

Aggregating decreases statistical power

since you throw away info, you have less info than the amount of obs. you actually collected

Multilevel analysis compared to
t test on aggregated observations

→ The higher the value is, the larger the diff. between these two tests.
The difference raises upto maximum 30%.



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

19

Traditional Approaches

- Disaggregate all variables to the lowest level   not a good idea
  - Do standard analyses (ANOVA, multiple regression)
- Aggregate all variables to the highest level   not a good idea
  - 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)

20

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. we can't generalize
- Suboptimal use of data: some research questions cannot be answered

Imagine that you have 100 clusters → means we have 100 values of u_j that you need to inspect.
& often poorly estimated when sample size within a cluster is rather small.

we cannot generalize to the larger population w.r.t our outcome.
we can only generalize to the particular samples that we collected if we take this approach.
& almost never it's our interest to say sth. about a sample. but a population

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 ~ often it's interesting to see the amount of variation between clusters. & to be able to statistically quantify this so that we can do sth. w/ it.

Disadvantages of ML over ANCOVA approach

- Assumptions about distributions of errors

  that assumption needs to fit the data ofc.

Emmeke Aarts Multilevel analysis - lecture 1 23

23

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)

Emmeke Aarts Multilevel analysis - lecture 1 24

24

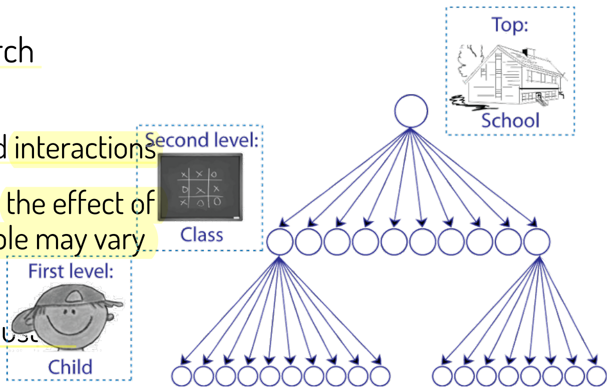
Why multilevel analysis?

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

25

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 clus
   - Contextual effects



26

Coffee break

Emmeke Aarts Multilevel analysis - lecture 1 27

27

Building the multilevel model

Formulating models at each level of the multilevel data structure

Emmeke Aarts Multilevel analysis - lecture 1 28

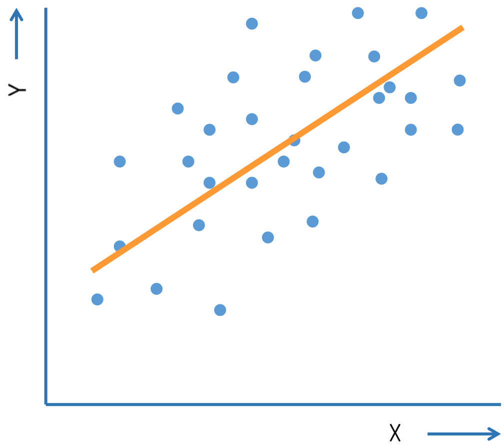
28

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



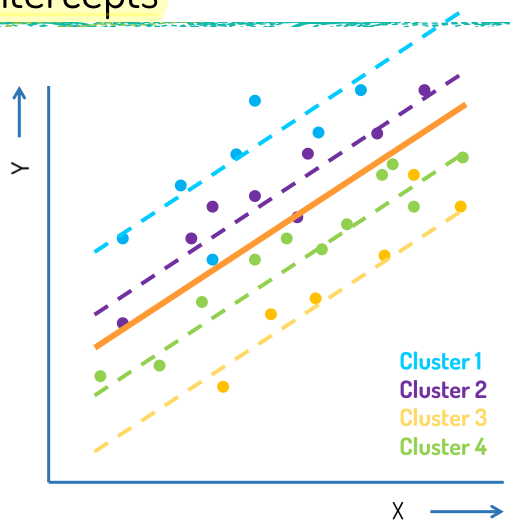
29

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 slope is fixed
- e_{ij} residual error term, are assumed to have mean zero and variance σ_e^2



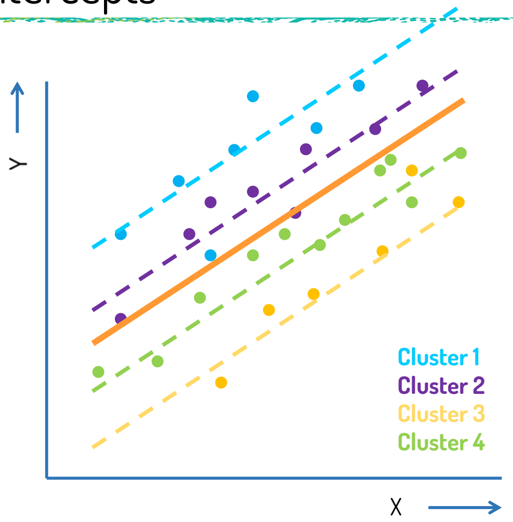
30

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



31

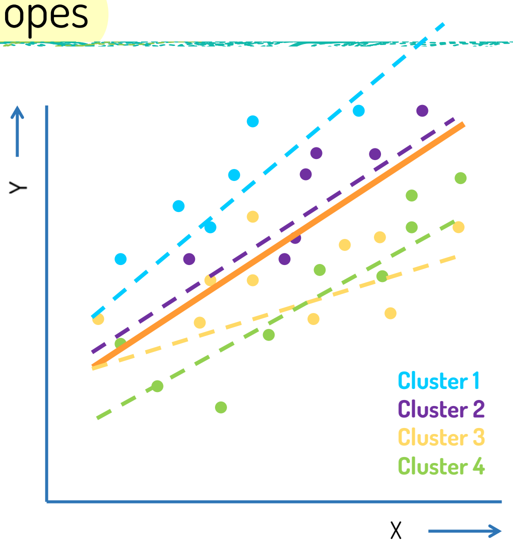
I can also let the slope be cluster-dependent.

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



32

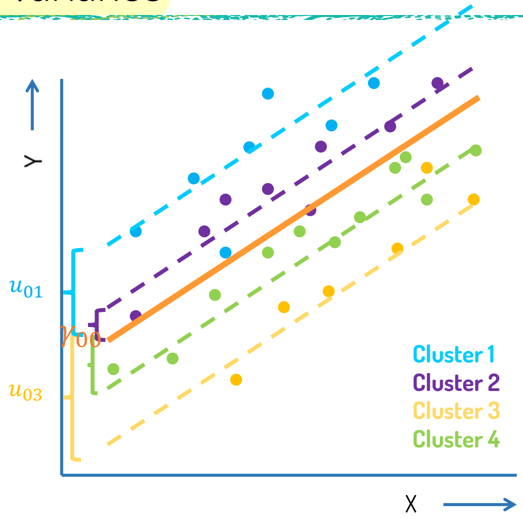
Now we have regressions at two levels:

Multilevel regression – intercept variance

Multilevel regression with one explanatory variable X:

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

cluster level — $\beta_{0j} = \underbrace{\gamma_{00}}_{\text{overall intercept across clusters}} + \underbrace{u_{0j}}_{\text{cluster-specific deviation to the overall intercept.}}$



Emmeke Aarts | Multilevel analysis - lecture 1 | 33

33

Multilevel regression – intercept variance

Multilevel regression with one explanatory variable X:

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

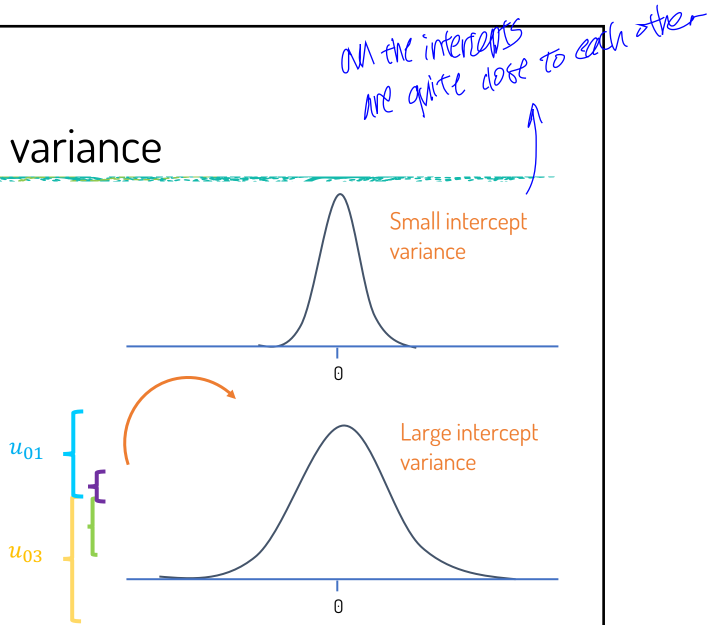
$\beta_{0j} = \gamma_{00} + u_{0j}$ we assume

$u_{0j} \sim N(0, \sigma^2_{u_0})$

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

So instead of estimating $U_{01}, U_{02}, \ldots U_{0j}$ we see the intercept variance

represents the amount of variance / heterogeneity that we observe in this cluster-specific deviations to the overall intercept.



Emmeke Aarts | Multilevel analysis - lecture 1 | 34

34

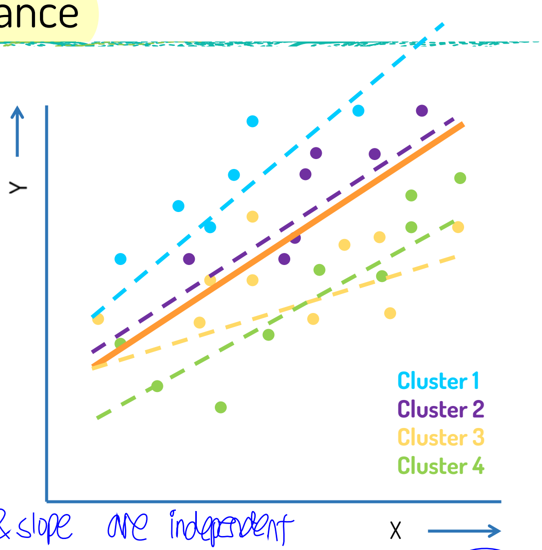
Multilevel regression – slope variance

Multilevel regression with one explanatory variable X:

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

also normal dist. $u_{1j} \sim N(0, \sigma^2_{u_1})$

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



* $\sigma^2_{u_1}$, $\sigma^2_{u_0}$ variances from intercept & slope are independent of the lower level variance $\sigma^2_{e_{ij}}$

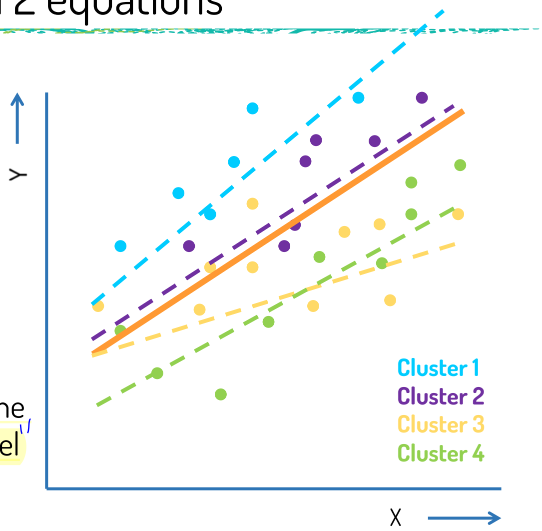
35

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)
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$ Level 2 equations (cluster level)

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



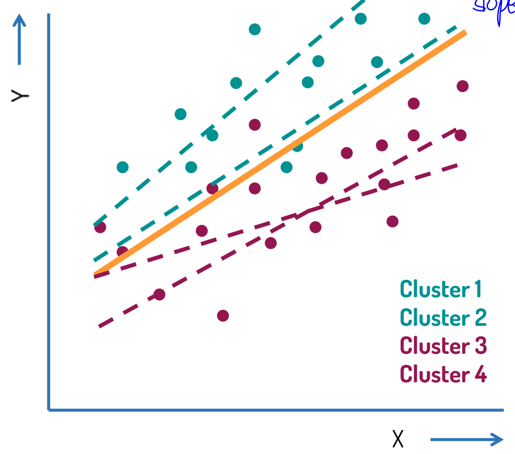
36

Multilevel regression – prediction at the second level

"cluster-level predictor": that partly explains why the intercepts or slopes differ.

Multilevel regression with one explanatory variable X:

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



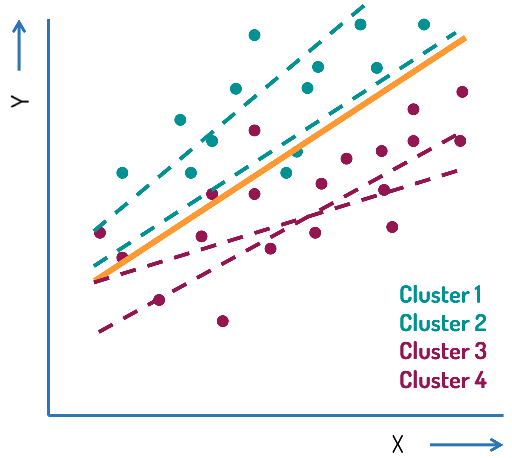
37

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



38

→ level2 predictors potentially reduce the intercept variance.

→ Same principle as intercepts, but keep in mind this will introduce a "cross-level interaction"

allows us to partly explain why the effect varies over the diff. clusters.

Multilevel regression – prediction at the second level

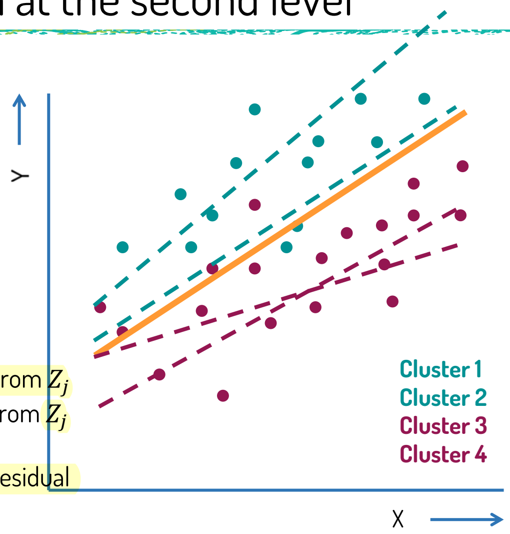
Multilevel regression with one explanatory variable X:

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

fixed effects

- 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} random effects

"all gammas are fixed"



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)

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

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*

$U_{ij} X_{ij}$ → implies that the residuals are larger for large value of X_{ij} & we're hence accommodating any potential heteroscedasticity in our data.

Emmeke Aarts Multilevel analysis - lecture 1

41

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} → independent from all the other errors in the model. It stands on its own.
  - 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}

all the level 2 variances
usually they are allowed to be correlated to each other.
So there's also a covariance between them (U_0 & U_1)

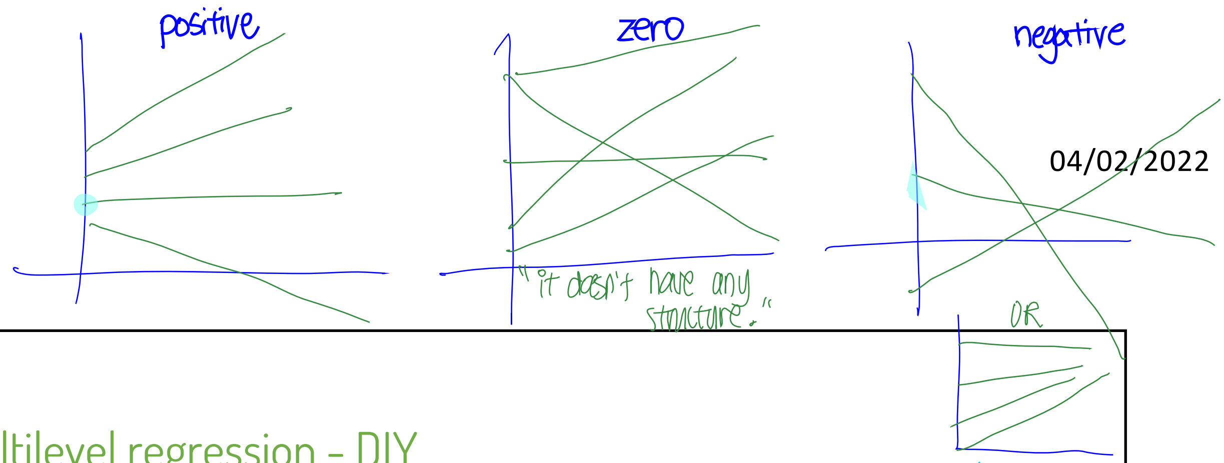
Emmeke Aarts Multilevel analysis - lecture 1 42

42

1. pos covariance → slope is also gonna have a positive deviation to the overall slope.
meaning that if we have a small pos. deviation to the intercept, we'll have a small positive deviation to the slope. If we have a larger pos. deviation to the intercept, we'll also have a larger pos. deviation to the slope. If we have a neg. deviation to the intercept, we'll have a neg. deviation to the slope.

3. neg covariance → if we have a pos deviation to the intercept, we have a neg deviation to slope: when higher intercept, the slope will go lower than the overall slope, and vice versa.

04/02/2022

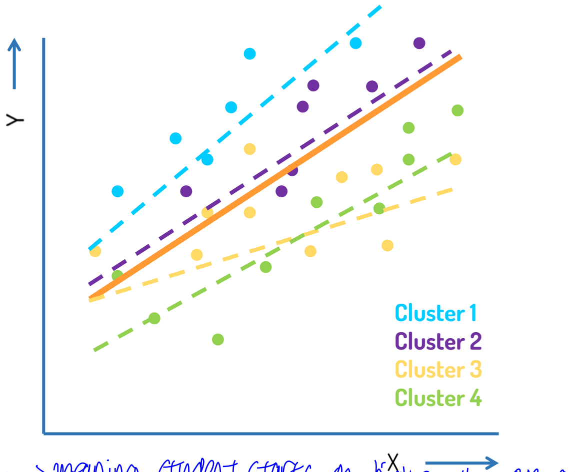


Multilevel regression - DIY

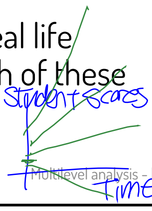
Draw three graphs that depict different scenarios of σ_{u01}, the covariance of u_{0j} and u_{1j}:

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



1. accelerating effect



→ meaning student starts on higher than average, they become even better & better & student starts on low, it becomes ever harder & harder for them and their scores go down more.

2. regression to mean: doesn't matter where you start, over time it will end up around the mean

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

in that sense, is more straightforward

(what sample size to use?)

Diff*
FML: likelihood estimate is based on both fixed component & the random component.
RML: likelihood is only based on variance components, the random components. so it ignores the fixed parts.

→ we always wanna use FML so that we can actually compare the likelihood of models that become more & more complex. So we can test differences between models & check the significant improvements... → FML approach is necessary

Analysis approach

Bottom up: from simple to complicated models

Steps in Model Exploration

1. *Intercept-only* model → we just estimate the "mean".
   calculate *intraclass correlation coefficient* (ICC)
2. *Fixed model*, 1st level predictor variables
   test slopes for significance γ_{10}
   calculate proportion explained variance for intercept and residual → if non-significant predictor, kick 'em out.
3. *Fixed model*, 1st and 2nd level predictor variables
   test slopes for significance γ_{01}
   calculate proportion explained variance for intercept only → "
4. *Random coefficient model*
   test if any 1st level slope has a significant variance component → introduce random slopes for the model. we use this one-by-one, so test them sequentially.
5. *Full Multilevel Regression Model*
   add predictors for random slopes, test for significance,
   calculate proportion explained variance for slopes

IFF we have random slopes, we introduce cross-level interactions to see if we can partly explain why the effects differ across clusters.

If we didn't have a sig. slope variance, this last step doesn't make any sense. Becuz if all effects are similar over the clusters, it doesn't make sense to try to explain why the effects differ. "nothing to explain"

☆ Remember, the kicked-out predictor in step 2 is tested for random slopes. So we re-introduce it to the model. becuz maybe there's nothing going on in average, but still interesting going on, maybe this slope varies hugely across the clusters so much so that the average becomes zero.
So we re-introduce non-sig. level-1 predictors & test for random slopes, if they're non-sig. for this, we kick it out completely again.

The intercept only model

- Intercept only model (null model, baseline model)
- Contains only intercept and corresponding error terms → allows us to disentangle the variance to 2 levels.
  - 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)
  $y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$

nothing more than a normalized version of the intercept variance. σ^2_{u0}

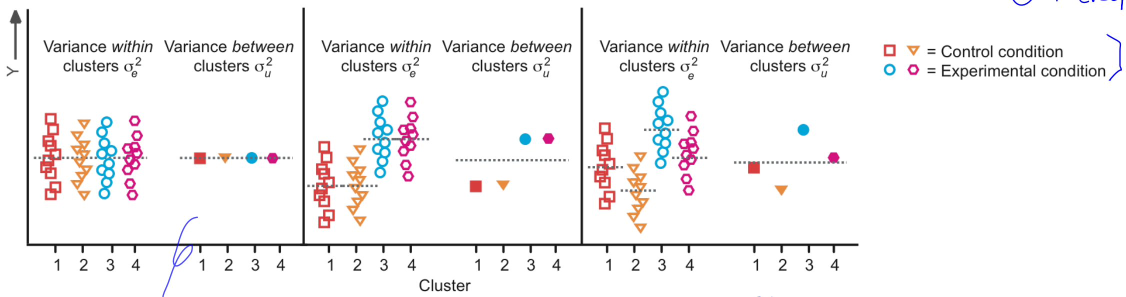
- Used to decompose the total variance and compute the *intraclass correlation* ρ ~ between 0 —— 1. 0 denoting no dependency & 1 denoting complete dependency
  - $\rho = \sigma^2_{u0}/(\sigma^2_e + \sigma^2_{u0})$
  - ρ = cluster level variance / total variance → combination of residual var. at level 1 σ^2_e and the intercept var. at level 2 σ^2_{u0}
  - Interpretation:
  1. Expected correlation between two randomly sampled individuals in *same* clsuter
  2. Percentage of variance at the cluster level : what percentage of variance is explained by the cluster differences

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? 0, 0.5, 0.5
2. What is the difference between the middle and right panel?

@ intercept-only model



no var. between cluster mean = 0
↓
ICC = 0

2. ICC are identical if you see intercept-only model in those two panels. But the diff. in cluster means in middle panel are fully explained by level2 predictor. In right panel, some of the intercept variance will remain after taking Cont/Exp conditions into account. So here the unexplained ICC is 0 (middle) and 0.3 (right panel).

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

  $\beta_1 = \gamma_{10}$
- hence

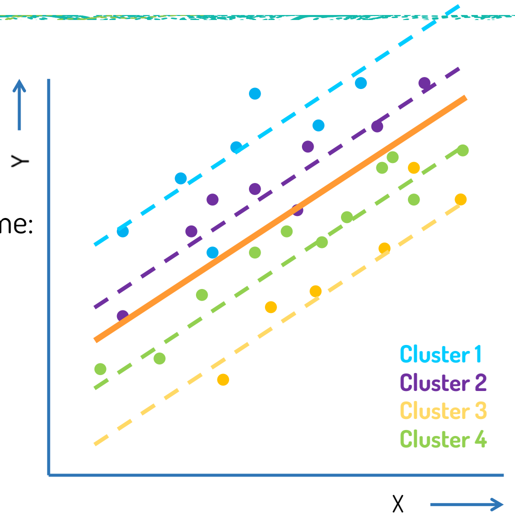
  $y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + e_{ij}$

The Fixed Model – level 1

Only fixed effects for explanatory variables

Intercepts vary across clusters, slopes are the same:

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



51

The Fixed Model – level 1 and 2

Fixed effects for individual and cluster level explanatory variables

Slopes are assumed not to vary across groups

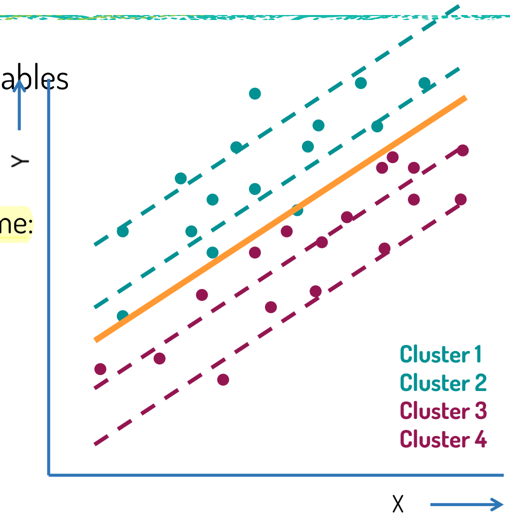
Slope is still fixed but we explain why intercepts vary w/ cluster level predictor.

- 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

52

The Fixed Model – level 1 and 2

Fixed effects for level 1 and 2 explanatory variables



Intercepts vary across clusters, slopes are the same:

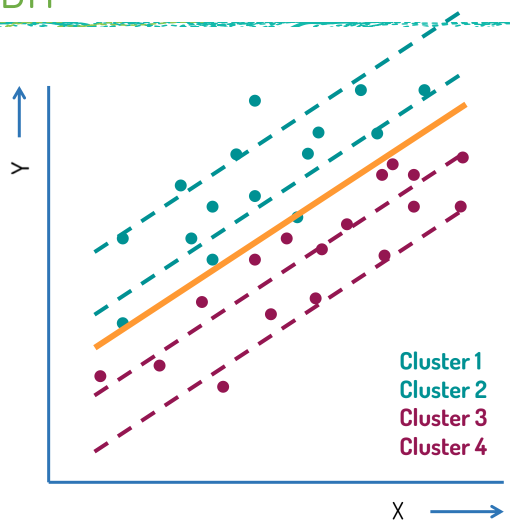
$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$= (\gamma_{00} + \gamma_{01} Z_j + u_{0j}) + \gamma_{10} X_{ij} + e_{ij}$$

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:

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



The Random Coefficient Model

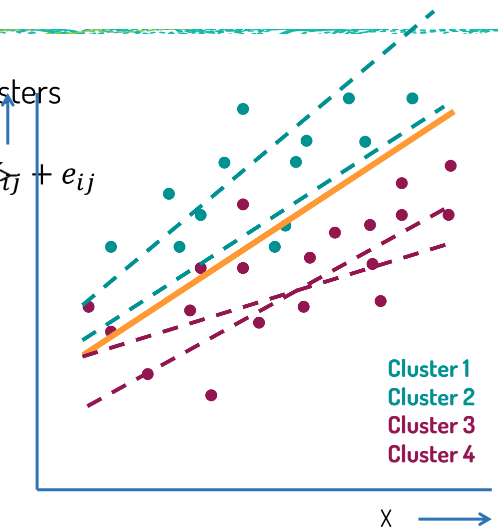
Assumes intercept **and slopes** vary across groups

- 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

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

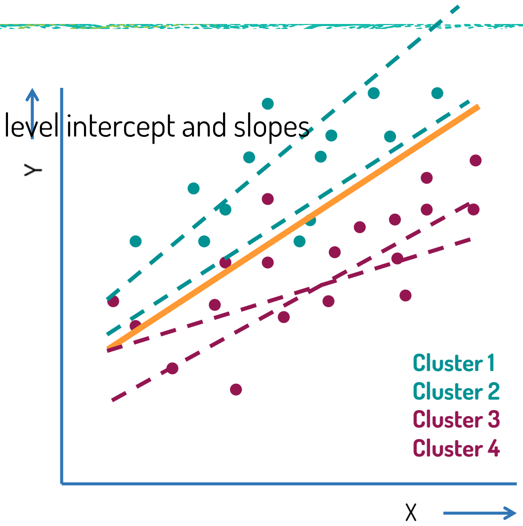
- Predicting the intercept implies a direct effect
- **Predicting slopes implies cross-level interactions**

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 R²

- 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: $R^2_{L1} = \frac{\sigma^2_{e|baseline} - \sigma^2_{e|model1}}{\sigma^2_{e|baseline}}$ ← mod. w/ predictors ← intercept only mod
  - Explained variance level 2: $R^2_{L2} = \frac{\sigma^2_{u_0|baseline} - \sigma^2_{u_0|model1}}{\sigma^2_{u_0|baseline}}$

When we re-calculate these explained variances for subsequent models, the baseline model doesn't change!

always stay as "intercept-only model."

becuz that's the ~~total~~ variance ~~that~~ we have in our data.

but then just separated for within / between variances.

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: $R^2_{L1} = \frac{\sigma^2_{e|baseline} - \sigma^2_{e|model2}}{\sigma^2_{e|baseline}}$~~ → we cannot explain variance within-cluster w/ level 2 predictors!
  - Explained variance level 2: $R^2_{L2} = \frac{\sigma^2_{u0|baseline} - \sigma^2_{u0|model2}}{\sigma^2_{u0|baseline}}$
  - For subsequent models we thus look at the **added** explained variance

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 !! → Bcuz we're decomposing the variance in diff. way
- When adding cross level interactions (i.e., full model), we only look at the explained variance of … ?

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?

overall popularity scores over all the children over all the classes.

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	
$\sigma_{u_0}^2$		

Emmeke Aarts Multilevel analysis - lecture 1 65

65

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 decomposed to	1.22
$\sigma_{u_0}^2$		0.69

Emmeke Aarts Multilevel analysis - lecture 1 66

66

Interpretation of coefficients } If you're a boy w/ 0 extraversion, then your predicted popularity score is 2.14. If you're a girl, your predicted score increases by 1.25. & every extra point on extraversion increases additional 0.44.

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:	M1: intercept only	M2: 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

σ_e^2: residual var. goes down quite a lot. almost becomes half. → w/ this level 1 predictors, we can explain a lot why there're differences within a class.

$\sigma_{u_0}^2$: intercept var. doesn't reduce much. so we're not really explaining the differences between classes.

reduce by a lot. explain lots of within

not really explaining

we explain 52% in level 1, & only 11% in level 2, by adding these predictors.

Emmeke Aarts Multilevel analysis - lecture 4 67

67

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:	M2: level 1 predictors	M3: 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

> w/ every year of teacher exp, in average the popularity scores of classes increase by 0.09. This only relates to the class average!

> σ_e^2 res. var stays the same, but we explain a lot of $\sigma_{u_0}^2$, why popularity differs between classes.

stays the same

explains additional 47% !!

Emmeke Aarts Multilevel analysis - lecture 4 68

68

4. Random Coefficient Model : Add random slopes!

```
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*

bcuz we have this sig. slope variance, we add the cross-level interaction!

Model:	M₄: 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:	M₄: with random slope		M₅: 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	

> This now means w/ 0 teacher exp, every extraversion point will give you 0.8 additional popularity scores.

> If you have 0 extraversion, every year on teacher exp, is gonna give class average 0.23 extra on popularity scores.

> The effect of extraversion on popularity w/ every extra year on teacher exp. will decrease by 0.03 ⇒ teacher can somehow mitigate the effect of extraversion on popularity 'within' the class!

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)