RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

Università della Svizzera italiana

Conceptual bases of Multi-level Modelling (MLM)

Lugano, 19th of October, 2020

Barbara Antonioli Mantegazzini

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n° 824091

Outiline of the presentation

RISIS

First of all: Who Am I?

This morning:

A. What is Multi-Level Modeling? An introduction.

B. Data structures and Multi-Level Modelling questions.

C. Why Multi-Level Modelling? Outline the motivations for the method ant its benefits.

D. Multi-Level Modelling at a glance:

- Basic assumptions of the paradigm

    I. Variance component model

    II. Random intercept model

    III. Random slope model.

E. How to use Multi-Level Modelling? A brief overview of estimation.

Università della Svizzera italiana

A. What is multilevel modelling?

Multi-level Modeling is:

- an approach/way of thinking
- a methodology
- a statistical technique or tool

It is also known as:

- hierarchical linear modeling
- mixed modeling
- random coefficient modeling
- variance component model

And is different from single-level methods such as:

- OLS
- ANOVA

A. What is multilevel modelling?

MLMs explicitly account for the complex (grouped/hierarchised) structure of data.

…..but, **what are hierarchies?**

A hierarchy consists of lower-level observations nested within higher level(s).

The lowest-level measurements are said to be at the *micro level*; all higher-level measurements at the *macro level*.

HIERARCHIES ARE ALL AROUND US**, real world generates multilevel structures:** population structures or naturally occurring hierarchies.

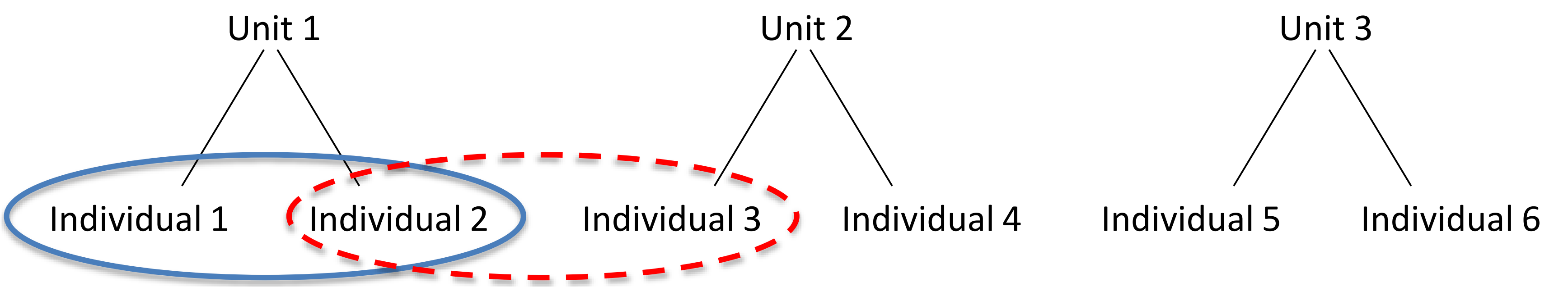
- Students in Classes in Schools; Employers within Firms; Patients in Clinics in Health authorities; Voters in Polling districts in Constituencies; People in Cities in Countries.
- But we also can get: strict hierarchies **imposed** during research design and / or during data collection……

A. What is multilevel modelling?

Hierarchies versus single level

Hierarchies usually imply "**dependence**" → individuals nested within a social unit **<u>are more alike</u>** than a random sample.

Multi-level is interesting because it addresses **dependence.**



Some examples of nested data…

Macro Level	Micro Level
schools	teachers
classes	pupils
doctors	patients
universities	professors
companies	teams
judges	suspects

→ Each level is a (potential) source of unexplained variability (and MLM might help in investigate on this variability)

A. What is multilevel modelling?

RISIS

- Multilevel analysis is a suitable approach to take into account the **social contexts** as well as the **individual** respondents or subjects.
- The hierarchical linear model is a type of **regression analysis** for multilevel data where the dependent variable is at the lowest level.
- Explanatory variables can be defined at any level (including aggregates of level-one variables).
- Also longitudinal data can be regarded as a nested structure; for such data the hierarchical linear model is likewise convenient.

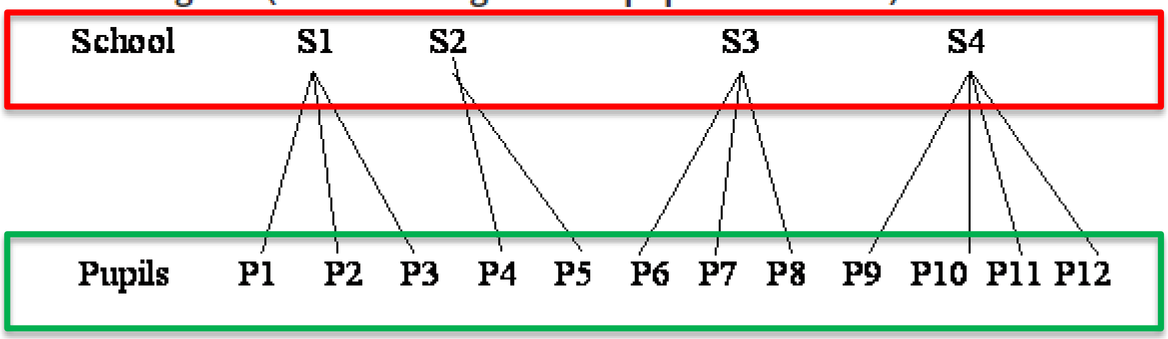
B. Data structure and MLM questions

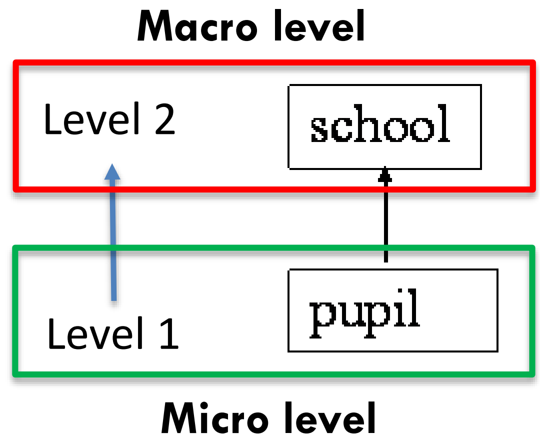
RISIS

Possible data structures with hierarchies

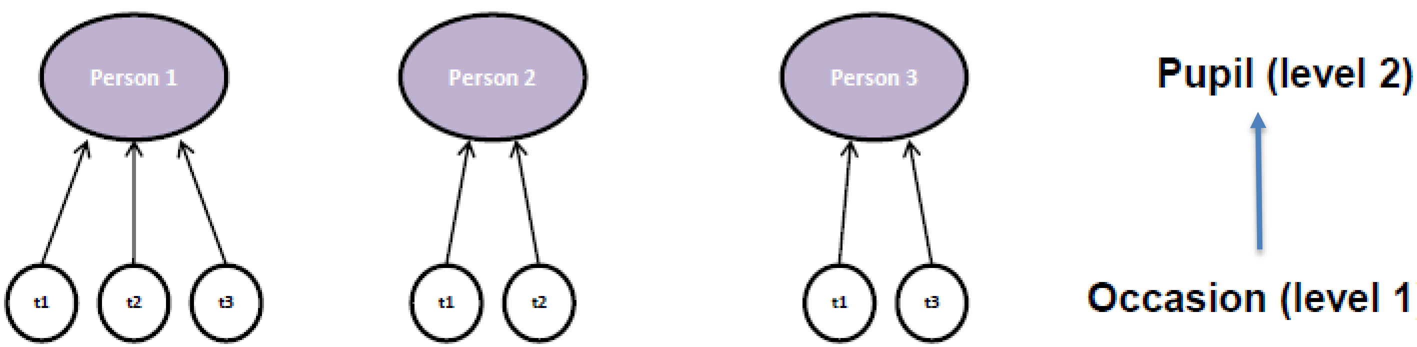
Pupils within schools. 2 level nested structure…

Unit diagram (nodes in diagram are population units)

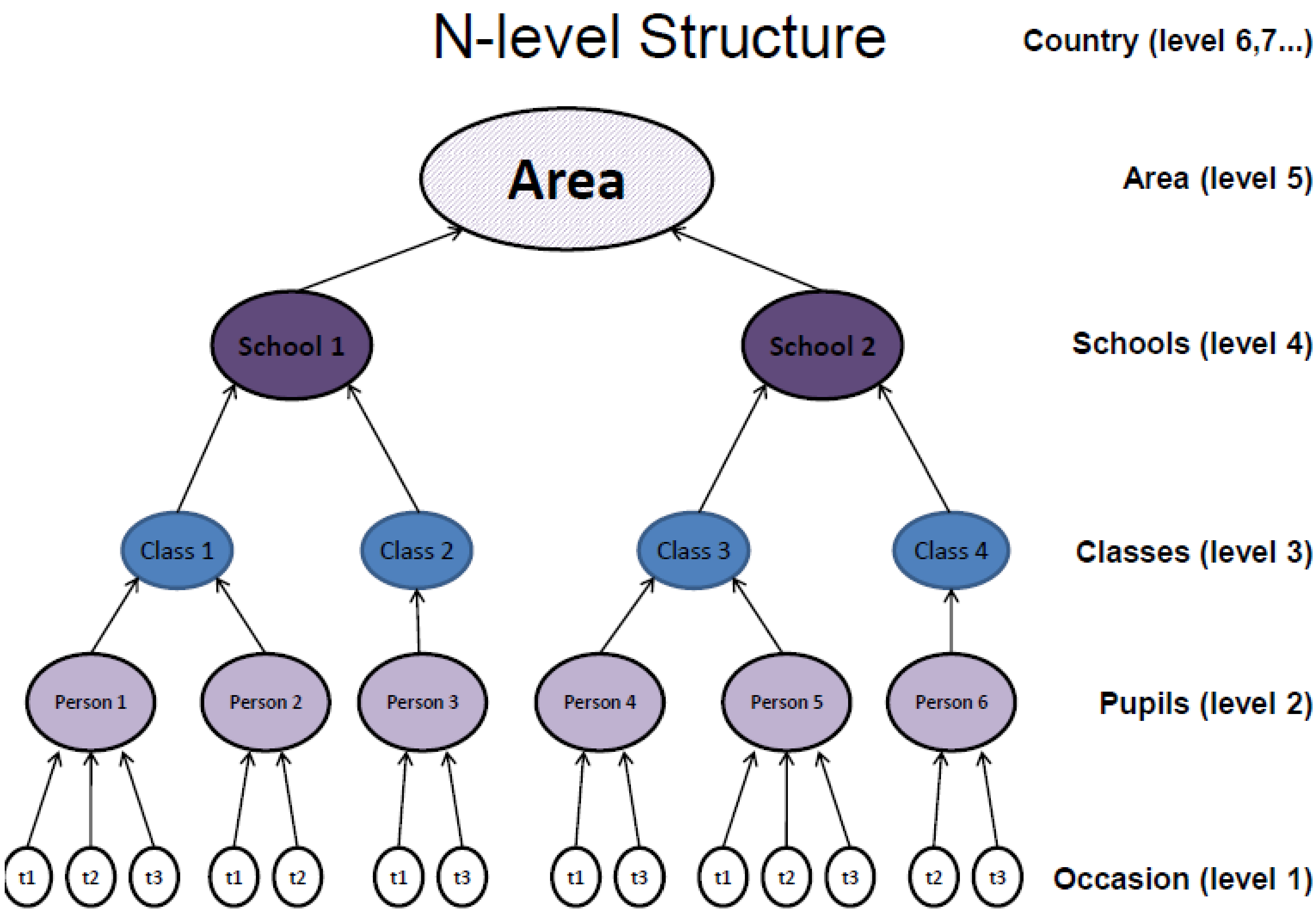




….but also 2 level repeated measure structure



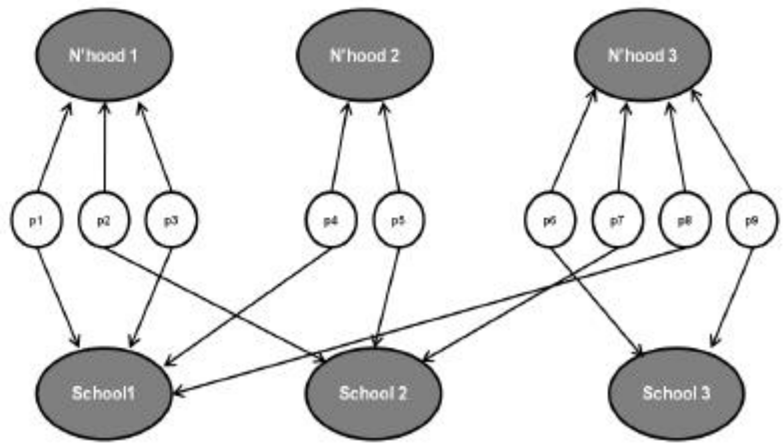
B. Data structure and MLM questions



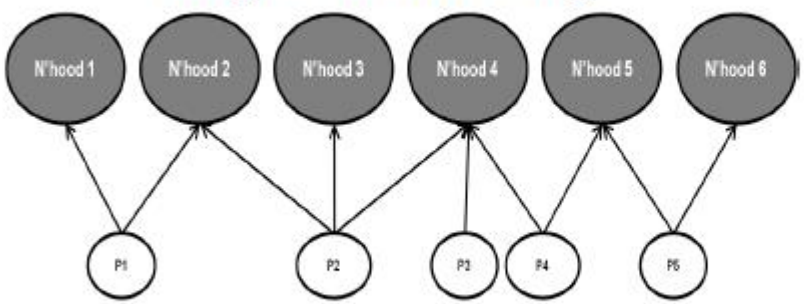
RISIS

Other non-nested structures also possible

- **Cross-classified**



- **Multiple membership**



Plus combinations...

Università della Svizzera italiana

C. Why multi-level modelling?

What do we gain with MLM?

RISIS

Before MLM, correlated observations was tackled in two ways:

- ignoring the problem or combining correlated observations into one value; (Naïve/disaggregation method);
- do not ignoring dependency but analyse group observation instead of individual observations (obtaining group averages that can then be used as an outcome in a standard regression analysis) – also called: aggregation method.

Then MLM was introduced.

Differences depend on whether or not the dataset is balanced:

- **balanced**: the only difference between the methods is observed in SE of the regression coefficients.
- - **unbalanced**: difference between the regression coefficients and the corresponding SE in the three methods.

Table 3.1. Results of a 'naive/disaggregated' analysis, a multilevel analysis, and an 'aggregated' analysis on a dataset in which a cluster randomisation is performed, i.e. the randomisation is carried out at the medical doctor level

	Intervention effect	Standard error	*p*-value
Balanced dataset[1]			
'Naive/disaggregation'	0.259	0.121	0.032
Multilevel analysis	0.259	0.213	0.224
'Aggregation'	0.259	0.225	0.265
Unbalanced dataset[2]			
'Naive/disaggregation'	0.176	0.137	0.199
Multilevel analysis	0.126	0.218	0.563
'Aggregation'	0.087	0.228	0.707

[1] In the balanced dataset 200 patients were included, equally divided among the 20 medical doctors.

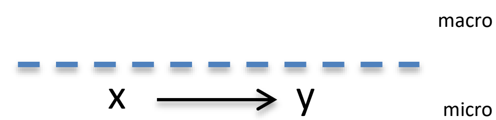
[2] In the unbalanced dataset, for half of the medical doctors only six patients were included, resulting in a total of 160 patients.

Università della Svizzera italiana

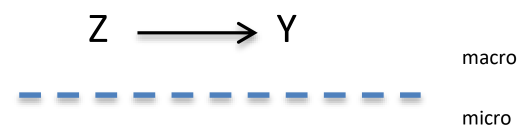
C. Why multi-level modelling?

Dependence as an interesting phenomenon..

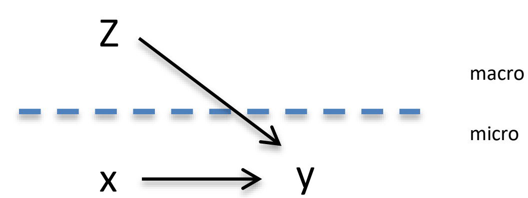
- Micro-level proposition: The effect of socio-economic status on pupil motivation



- Macro-level proposition: The effect of increased governmental spending on overall student performance



- Multi-level proposition: The effect of teacher efficacy (macro-level) on pupil motivation (micro-level)



In social sciences, ecologic (context/macro) data a usually include variables presenting a territory area a unit of analysis (region, province, electoral college, etc…) or, more generally, territorial aggregation.

C. Why multi-level modelling?

Why can't we just use (and compare) group averages?

Because it does not allow to consider relationships **across different levels** → The feature that distinguishes multilevel models from classical regression is in the **modeling of the variation between groups**.

See the "Robinson effect" example:

The relationship between race and illiteracy from Robinson (1950):

- He investigated on illiteracy rates in the US in 1930;
- A high correlation between the percentage of blacks living in a state and the state's illiteracy rates
- So, from this can we conclude that blacks are much likely than non-blacks to be illiterate?

Table 5. Percent Black and Percent Illiterate, South/Non-South: 1930 U.S.[a]

Region	Percent Black[b]	Percent Illiterate[b]
South	24.7	8.3
Non-South	3.0	2.7

[a] Source: U.S. Census, 1930.
[b] Population ten years and older.

C. Why multilevel modelling?

Ecological fallacy

Robinson argued that results of research in which aggregates are the units of analysis **cannot be used to infer relationships among individuals** and that a large proportion of research that employs aggregate data does so only because information on individuals is not available.

Correlations for Foreign Born and Illiteracy:

- Individual level analysis shows that there is a positive relationship between being foreign-born and illiterate;

  → Logical conclusion since natives can have a better command of the language
- If aggregate this to the state level, you may find that states with more foreign born have lower illiteracy rates.

C. Why multilevel modelling?

RISIS

At the opposite: why can't we just use individual-level analysis?

Associations between variables at the individual level may differ from associations between analogous variables measured at the group level:

- e.g. increasing individual level income is associated with decreasing coronary heart disease mortality;
- At the country level, increasing per capita income would be associated with decreasing coronary heart disease mortality.

But this is a fallacy!

- across countries, increasing per capita income may actually be associated with *increasing* coronary heart disease mortality

….Atomistic fallacy

Università della Svizzera italiana

C. Why multilevel modelling?

RISIS

Why can't we just use Dummies?

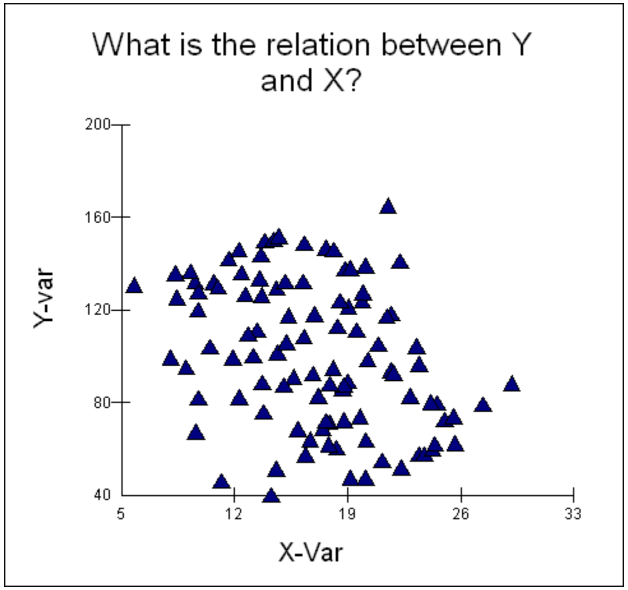
Analysis of covariance (fixed effects model) → Includes dummy variables for each and every group.

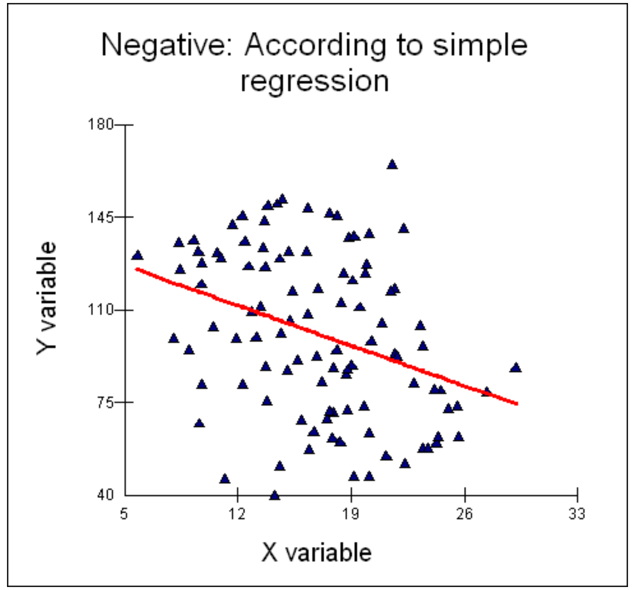
Problems:

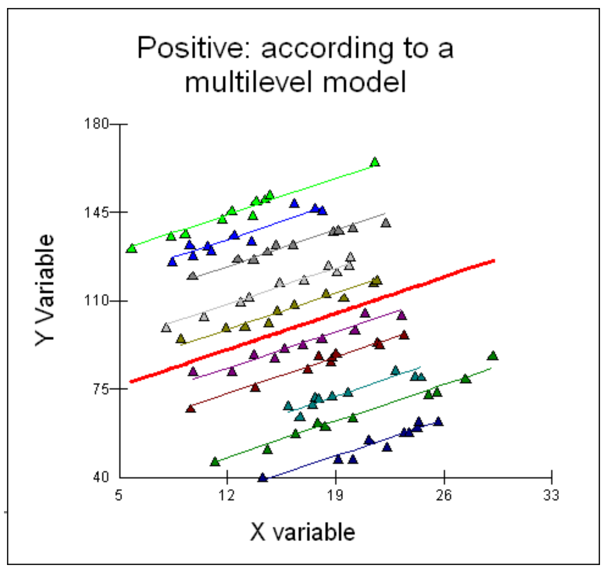
- What if number of groups very large, eg households?
- No single parameter assesses between group differences;
- Cannot make inferences beyond groups in sample;
- Cannot include group-level predictors as all degrees of freedom at the group-level have been consumed (not parsimonious);
- Acknowledge group differences but treat it is a nuisance;
- Cannot help us understand causal heterogeneity between groups.

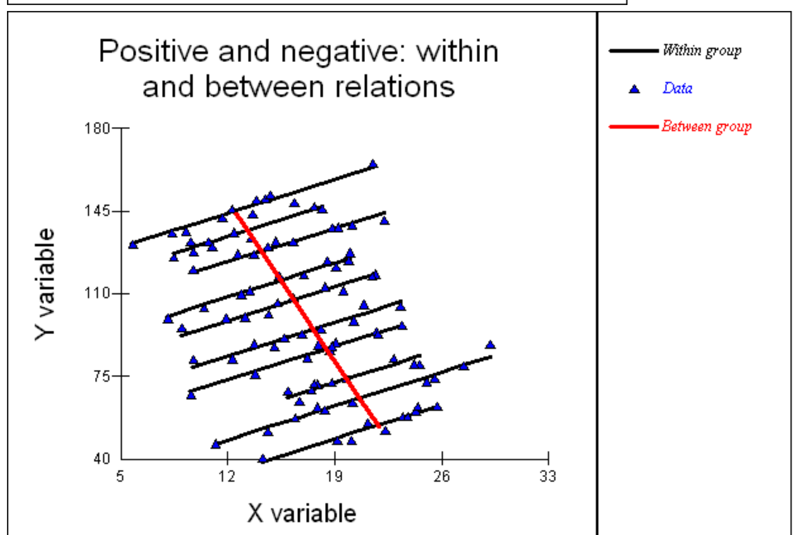
Università della Svizzera italiana

C. Why multilevel modelling?





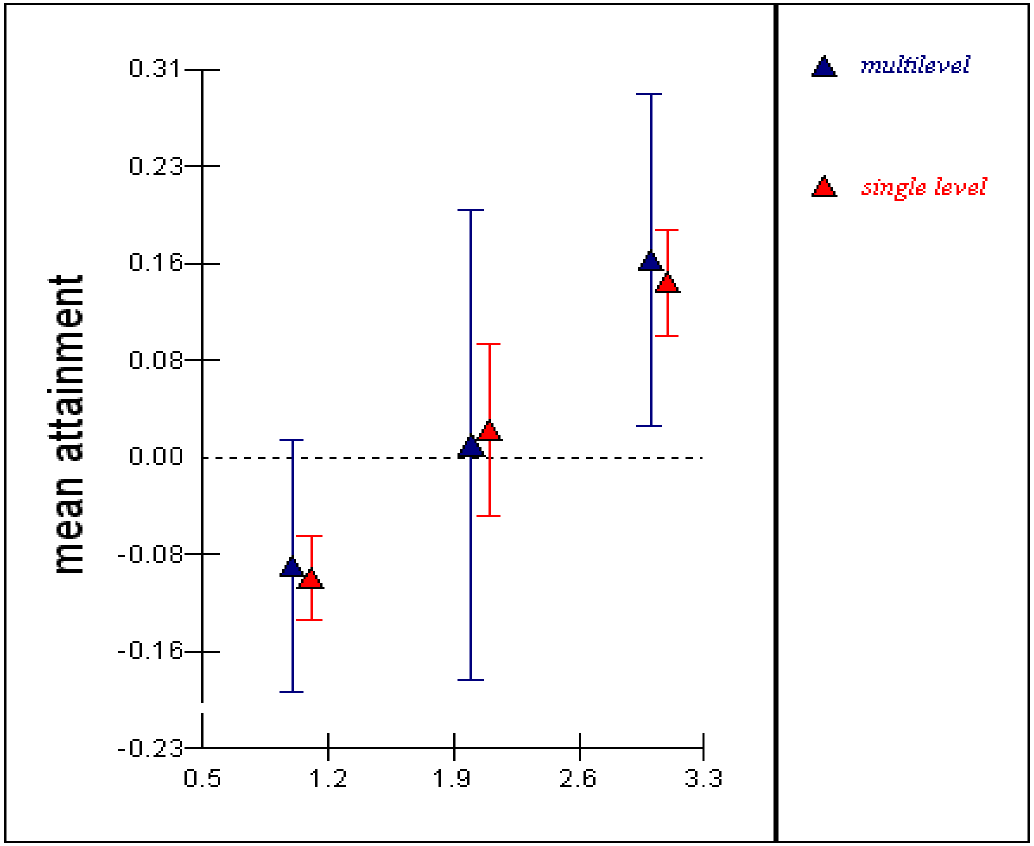




Sometimes: single level models can be seriously misleading!

C. Why multilevel modelling?

RISIS



If we fit a single level model we get incorrect uncertainty intervals leading to incorrect inferences

Single level model ignores the clustering effect of schools

Multilevel model automatically corrects the SE

Università della Svizzera italiana

D. Multi-Level Model at a glance

Basic assumptions of the paradigm

- **Differentiate levels of variables**

Variables may distinguish between individuals but also between groups.

Most level-1 variables have an individual-level as well as a group-level aspect.

Total variability = within-group variability + between-group variability

- **Groups differ**

The effect of "X" (individual-level explanatory variables) on Y (outcome variable) can differ from group to group (macro level) and this group differences are interesting and informative

(fixing of ECOLOGICAL and ATOMISTIC FALLACIES)

- **MLM model interaction between individual relationships and group relationships** (linking context to individual)

D. Multi-Level Model at a Glance

Multi-level is a "Mixed Effect" Model with:

- Fixed Effect Component
  - intercepts/slopes describe whole population
  - (i.e. the population mean).
- Random Effect Component
  - intercepts/slopes can vary across subgroups

    (different level groups means)

D. Multi-Level Model at a Glance

More detailed:

West, Welch, and Gatecki (2015) provide a good definition of fixed-effects and random-effects

"**Fixed-effect parameters** describe the relationships of the covariates to the dependent variable **for an entire population**, **random effects** are specific to clusters of subjects within a population."

→ the covariates are the independent variables in the model.

- **Fixed factors** are the independent variables that are of interest to the study, e.g. treatment category, sex or gender, categorical variable, etc.
- **Random factors** are the classification variables that the unit of analysis is grouped under, i.e. these are typically the variables that define level 2, level 3, or level N. These factors can also be thought of as a population where the unit measures were randomly sampled from, i.e. (using a school example) the school (level 2) had a random sample of students (level 1) scores selected to be used in the study.

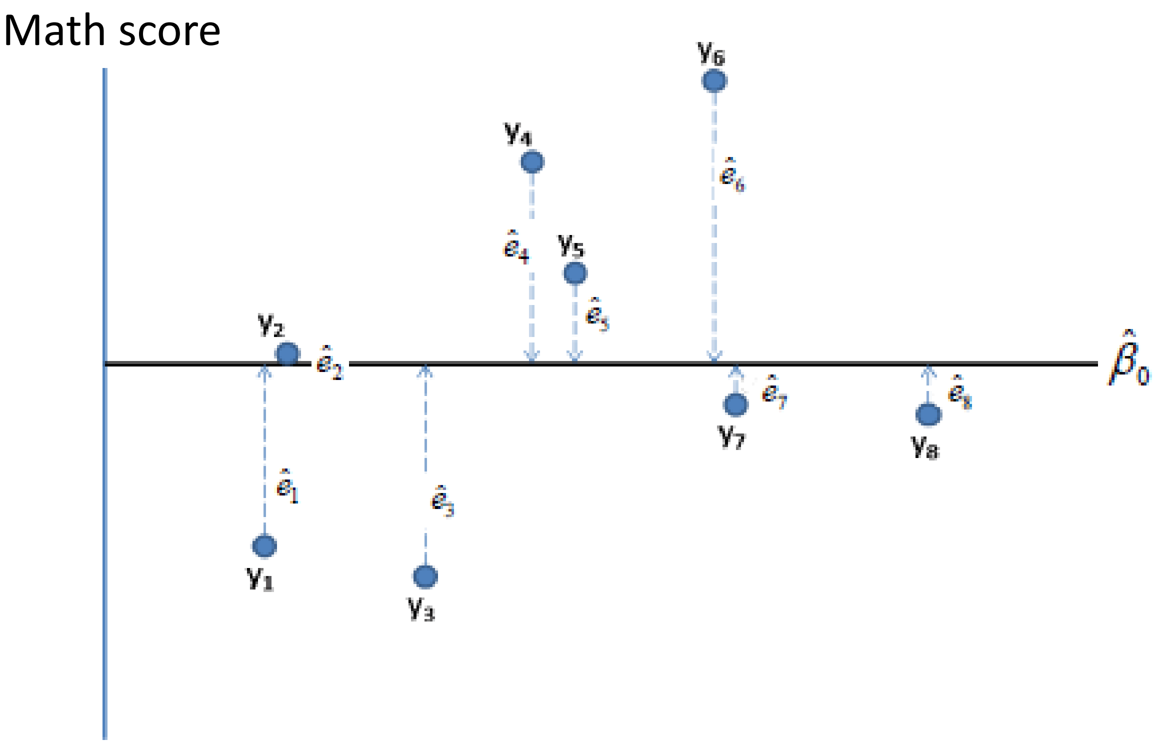
D. Multi-Level Model at a glance

Basic assumptions of the paradigm

(focusing on linear model) → If we consider linear MLM as an extension of linear regression, all the assumptions for the latter must hold for the first:

- Continuous outcome is normally distributes (i.e. residuals should be normally distributed) → investigation by plot (Goldstein and Healy, 1995);
- Random intercepts and random slopes must be normally distributes;
- Pay attention to outliers and data-points;
- Residuals are usually correlated due to the use of a single level instead of multi-level (and this will be fixed with MLM).

Single-level model for math score



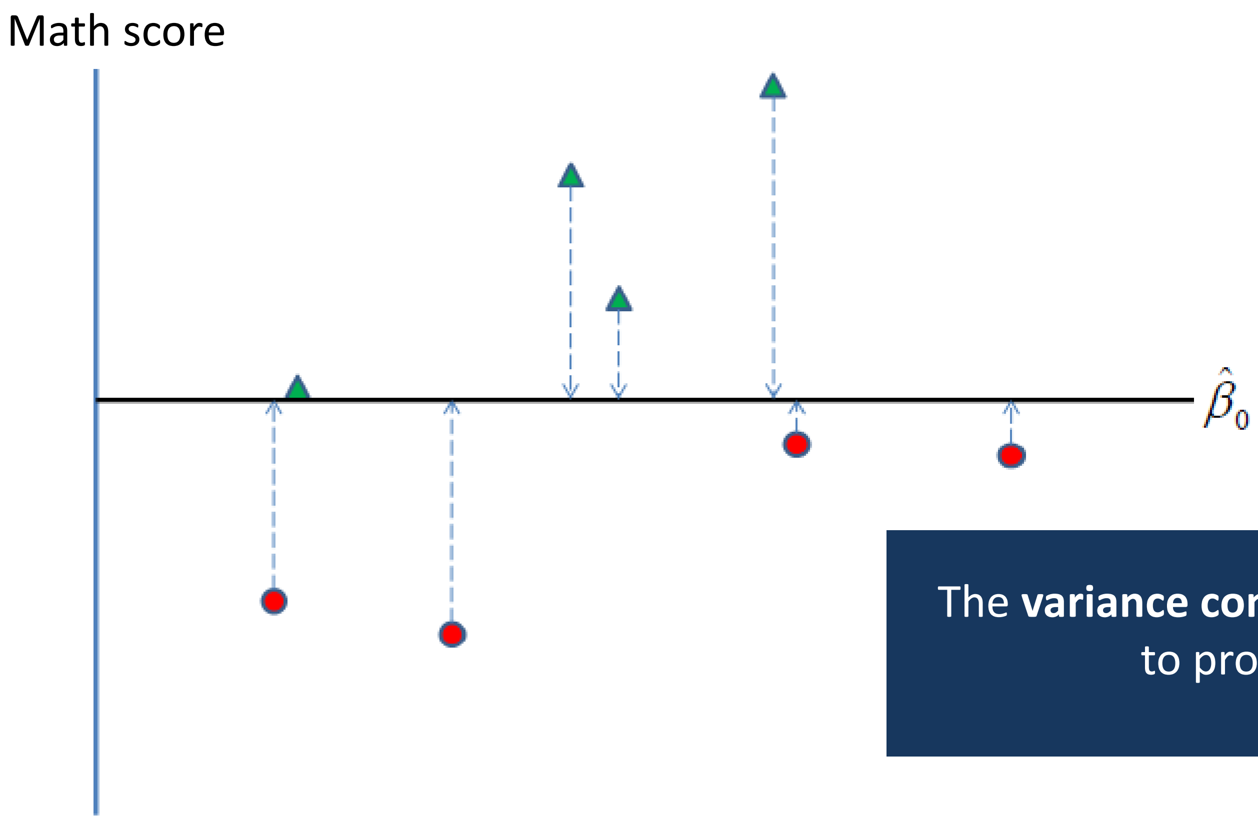
$$y_i = \beta_0 + e_i$$

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

y_i = the math score for the i_{th} individual

- β_0 is the mean math score in population and ε_i is residual for i_{th} individual (I = 1, 2...,n)
- Assume e_i are approximately normal with mean 0. The variance summarises distribution around the mean.

Single-level model for mean math score

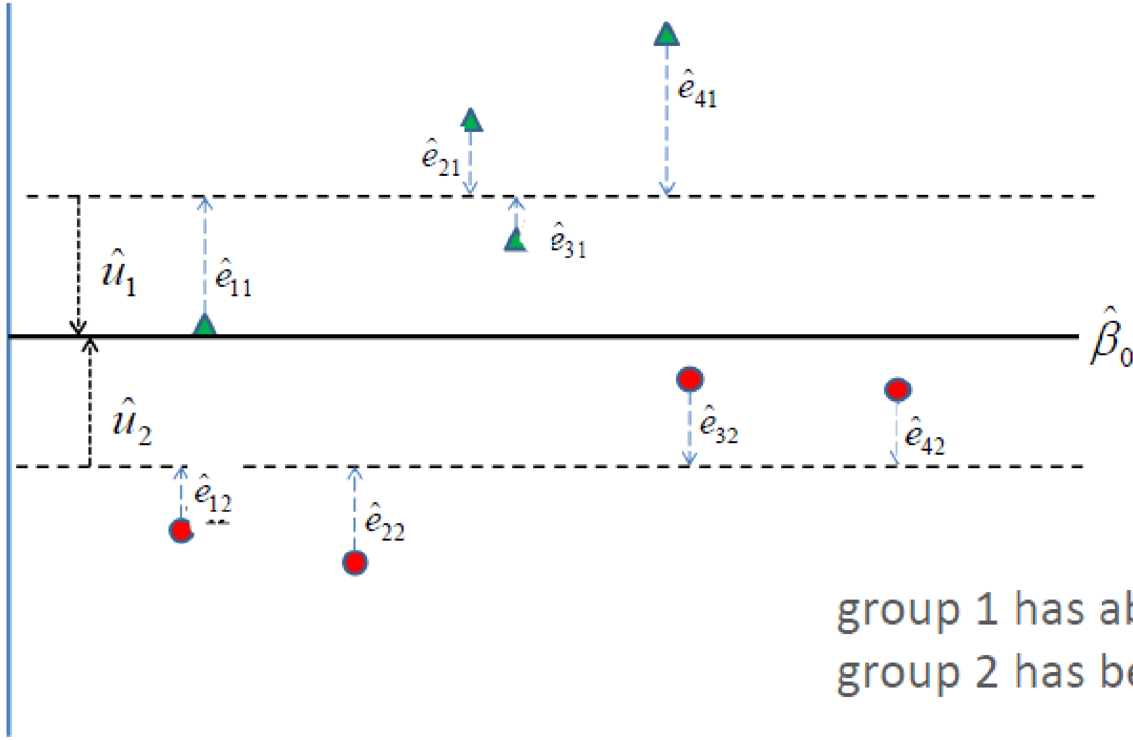


The **variance components model** can be used to provide an initial estimate of the contribution of groups

- But suppose we know our observations come from different groups (e.g. schools), j=1,…, J
- Shown here are two groups (in practice, there will be many more)
- We can capitalise on this additional information and improve our model

I. Variance components model (aka "*Empty model*")

Math score



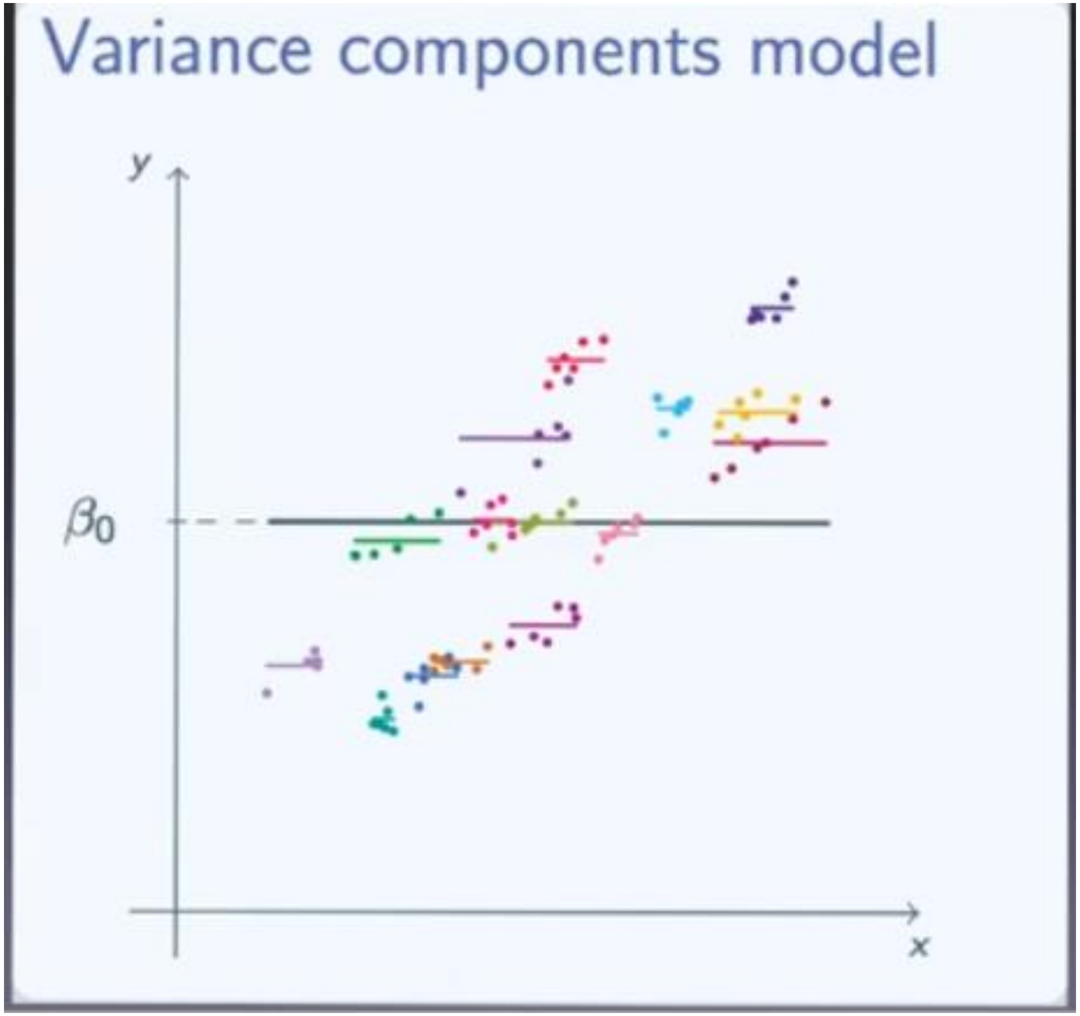
$$y_{ij} = (\beta_0 + u_j) + e_{ij}$$
$$u_j \sim N(0, \sigma_u^2)$$
$$e_{ij} \sim N(0, \sigma_e^2)$$

Basic linear regression model with no explanatory variables

group 1 has above-average mean (positive u)
group 2 has below-average mean (negative u)

- y_{ij} math score for the i_{th} individual in j_{th} group (1,2,...n).
- β_0 is the average math score across all groups
- u_j is the group mean deviations from overall average math score
- e_{ij} is the individual deviations from group means
- $\beta_0 + u_j$ is the average math score in group j

I. Variance components model (aka "*Empty model*")

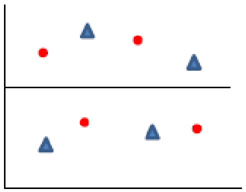


Variance Partition Coefficient (VPC)/Interclass Correlation Coefficient (ICC)

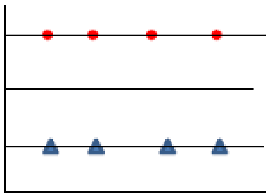
RISIS

$$VPC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

- VPC/ICC tells us how important group level differences are (e.g. what proportion of variance is at the group level? – between groups)



- VPC = 0 if no group effect (no between groups effects)



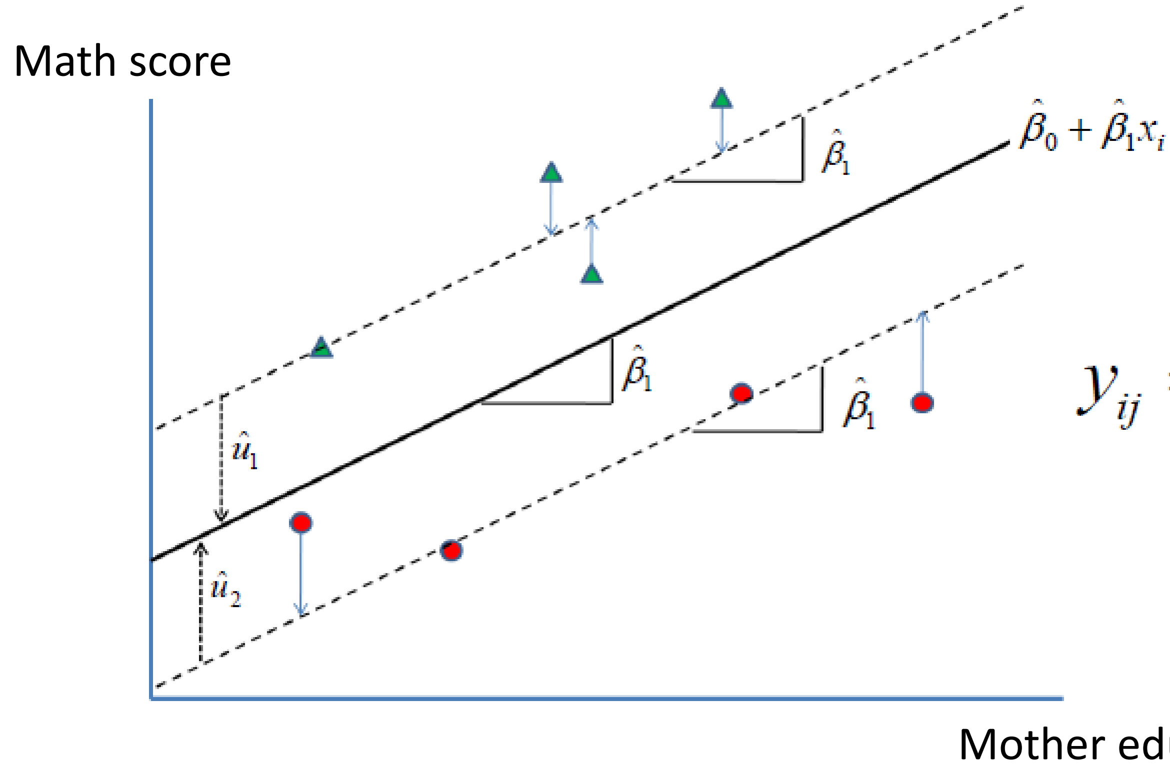
- VPC =1 if no within group differences

$$r = \frac{s_s^2}{s_s^2 + s_e^2}$$

Rule of thumbs: substatial clusterization if ICC>0.5
In general: any value above zero is related to hierarchies

Adding an explanatory variable: random intercept model

RISIS



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

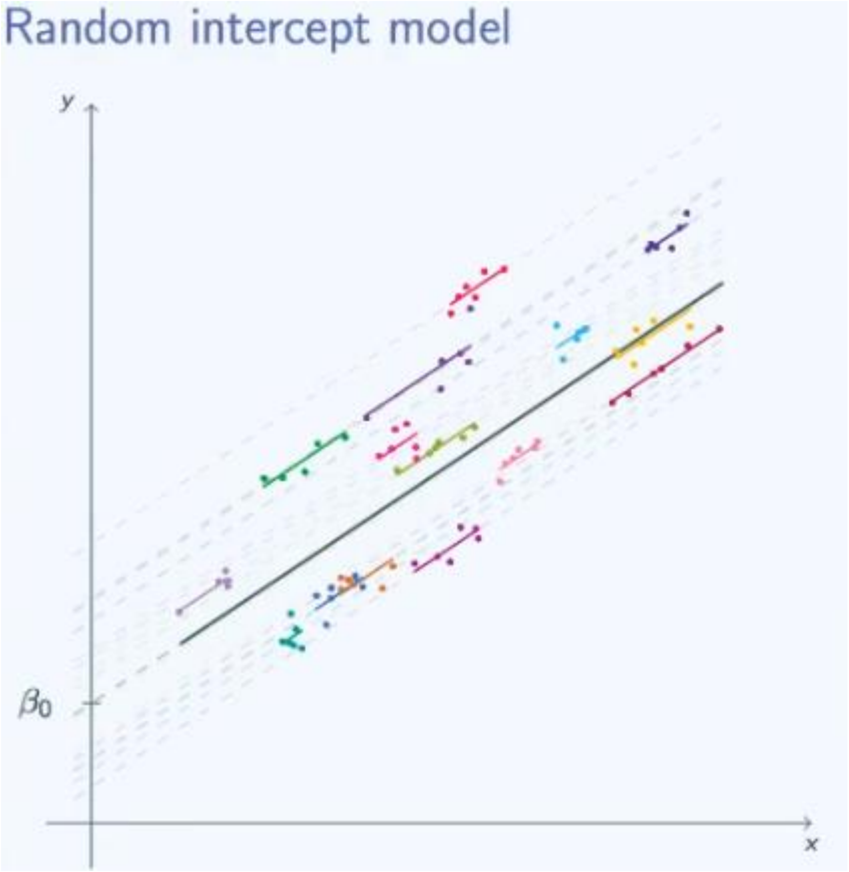
$$u_j \sim N(0, \sigma_u^2)$$

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

Overall relationship between average math score and parent education families is represented by the intercept β_0 and slope β_1 (fixed part)

- For group j, the intercept is $\beta_0 + u_j$ (either above or below average)
- Individual deviations from group line e_{ij} and group deviations from average line u_j (random part, with means 0 and variances σ^2_e and σ^2_e

Adding an explanatory variable: random intercept model



RISIS

II. The Random intercept model

A **random intercepts model** is a **model** in which **intercepts** are allowed to vary and, therefore, the scores on the dependent variable for each individual observation are predicted by the **intercept** that varies (random) across groups.

$$Y_{ij} = B_{0j} + B_1 x_{ij} + U_{0j} + R_{ij}$$

- Y_{ij}: Dependent Variable (outcome)
- B_{0j}: Average group intercept plus group-dependent deviation **(random differences between groups)**
- B_1: **B1 is common to all groups**
- x_{ij}: Explanatory variable
- U_{0j}: Random effect at group level
- R_{ij}: Random effect at individual level

With Multi-Level analysis we do not estimate separate intercept for each group but only for their variance: → **only a variance parameter is estimated!**

Note: *i* denotes the individual level, *j* denotes the group level

D. Multi-Level Model at a glance

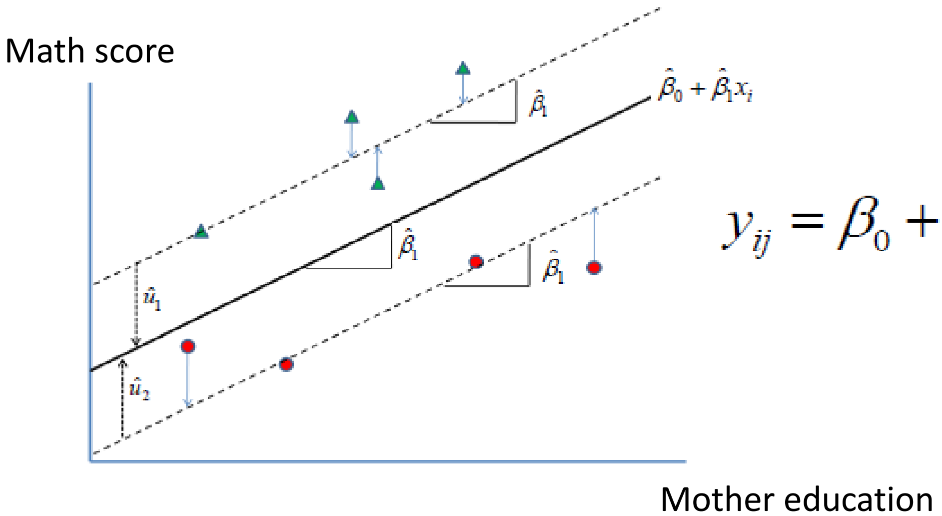
- Variance component model: can be used to provide an initial estimate of the contribution of groups.
  - ("*empty model*")
- Random intercept models: allows us to include explanatory variables at the individual and group level to explain variation in our dependent variable.
  - Explores group level variables simultaneously with individual

D. Multi-Level Model at a glance

RISIS

Random coefficient models

- Allowing individual-level relationships to vary across groups
- Linking individual and group level explanations
- Cross level interactions



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

$$u_j \sim N(0, \sigma_u^2)$$

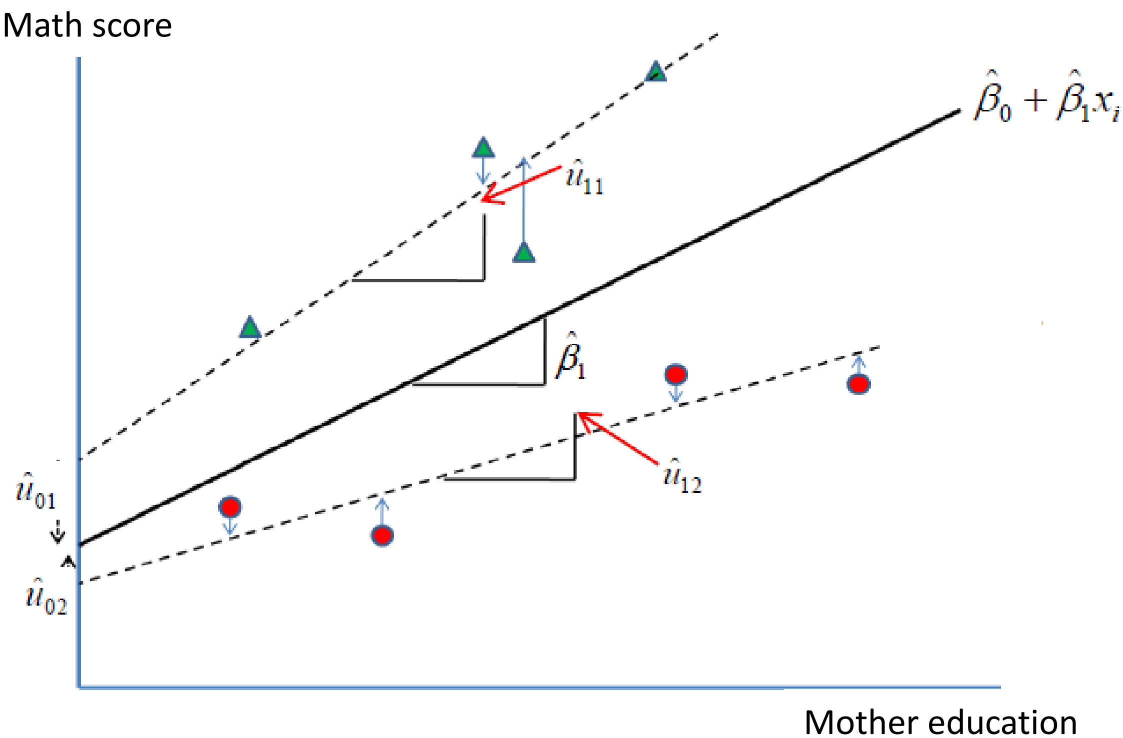
$$e_{ij} \sim N(0, \sigma_e^2)$$

- Previously, we allowed average math scores might to be different in each school….
- ….but assumed relationship with mother education constant.
- Now, what if a unit increase in education leads to **different** increases in math score?

Università della Svizzera italiana

Random slope model (Random coefficient model) of math score and mother education

RISIS

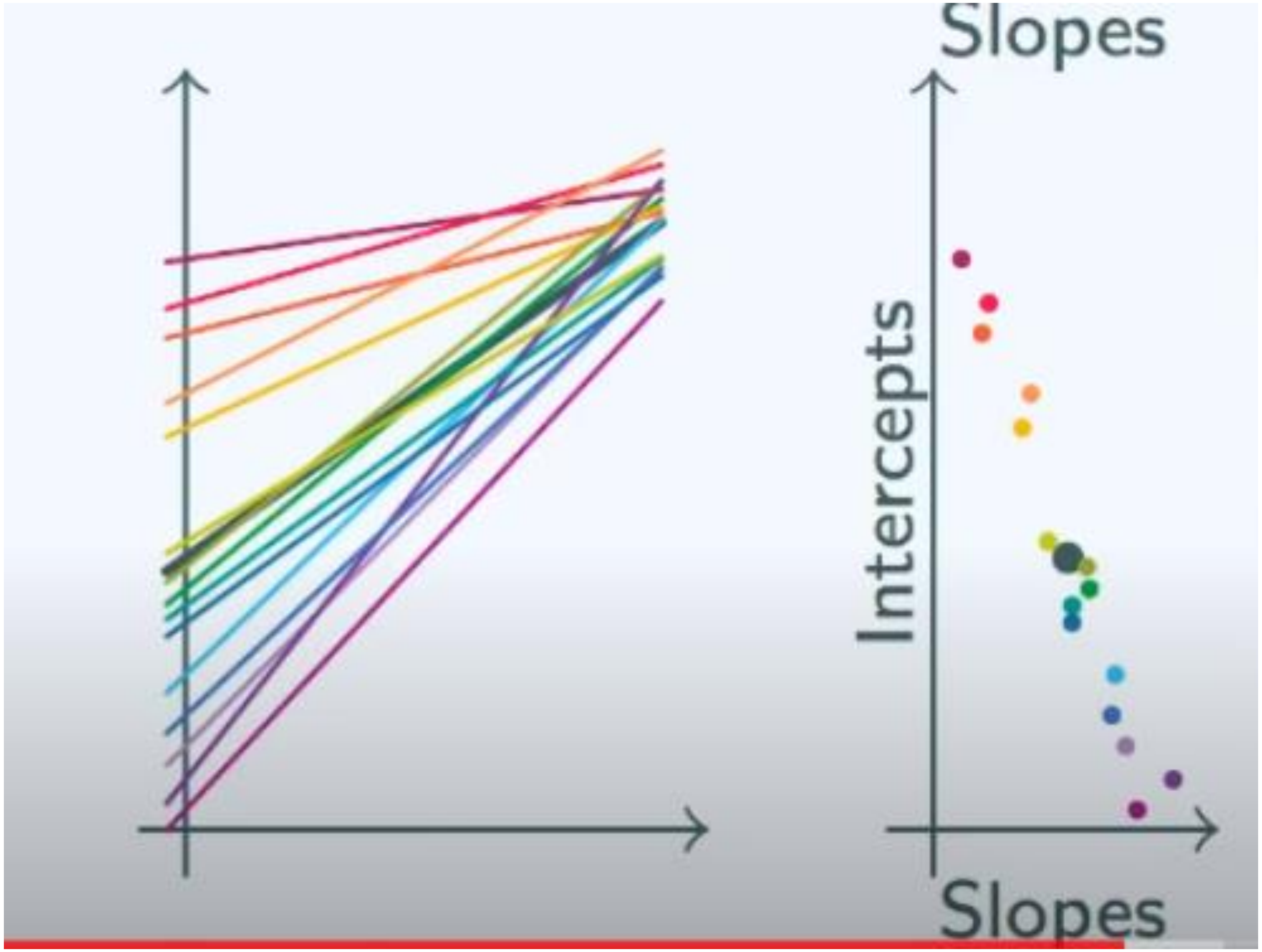


is the Slope of group j

Is the average relationship between math score and mother education

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

Random slope model (Random coefficient model) of math score and mother education



D. Multi-Level Model at a glance

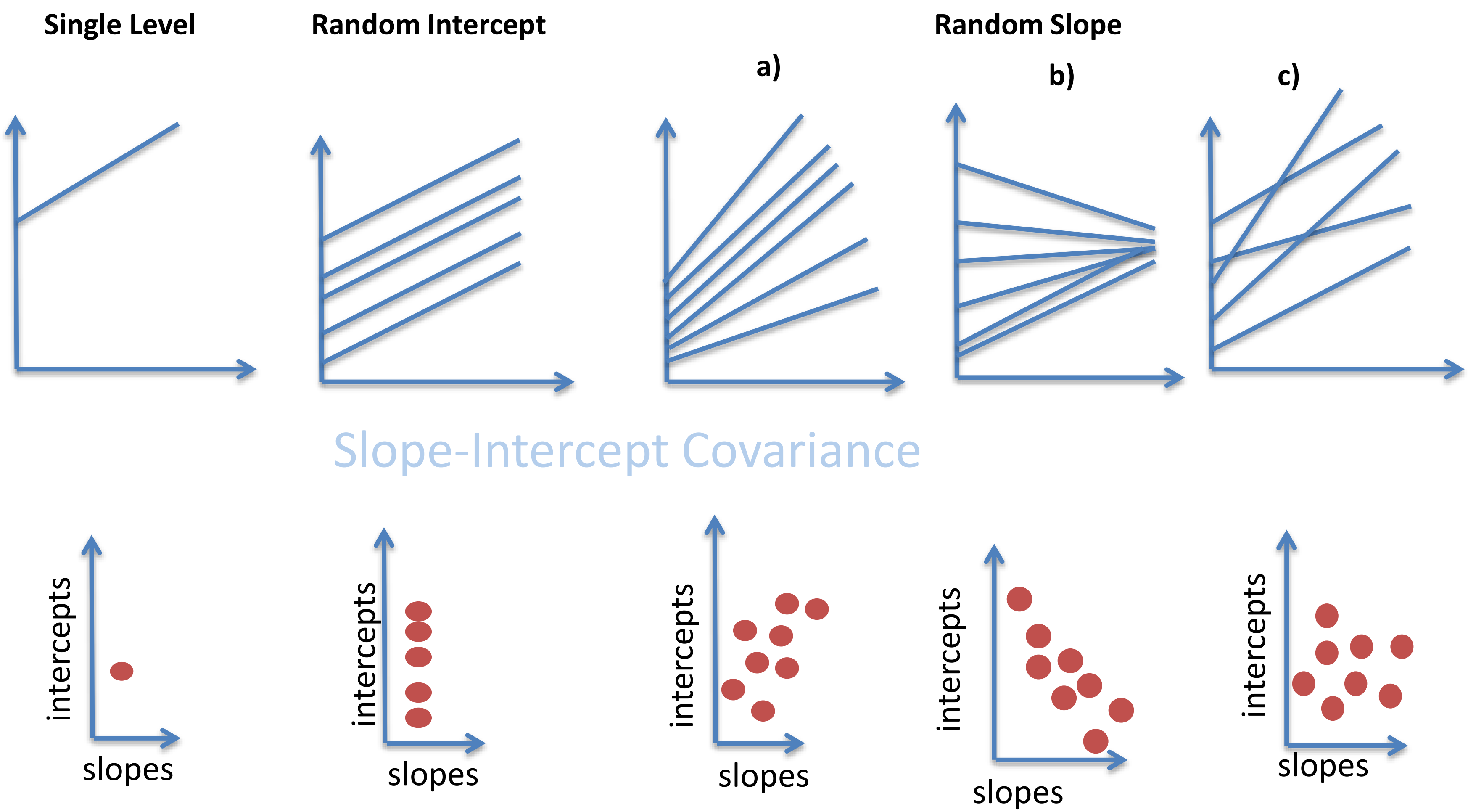
II. The Random coefficient model

RISIS

$$Y_{ij} = B_{0j} + B_1 x_{ij} + U_{0j} + U_{1j} x_{ij} + R_{ij}$$

- Builds on the Random Intercept Model
- Unlike a random intercept model, a random slope allows each group line to have a different slope
- So, the random slope model **allows the explanatory variable to have a different effect for each group**.
- Example: The effect of IQ is positive in all schools but, the size of the effects varies across schools
- How do we achieve this? By adding a random term to the coefficient of xij, so it can be different for each group.

Random intercepts and slopes



Cross level interactions

RISIS

- Allow the effect of an explanatory variable on y to depends on the value of another (grouping) variable

- Do people experience context differently?
  - Place individuals directly within their context

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_j + \beta_3 x_{1ij} x_{2j} + u_j + u_{1j} x_{ij} + e_{ij}$$

- Often included when individual level variables has a random coefficient

Università della Svizzera italiana

Summary

RISIS

- In linear random-intercept models, the overall level of the response, conditional on X, could vary across clusters;
- In random coefficients models, we also allow the marginal effect of the covariates to vary across clusters;
- This is exactly analogous to the different slopes
- Random coefficient models can be used to more accurately account for differential associations between x and y across groups
- Cross-level interactions connect individual and group level explanations
- Models also available for non-normal data (e.g. binary, categorical, ordered categories, Poisson)

Università della Svizzera italiana

E. How to use Multi-Level Model

When building your ML Model, you will need to tailor it:

- Type of response variable
- Data structure
- Variance structure

More generally, you have to take into consideration that:

- Explanatory, approximation of reality
- There are no "correct models"
- Models are not "kitchen sink" approach
- It is crucial to combine data with theoretical framework framing the question;
- Many models are better than one

E. How to use Multi-Level Model

RISIS

Multi-level Models can be fitted for a variety of response variables…..

- continuous
  - e.g. student test scores
- Binary
  - e.g. unemployed/employed
- nominal categorical
  - e.g. vote for party A, B, or C
- ordinal categorical
  - e.g. attitudinal scale (strongly disagree, disagree…)
- Counts (Poisson)
  - e.g. mortality rate
- Duration or Survival
  - e.g. duration of marriage or unemployment

Università della Svizzera italiana

E. How to use Multi-Level Model

RISIS

... and can be fitted for a variety of data structure

- multivariate
- three-levels
- longitudinal
- small groups
- cross/multiple classification

Università della Svizzera italiana

E. How to use Multi-Level Model

RISIS

Estimation strategies

- Maximum likelihood
  - frequentist procedure
  - Restricted maximum likelihood (REML)
- More advanced estimations use Bayesian simulation methods (Markov Chain Monte Carlo)
  - prior parameter distribution
  - more time-consuming but more flexible
  - better for more complex models

Università della Svizzera italiana

RISIS

Model Diagnostics

- In OLS regression, the explained proportion of variance is captured by R-squared
- Calculating R-squared for each level is not intuitive to interpret
- So multilevel models can use the following diagnostics:
  - Estimated Variance Parameters
  - log pseduo-likelihood
  - ICC, AIC and BIC

Università della Svizzera italiana

CREDITS

Slides partially based on previous presentation by

Dr. Anne-Marie Jeannet and on free available documentation by NCRM – National Centre for Research Methods and by Dr. Mike Crowson.

RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

THANK YOU !

WWW.RISIS2.EU

ZENODO.ORG/COMMUNITIES/RISIS

CONTACT@RISIS2.EU

@RISIS_EU

FACEBOOK.COM/RISIS.EU

RISIS2 EU PROJECT