Introduction to mixed-effects models

(for hierarchical data)

Nora Wickelmaier

Leibniz-Institut für Wissensmedien

2025-06-24

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)
- I will explain the general concepts with the slides

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)
- I will explain the general concepts with the slides
- We will switch to R and use the lme4 package to fit the models

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)
- I will explain the general concepts with the slides
- We will switch to R and use the lme4 package to fit the models
- You will use R to fit an extension of the model

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)
- I will explain the general concepts with the slides
- We will switch to R and use the lme4 package to fit the models
- You will use R to fit an extension of the model
- We will discuss the results

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)
- I will explain the general concepts with the slides
- We will switch to R and use the lme4 package to fit the models
- You will use R to fit an extension of the model
- We will discuss the results
- All the materials are here:
  https://gitea.iwm-tuebingen.de/nwickelmaier/lead_lmm

Plan for today

- We will walk through an example for a hierarchical data set (students in schools)
- I will explain the general concepts with the slides
- We will switch to R and use the lme4 package to fit the models
- You will use R to fit an extension of the model
- We will discuss the results
- All the materials are here:
  https://gitea.iwm-tuebingen.de/nwickelmaier/lead_lmm

→ Try to go along in R! Ask as many questions as possible, also the ones you usually do not dare to ask (because you are supposed to know them already or something. . .)

Outline

1. Introduction to random effects

2. Hierarchical modeling

1 Introduction to random effects

Hierarchical data

- Observations often do not come from a simple random sample, but result from a hierarchical structure

Hierarchical data

- Observations often do not come from a simple random sample, but result from a hierarchical structure
  - Individuals are organized in groups (e.g., students nested in classes, or schools)

Hierarchical data

- Observations often do not come from a simple random sample, but result from a hierarchical structure
  - Individuals are organized in groups (e.g., students nested in classes, or schools)
  - Persons are observed multiple times (observations nested in persons, longitudinal data)

Hierarchical data

- Observations often do not come from a simple random sample, but result from a hierarchical structure
  - Individuals are organized in groups (e.g., students nested in classes, or schools)
  - Persons are observed multiple times (observations nested in persons, longitudinal data)
- Statistical models for this kind of data are called multilevel models, mixed-effects models, random-effects models, covariance components models, or hierarchical models

Example: Mathematics achievement study

- The `hsbdataset.txt` file contains data from the National Center for Education Statistics' (NCES) "High School & Beyond" national survey of U.S. public and Catholic high schools (Raudenbush & Bryk, 2002)

Example: Mathematics achievement study

- The `hsbdataset.txt` file contains data from the National Center for Education Statistics' (NCES) "High School & Beyond" national survey of U.S. public and Catholic high schools (Raudenbush & Bryk, 2002)
- The data set consists of information on 7,185 students from 160 schools on student performance on a mathematics test and information concerning their socioeconomic status

Example: Mathematics achievement study

- The `hsbdataset.txt` file contains data from the National Center for Education Statistics' (NCES) "High School & Beyond" national survey of U.S. public and Catholic high schools (Raudenbush & Bryk, 2002)
- The data set consists of information on 7,185 students from 160 schools on student performance on a mathematics test and information concerning their socioeconomic status
- Hierarchical data structure
    - Students are organized in schools

      y_{ij} mathematics achievement of student j in school i

      x_{ij} (relative) socioeconomic status of student j in school i (overall mean 0, centered)

Example: Mathematics achievement study

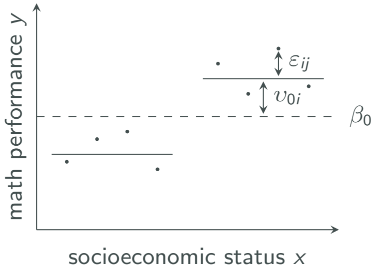
- The `hsbdataset.txt` file contains data from the National Center for Education Statistics' (NCES) "High School & Beyond" national survey of U.S. public and Catholic high schools (Raudenbush & Bryk, 2002)
- The data set consists of information on 7,185 students from 160 schools on student performance on a mathematics test and information concerning their socioeconomic status
- Hierarchical data structure
  - Students are organized in schools
    y_{ij} mathematics achievement of student j in school i
    x_{ij} (relative) socioeconomic status of student j in school i (overall mean 0, centered)
  - Two levels
    - Level 1: Student attributes
    - Level 2: School attributes

Regression with random school effects

- What is the mean math achievement of the students?
- How much do schools vary in mean math achievement?

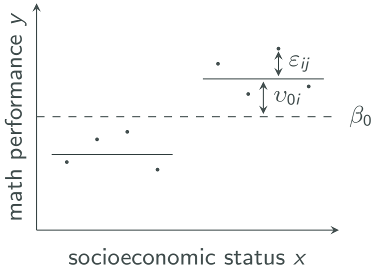
Regression with random school effects

- What is the mean math achievement of the students?
- How much do schools vary in mean math achievement?



Regression with random school effects

- What is the mean math achievement of the students?
- How much do schools vary in mean math achievement?



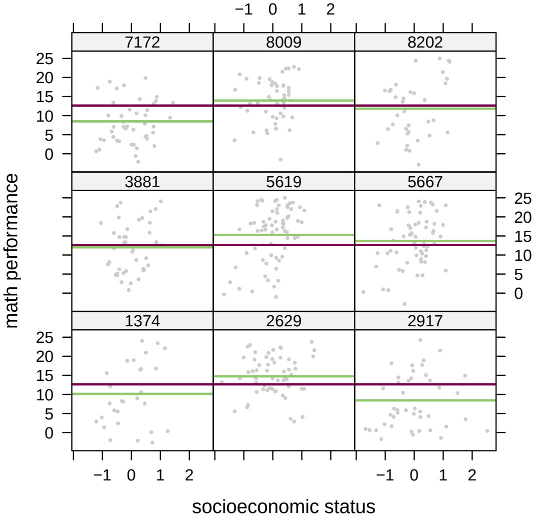
Model equation

$$
\begin{aligned}
&\text{(Level 1)} \quad && y_{ij} = b_{0i} + \varepsilon_{ij} \\
&\text{(Level 2)} \quad && b_{0i} = \beta_0 + \upsilon_{0i} \\
&\text{(2) in (1)} \quad && y_{ij} = \beta_0 + \upsilon_{0i} + \varepsilon_{ij}
\end{aligned}
$$

with $\upsilon_{0i} \sim N(0, \sigma_\upsilon^2)$ i.i.d, $\varepsilon_{ij} \sim N(0, \sigma^2)$ i.i.d, υ_{0i} and ε_{ij} independent

Null model with random intercepts

Subset of 9 schools



Random effects

- The problem of grouping observations in schools and the thereby induced dependencies is solved by introducing school effects

Random effects

- The problem of grouping observations in schools and the thereby induced dependencies is solved by introducing school effects
- For many schools this calls for (too) many parameters

Random effects

- The problem of grouping observations in schools and the thereby induced dependencies is solved by introducing school effects
- For many schools this calls for (too) many parameters
- School effects are therefore modeled as random effects (random variables) υ_{0i}

Random effects

- The problem of grouping observations in schools and the thereby induced dependencies is solved by introducing school effects
- For many schools this calls for (too) many parameters
- School effects are therefore modeled as random effects (random variables) υ_{0i}
- Only their variance σ_υ^2 has to be estimated in the model

Random effects

- The problem of grouping observations in schools and the thereby induced dependencies is solved by introducing school effects
- For many schools this calls for (too) many parameters
- School effects are therefore modeled as random effects (random variables) υ_{0i}
- Only their variance σ_υ^2 has to be estimated in the model
- The total variance of y_{ij} is decomposed into the variance between schools σ_υ^2 and within schools σ^2

Results

- The above posed research questions can be answered based on the parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$

Results

- The above posed research questions can be answered based on the parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$
  - The estimated mean math achievement of students is $\hat{\beta}_0$

Results

- The above posed research questions can be answered based on the parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$
  - The estimated mean math achievement of students is $\hat{\beta}_0$
  - The estimated variance of schools in mean math achievement is $\hat{\sigma}_v^2$

Results

- The above posed research questions can be answered based on the parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$
  - The estimated mean math achievement of students is $\hat{\beta}_0$
  - The estimated variance of schools in mean math achievement is $\hat{\sigma}_v^2$
  - The proportion of the total variance accounted for by the variance between schools is

$$\mathrm{ICC} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

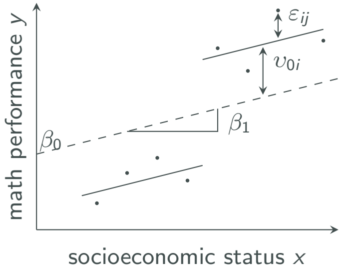
(Intra-class correlation)

Adding socioeconomic status as a predictor

- How strong is the relationship between students' socioeconomic status and their math achievement on average?
- How much do schools vary in mean math achievement for students with average socioeconomic status?

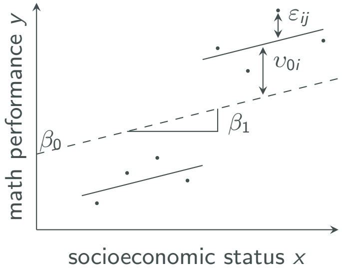
Adding socioeconomic status as a predictor

- How strong is the relationship between students' socioeconomic status and their math achievement on average?
- How much do schools vary in mean math achievement for students with average socioeconomic status?



Adding socioeconomic status as a predictor

- How strong is the relationship between students' socioeconomic status and their math achievement on average?
- How much do schools vary in mean math achievement for students with average socioeconomic status?



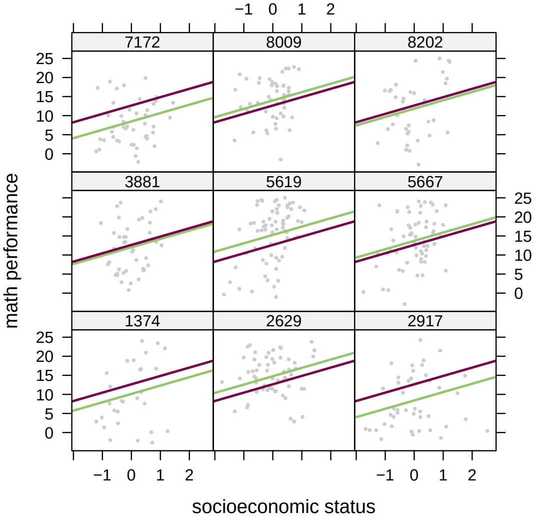
Model equation

$$
\begin{aligned}
&\text{(Level 1)} \quad & y_{ij} &= b_{0i} + b_{1i}\, x_{ij} + \varepsilon_{ij} \\
&\text{(Level 2)} \quad & b_{0i} &= \beta_0 + \upsilon_{0i} \\
& & b_{1i} &= \beta_1 \\
&\text{(2) in (1)} \quad & y_{ij} &= \beta_0 + \beta_1\, x_{ij} + \upsilon_{0i} + \varepsilon_{ij}
\end{aligned}
$$

with $\upsilon_{0i} \sim N(0, \sigma_\upsilon^2)$ i.i.d, $\varepsilon_{ij} \sim N(0, \sigma^2)$ i.i.d, υ_{0i} and ε_{ij} independent

Model with covariate and random intercepts

Subset of 9 schools



Exercise

- What would be the next possible extension of this model?

Exercise

- What would be the next possible extension of this model?
- Write down the model equations
  - What changes for the fixed effects?
  - How do the variance components for the random effects change?

Exercise

- What would be the next possible extension of this model?
- Write down the model equations
  - What changes for the fixed effects?
  - How do the variance components for the random effects change?
- How can we interpret the random slopes for this model?

Exercise

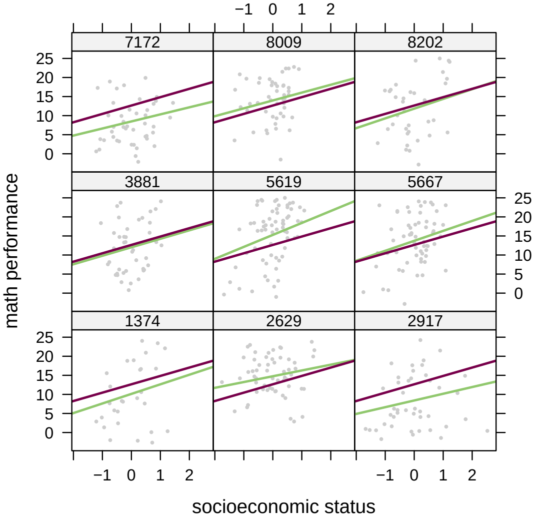
- What would be the next possible extension of this model?
- Write down the model equations
  - What changes for the fixed effects?
  - How do the variance components for the random effects change?
- How can we interpret the random slopes for this model?
- How do we add random slopes to a random intercept model using `lme4::lmer()`?

Exercise

- What would be the next possible extension of this model?
- Write down the model equations
  - What changes for the fixed effects?
  - How do the variance components for the random effects change?
- How can we interpret the random slopes for this model?
- How do we add random slopes to a random intercept model using `lme4::lmer()`?
- Fit a model with random slopes for socioeconomic status in R

Model with covariate and random slopes

Subset of 9 schools



2 Hierarchical modeling

HSB data set

Level	Variable	Description
1	`mathach`	Performance in mathematics test
1	`ses`	(relative) socioeconomic status (overall mean 0)
2	`meanses`	mean socioeconomic status of the school (overall mean 0)
1	`cses`	Centered socioeconomic status of the student (mean for each school 0, difference `ses - meanses`)
2	`school`	school ID
2	`sector`	Public (0) or Catholic High School (1)

Hierarchical regression model

Model equation

$$
\begin{aligned}
\text{(Level 1)} \quad y_{ij} &= b_{0i} + b_{1i}\, cses_{ij} + \varepsilon_{ij} \\
\text{(Level 2)} \quad b_{0i} &= \beta_0 + \beta_2 meanses_i + \beta_4 sector_i + \upsilon_{0i} \\
b_{1i} &= \beta_1 + \beta_3 meanses_i + \beta_5 sector_i + \upsilon_{1i} \\
\text{(2) in (1)} \quad y_{ij} &= \beta_0 + \beta_1\, cses_{ij} + \beta_2 meanses_i + \beta_4 sector_i \\
&\quad + \beta_3 (cses_{ij} \times meanses_i) + \beta_5 (cses_{ij} \times sector_i) \\
&\quad + \upsilon_{0i} + \upsilon_{1i} cses_{ij} + \varepsilon_{ij}
\end{aligned}
$$

with

$$
\begin{pmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_\upsilon = \begin{pmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} \end{pmatrix} \right) \text{ i.i.d}
$$

$$
\varepsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}) \text{ i.i.d}
$$

Decomposing socioeconomic status

- In this model, by decomposing the socioeconomic status according to the equation

$$ses = cses + meanses$$

its differential effectiveness is considered at each of the levels

Decomposing socioeconomic status

- In this model, by decomposing the socioeconomic status according to the equation

$$ses = cses + meanses$$

  its differential effectiveness is considered at each of the levels
- At the same time, the effect of the type of school is examined via the variable `sector`

Decomposing socioeconomic status

- In this model, by decomposing the socioeconomic status according to the equation

$$ses = cses + meanses$$

  its differential effectiveness is considered at each of the levels
- At the same time, the effect of the type of school is examined via the variable `sector`
- Notice that the formulation of the model assumes dependencies of the slope b_{1i} on both mean socioeconomic status and school type, which is captured by the interactions of `cses` with `meanses` and `sector`, respectively

Exercise

1. Compute the model in R using `lme4::lmer()`

$$
\begin{aligned}
\text{(Level 1)} \quad & y_{ij} = b_{0i} + b_{1i}\, cses_{ij} + \varepsilon_{ij} \\
\text{(Level 2)} \quad & b_{0i} = \beta_0 + \beta_2 meanses_i + \beta_4 sector_i + \upsilon_{0i} \\
& b_{1i} = \beta_1 + \beta_3 meanses_i + \beta_5 sector_i + \upsilon_{1i} \\
\text{(2) in (1)} \quad & y_{ij} = \beta_0 + \beta_1\, cses_{ij} + \beta_2 meanses_i + \beta_4 sector_i + \beta_3 (cses_{ij} \times meanses_i) + \beta_5 (cses_{ij} \times sector_i) \\
& \qquad + \upsilon_{0i} + cses_{ij} \upsilon_{1i} + \varepsilon_{ij}
\end{aligned}
$$

with $\begin{pmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_\upsilon = \begin{pmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} \end{pmatrix} \right)$ i.i.d, $\varepsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$ i.i.d

2. Interpret the parameters

Results

Fixed effects

- Mean math achievement (i.e., for a student with a mean `cses` score in a school with a mean `meanses` score) is 12.11 in Public High Schools and 13.33 in Catholic High Schools

Results

Fixed effects

- Mean math achievement (i.e., for a student with a mean `cses` score in a school with a mean `meanses` score) is 12.11 in Public High Schools and 13.33 in Catholic High Schools
- Effects of socioeconomic status at the two levels

Results

Fixed effects

- Mean math achievement (i.e., for a student with a mean `cses` score in a school with a mean `meanses` score) is 12.11 in Public High Schools and 13.33 in Catholic High Schools
- Effects of socioeconomic status at the two levels
  - The effect at the student level depends on the type of school: math achievement increases by 2.94 points in Public High Schools and by $2.94 - 1.64 = 1.30$ points in Catholic High Schools for a unit increase in cses

Results

Fixed effects

- Mean math achievement (i.e., for a student with a mean `cses` score in a school with a mean `meanses` score) is 12.11 in Public High Schools and 13.33 in Catholic High Schools
- Effects of socioeconomic status at the two levels
  - The effect at the student level depends on the type of school: math achievement increases by 2.94 points in Public High Schools and by $2.94 - 1.64 = 1.30$ points in Catholic High Schools for a unit increase in cses
  - Higher math achievements are obtained in schools with higher mean socioeconomic status

Results

Fixed effects

- Mean math achievement (i.e., for a student with a mean `cses` score in a school with a mean `meanses` score) is 12.11 in Public High Schools and 13.33 in Catholic High Schools
- Effects of socioeconomic status at the two levels
  - The effect at the student level depends on the type of school: math achievement increases by 2.94 points in Public High Schools and by $2.94 - 1.64 = 1.30$ points in Catholic High Schools for a unit increase in cses
  - Higher math achievements are obtained in schools with higher mean socioeconomic status
  - In addition, the dependence of math achievement on `cses` scores is more pronounced in schools with higher `meanses` scores (estimated interaction > 0)

Results

Random effects

- The estimate $\hat{\sigma}^2_{\upsilon_0} = 2.32$ of the variance of mean school performance provides room for improving prediction by including additional predictors

Results

Random effects

- The estimate $\hat{\sigma}^2_{\upsilon_0} = 2.32$ of the variance of mean school performance provides room for improving prediction by including additional predictors
- However, there is virtually no variation in the dependence of math achievement on cses across schools ($\hat{\sigma}^2_{\upsilon_1} = 0.07$), which should also be noted when interpreting the reported correlation of 0.48

Results

Random effects

- The estimate $\hat{\sigma}^2_{\upsilon_0} = 2.32$ of the variance of mean school performance provides room for improving prediction by including additional predictors
- However, there is virtually no variation in the dependence of math achievement on cses across schools ($\hat{\sigma}^2_{\upsilon_1} = 0.07$), which should also be noted when interpreting the reported correlation of 0.48
- The corresponding covariance has an estimated value of $\hat{\sigma}_{\upsilon_0\upsilon_1} = 0.48 \cdot \hat{\sigma}_{\upsilon_0} \cdot \hat{\sigma}_{\upsilon_1} = 0.19$

Results

Random effects

- The estimate $\hat{\sigma}^2_{\upsilon_0} = 2.32$ of the variance of mean school performance provides room for improving prediction by including additional predictors
- However, there is virtually no variation in the dependence of math achievement on `cses` across schools ($\hat{\sigma}^2_{\upsilon_1} = 0.07$), which should also be noted when interpreting the reported correlation of 0.48
- The corresponding covariance has an estimated value of $\hat{\sigma}_{\upsilon_0 \upsilon_1} = 0.48 \cdot \hat{\sigma}_{\upsilon_0} \cdot \hat{\sigma}_{\upsilon_1} = 0.19$
- These results suggest a simplified model of the dependence of math achievement on `cses`, where the intercept, but not the slope varies across schools

Summary

- Regression models with fixed and random effects

Summary

- Regression models with fixed and random effects
    - allow for adequately modeling hierarchical data structures
        - longitudinal data
        - individuals organized in groups (e.g., students in classes, or schools)

Summary

- Regression models with fixed and random effects
  - allow for adequately modeling hierarchical data structures
    - longitudinal data
    - individuals organized in groups (e.g., students in classes, or schools)
  - allow for adequately modeling the sources of error occurring in this context

Summary

- Regression models with fixed and random effects
    - allow for adequately modeling hierarchical data structures
        - longitudinal data
        - individuals organized in groups (e.g., students in classes, or schools)
    - allow for adequately modeling the sources of error occurring in this context
    - offer an optimal trade-off between individual and aggregate data analysis
        - while individual differences are modeled, information aggregated over the sample is exploited, too

Summary

- Regression models with fixed and random effects
  - allow for adequately modeling hierarchical data structures
    - longitudinal data
    - individuals organized in groups (e.g., students in classes, or schools)
  - allow for adequately modeling the sources of error occurring in this context
  - offer an optimal trade-off between individual and aggregate data analysis
    - while individual differences are modeled, information aggregated over the sample is exploited, too
- Therefore, linear mixed-effects models allow for integrating differential and general psychological aspects within a common theoretical framework

What we learned today...

. . . and how to go on

What we learned today...

...and how to go on

1. We learned

What we learned today...

...and how to go on

1. We learned
    - The basic concept of random effects and why to include them in a model

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package
   - How to use a hierarchical model to separate individual and school differences

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package
   - How to use a hierarchical model to separate individual and school differences
   - How to interpret parameters in a linear mixed-effects model

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package
   - How to use a hierarchical model to separate individual and school differences
   - How to interpret parameters in a linear mixed-effects model
2. Next steps

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package
   - How to use a hierarchical model to separate individual and school differences
   - How to interpret parameters in a linear mixed-effects model
2. Next steps
   - Do this exercise https://gitea.iwm-tuebingen.de/nwickelmaier/lead_lmm/src/branch/master/exercises/jsp.md using the JSP data set in R

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package
   - How to use a hierarchical model to separate individual and school differences
   - How to interpret parameters in a linear mixed-effects model
2. Next steps
   - Do this exercise https://gitea.iwm-tuebingen.de/nwickelmaier/lead_lmm/src/branch/master/exercises/jsp.md using the JSP data set in R
   - It has a very similar structure than the HSB data set and this will help you to generalize the concepts we learned today

What we learned today...

...and how to go on

1. We learned
   - The basic concept of random effects and why to include them in a model
   - How to compute a linear mixed-effects model in R using `lmer()` from the lme4 package
   - How to use a hierarchical model to separate individual and school differences
   - How to interpret parameters in a linear mixed-effects model
2. Next steps
   - Do this exercise https://gitea.iwm-tuebingen.de/nwickelmaier/lead_lmm/src/branch/master/exercises/jsp.md using the JSP data set in R
   - It has a very similar structure than the HSB data set and this will help you to generalize the concepts we learned today
   - You can send questions to me and even make an appointment with me to go over your solution

References

Hedeker, D. R., & Gibbons, R. D. (2006). *Longitudinal data analysis*. John Wiley.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.