|  |
| --- |
| Assignment 2 |
| Multilevel regression models with random intercepts, random slopes, and explanatory variables |
| MAE4112 Multilevel Models |
| Centre for Educational Measurement at the University of Oslo (CEMO) |
| Autumn semester |

|  |  |
| --- | --- |
| Name: |  |

# Results

|  |  |  |
| --- | --- | --- |
| Task | Credits | Max. credits |
| C1 |  | 5 |
| C2 |  | 10 |
| A1 |  | 22 |
| A2 |  | 7 |
| D |  | 19 |
| **TOTAL:** |  | **63** |

# Conceptual problems

## C1. Multilevel modeling terms and conditions

Explain the following terms and conditions.

|  |  |
| --- | --- |
| Term and conditions | Explanation |
| Cross-level interaction effect |  |
| Contextual effect |  |
| Group mean-centered level-1 predictor |  |
| Multivariate normal distribution of the level-2 random effects |  |
| Reliability of aggregated variables ICC[2] |  |

## C2. Path diagram representation of multilevel models

The following path diagram represents a multilevel model with two levels (L1 and L2):



1. Formulate the L1 and L2 model equations, including the assumptions on the distributions of the L1 and L2 residuals.

|  |
| --- |
| Model specification |
|  |

1. How many random and fixed effects are estimated in this model?

|  |  |
| --- | --- |
| Number of random effects: |  |
| Number of fixed effects: |  |

1. Formulate a research question that researchers could address with this model. Please define the levels of analysis (L1 and L2) and what the variables in the model may represent.

|  |
| --- |
| Research Question |
|  |

# Applied problems

## A1. Grade retention and mathematics achievement

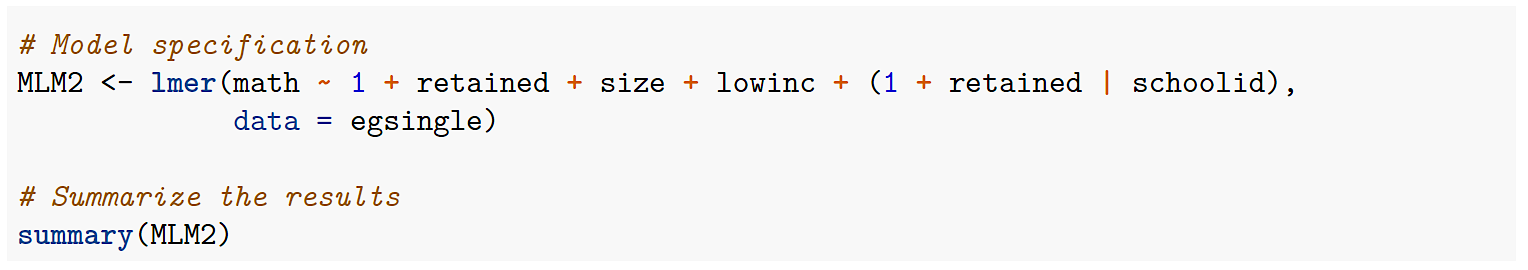
The U.S. Sustaining Effects Study examined the relation between students’ grade retention and their achievement in mathematics for large and representative samples. The data (egsingle) contain the following variables:

* schoolid: Identifier of the school
* childid: Identifier of the student
* math: Mathematics achievement score (based on an IRT scale)
* retained: Binary variable indicating whether or not the student has retained in a grade (*1 = retention*, *0 = no retention*)
* size: Number of students in the school
* lowinc: Percentage of students with a low-income background in the school

1. Determine the level at which these variables were measured.

|  |  |  |
| --- | --- | --- |
| Variable | Student level | School level |
| schoolid | 🞏 | 🞏 |
| childid | 🞏 | 🞏 |
| math | 🞏 | 🞏 |
| retained | 🞏 | 🞏 |
| size | 🞏 | 🞏 |
| lowinc | 🞏 | 🞏 |

Analyzing a subset of the data (*N* = 7230 students in 60 schools), a researcher specified the following multilevel model (MLM2) in lme4:



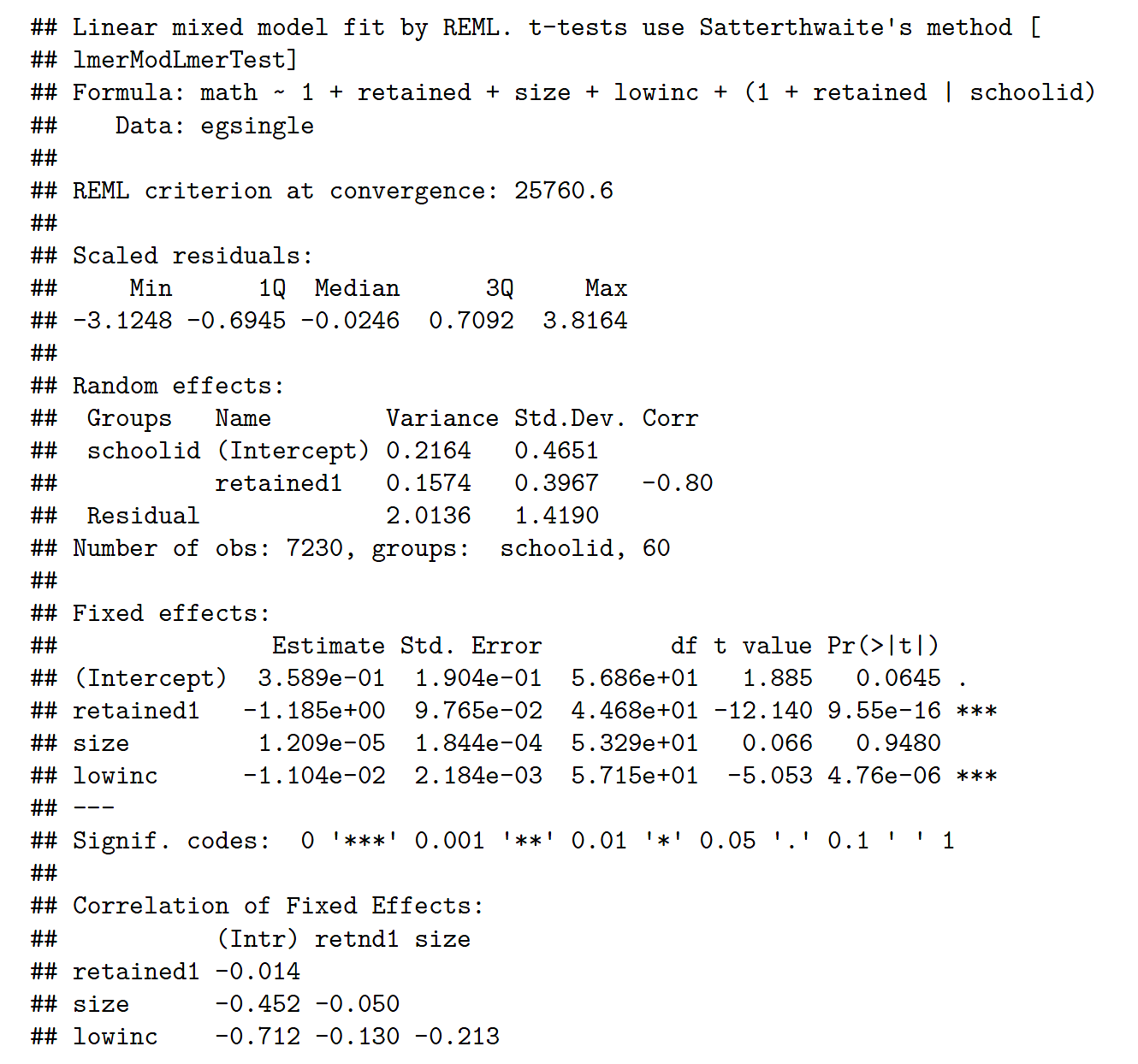
1. Formulate the L1 and L2 model equations underlying this model.

|  |
| --- |
| Model specification |
|  |

1. Draw the corresponding path diagram, including all variances, covariances, and means.

|  |
| --- |
| Path diagram |
|  |

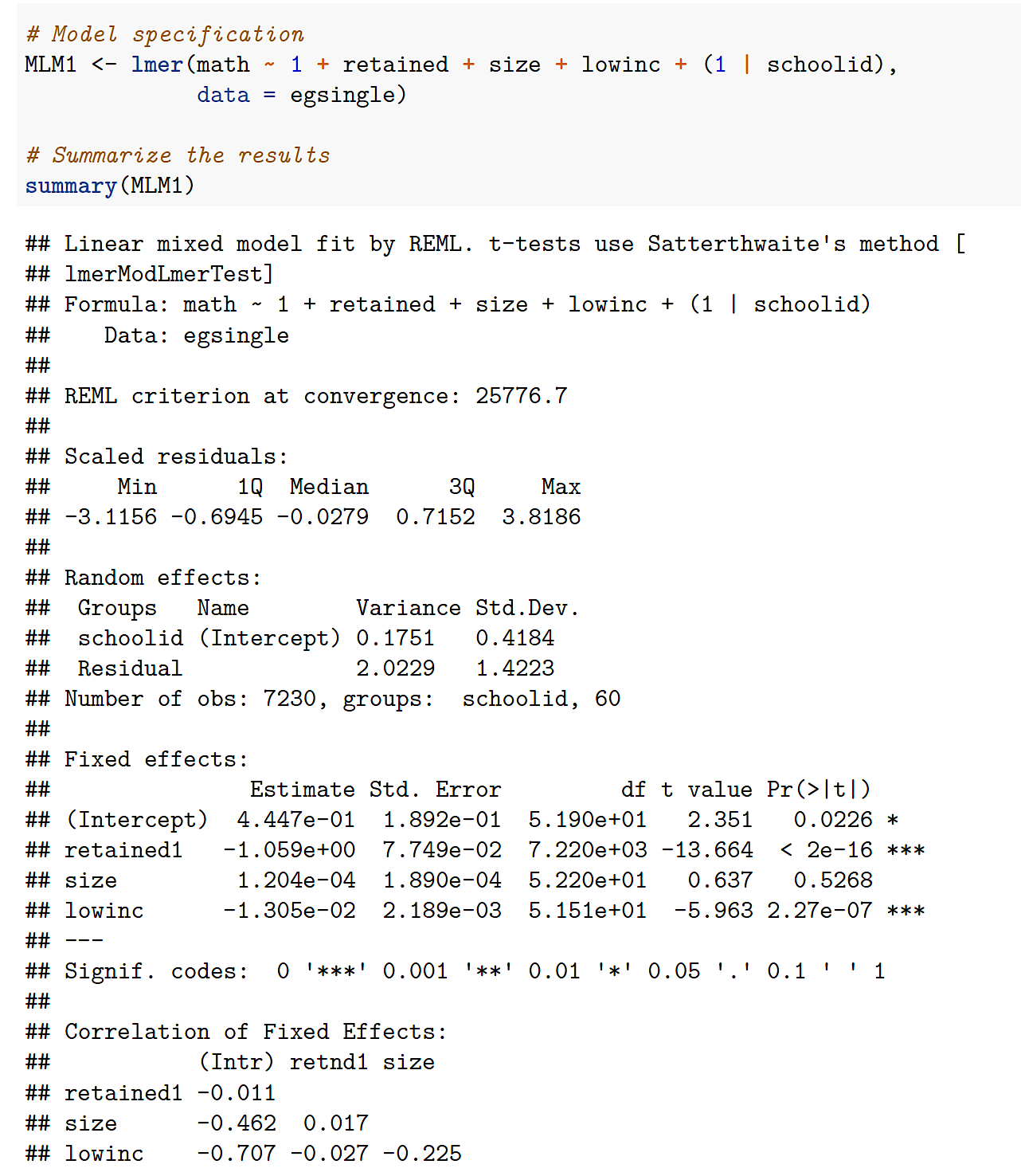
1. Estimating this model (MLM2) in lme4, the researchers obtained the following output:



Which of the following conclusions can be drawn (i.e., are true)?

|  |  |  |
| --- | --- | --- |
| Conclusion | True | False |
| 1. Students who retained in a grade scored significantly lower on the math achievement test (math) than those who did not retain, after controlling for school size (size) and the proportion of low-income students in the school (lowinc). | 🞏 | 🞏 |
| 1. Neither the proportion of low-income students (lowinc) nor the size of the school (size) were significantly related to the math achievement test score (math) between schools after controlling for retention (retained). | 🞏 | 🞏 |
| 1. School size (size) explains significant variation in the random slopes (i.e., the relation between math and retained). | 🞏 | 🞏 |

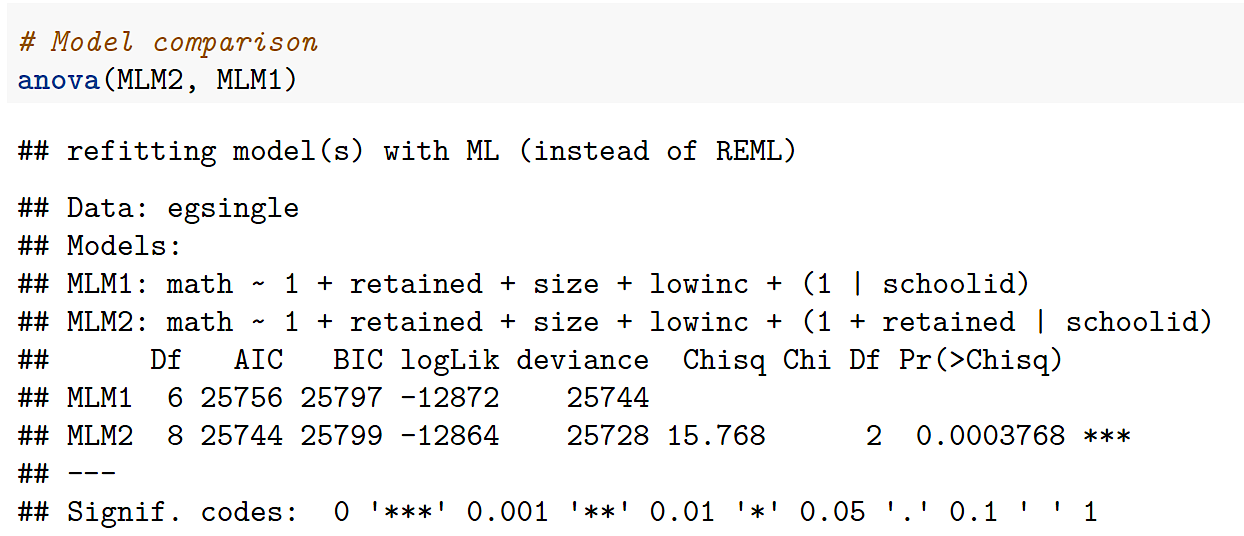
1. In a second run of the analyses, the researchers modified the model (MLM2) slightly into model MLM1 as follows:



Describe the difference between models MLM1 and MLM2 (conceptually).

|  |
| --- |
| Difference between models MLM1 and MLM2 |
|  |

1. To determine which of these two models (MLM1, MLM2) represents the data better, the researchers performed a likelihood-ratio test and obtained the following results:



Decide on which of these two models represents the data better.

|  |
| --- |
| Model comparison |
|  |

1. Comparing model MLM1 with the null model of math (Null), the researcher wanted to compute the variance explanation at level 1 (i.e., the student level). They obtained the following model outputs:

|  |  |
| --- | --- |
| Model MLM1: | Null model: |
|  |  |

Compute the level-1 variance explanations using the Raudenbush and Bryk (1992) and the Snijders and Bosker (2012) formulas. Provide not only the final result but also the calculation.

|  |  |
| --- | --- |
| Raudenbush and Bryk (1992), | Snijders and Bosker (2012), |
|  |  |

## A2. Evaluating the quality of teaching

Taut and her colleagues (2019) conducted a study of the relation between student achievement and the competence of the teachers in classrooms. The overall sample contained = 1124 Chilean students () in 48 classrooms ().

Taut et al. included three variables in their analyses:

* : Standardized test score in Mathematics and Spanish (pre-test)
* : Standardized test score in Mathematics and Spanish (post-test)
* : Teacher competence coded as 0 (*basic*) and 1 (*competent*) for each classroom

The authors estimated three multilevel regression models:

|  |
| --- |
| Model 1 |
| where and |
| Model 2 |
| where and |
| Model 3 |
| where , , , and |

1. Assign all variables in these models to either the student or the classroom level.

|  |  |  |
| --- | --- | --- |
| Variable | Student level | Classroom level |
|  | 🞏 | 🞏 |
|  | 🞏 | 🞏 |
|  | 🞏 | 🞏 |
|  | 🞏 | 🞏 |

1. Which of the following statements about these models are true?

|  |  |  |
| --- | --- | --- |
| Statement | True | False |
| 1. Model 1 assumes that the intercept of varies between classrooms. | 🞏 | 🞏 |
| 1. In all models, the relation between and is fixed (i.e., constant). | 🞏 | 🞏 |
| 1. The L2 random effects in Model 3 are assumed to be independent. | 🞏 | 🞏 |
| 1. Model 1 represents a contextual model. | 🞏 | 🞏 |
| 1. In Model 3, the coefficient represents the average classroom score. | 🞏 | 🞏 |

# Data-analytic problem

## D. High School and Beyond Study 1982

The High School and Beyond Study was conducted among *N* = 7185 students from 160 schools in 1982. The following variables were measured:

* school: Identifier of the school
* minrty: Binary variable indicating whether students belong to a minority (Yes, No)
* sx: Binary variable indicating students’ gender (Male, Female)
* ses: Index of students’ socioeconomic status
* mAch: Students’ mathematics achievement test score
* meanses: School-average socioeconomic status (i.e., the group mean of ses)
* sector: Binary variable indicating the type of school (Public, Catholic)
* cses: Group mean-centered ses variable (i.e., cses = ses – meanses)

The full data set is stored in the R package `mlmRev` under the name `Hsb82`. You may call the data using the following R code:

# Load the R package

library(mlmRev)

# Read and attach the data

data(“Hsb82”)

attach(Hsb82)

In the following tasks, make sure to use the group mean-centered SES variable (cses).

**Contextual effect of SES using the school-average SES:**

Test the hypothesis that the school-average socioeconomic status (meanses) has a contextual effect on students’ individual mathematics achievement (mAch).

1. Provide the R code of the model you have used to test this hypothesis in **lme4**:

|  |
| --- |
| Contextual effect—R code in lme4: |
|  |

1. Provide the coefficients of the fixed effects, including their standard errors, and *p*-values.

|  |  |  |  |
| --- | --- | --- | --- |
| Fixed effect | Estimate |  |  |
| Intercept |  |  |  |
| Slope of group mean-centered SES () |  |  |  |
| Slope of school-average SES () |  |  |  |

1. Specify and estimate the same model (from (a)) that tests the contextual effect of school-average SES on students’ individual mathematics test performance using the R package **lavaan**.

Provide the model specification and estimation syntax and compute the contextual effect.

|  |
| --- |
| Contextual effect using school-average SES—R code in lavaan: |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Contextual effect | Estimate |  |  |
|  |  |  |  |

**Contextual effect of SES using the latent decomposition:**

Test the same contextual model as you have in task (c) with a latent decomposition of the variable SES. Remember that this decomposition does not need to include school-average SES. Estimate the resultant contextual effect.

1. Provide the model specification and estimation syntax and compute the contextual effect **in lavaan**.

*Note. There is no need to use the centered ses variables.*

|  |
| --- |
| Contextual effect using latent decomposition—R code in lavaan: |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Contextual effect | Estimate |  |  |
|  |  |  |  |

**Cross-level interaction effect of sector:**

Test the hypothesis that the type of school (sector) moderates the relation between students’ individual mathematics achievement (mAch) and SES (cses).

1. Provide the R code of the model you have used to test this hypothesis in **lme4**:

|  |
| --- |
| Cross-level interaction effect—R code in lme4: |
|  |

1. Decide whether this model provides evidence for a cross-level interaction effect and report this effect.

|  |
| --- |
| Cross-level interaction effect—Evidence and Interpretation: |
|  |