|  |
| --- |
| Assignment 1 |
| The Null Model and the Random-Intercept Model |
| MAE4112 Multilevel Models |
| Centre for Educational Measurement at the University of Oslo (CEMO) |
| Autumn semester |

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

# Results

|  |  |  |
| --- | --- | --- |
| Task | Points | Max. points |
| C1 |  | 6 |
| C2 |  | 9 |
| C3 |  | 15 |
| A1 |  | 10 |
| A2 |  | 2 |
| A3 |  | 8 |
| D |  | 8 |
| **TOTAL:** |  | **58** |

# Conceptual problems

## C1. Key concepts of multilevel modeling (6 credits)

Explain the following concepts.

|  |  |
| --- | --- |
| Concept | Explanation |
| Intraclass correlation |  |
| Design effect |  |
| Aggregated analysis |  |
| Cross-classified multilevel structure |  |
| Fixed effect |  |
| Random effect |  |

## C2. Path diagram representation of multilevel models (9 credits)

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. Provide an overall equation.

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

1. Which of the following statements about this model are true?

|  |  |  |
| --- | --- | --- |
| Statement | True | False |
| 1. The L1 model allows researchers to test whether moderates the relation between and . | 🞏 | 🞏 |
| 1. This model quantifies the variation of the variable across clusters . | 🞏 | 🞏 |
| 1. The relations between the predictors and the outcome variable are assumed to be fixed across clusters at L2. | 🞏 | 🞏 |
| 1. This model includes , , , and as random effects. | 🞏 | 🞏 |
| 1. This model represents a random-intercept model. | 🞏 | 🞏 |
| 1. The relation between and can be interpreted as follows: After controlling for and , the coefficient indicates how changes in one unit of result in changes of . | 🞏 | 🞏 |

## C3. Interpreting the intraclass correlation ICC[1] (15 credits)

The intraclass correlation represents an indicator of homogeneity within clusters and is based on the null (empty) model.

1. Formulate the two-level null model for a given variable of persons in clusters , including its assumptions on the distributions of the L1 and L2 residuals.

|  |
| --- |
| Null Model |
|  |

1. The intraclass correlation can be interpreted as the correlation between the scores of two randomly drawn but different persons and within the same cluster .

In other words, for persons .

Using the null model, its assumptions, and the rules for (co-)variances, prove that this is the case by taking the following steps:

1. Show that .
2. Show that for persons for .
3. Using (1) and (2), show that for persons for .

|  |  |  |
| --- | --- | --- |
| Step 1: | | |
|  | | |
|  |  |  |

|  |
| --- |
| Step 2: for |
|  |

|  |
| --- |
| Step 3: for |
|  |

# Applied problems

## A1. Language skills and intelligence (10 credits)

In a large-scale study of eight-grade students in 211 schools, students’ scores on a language test and on a verbal intelligence test () were reported. Examining the extent to which intelligence explains variation in language skills, a team of researchers specified and estimated a series of models. The corresponding model parameters are shown in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 |
|  | **Null model** | **Single-level regression** | **Random-intercept model** |
| Parameter estimate | Coefficient () | Coefficient () | Coefficient () |
| Intercept () | 41.00 (0.32) | 41.30 (0.12) | 41.06 (0.24) |
| Regression coefficient of () |  | 2.65 (0.06) | 2.51 (0.05) |
| L2 variance () | 18.12 (2.16) |  | 9.85 (1.21) |
| L1 variance () | 62.85 (1.49) | 49.80 (1.15) | 40.47 (0.96) |
| Deviance () | 26595.3 | 25351.0 | 24912.2 |
| Number of parameters () | 3 | 3 | 4 |

1. Using the information in the table, calculate the intraclass correlation of the language skills scores and decide whether multilevel modeling is needed.

|  |  |
| --- | --- |
|  |  |
| Decision |  |

1. Interpret the regression coefficient of () in Models 2 and 3.

|  |
| --- |
| Interpretation |
|  |

1. Calculate the Akaike’s Information Criterion (AIC) and the Bayesian Information Criterion (BIC) from the deviances for Models 1-3.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 |
|  | **Null model** | **Single-level regression** | **Random-intercept model** |
| AIC |  |  |  |
| BIC |  |  |  |

1. Calculate the L1 variance explanation according to Hox in Model 3.

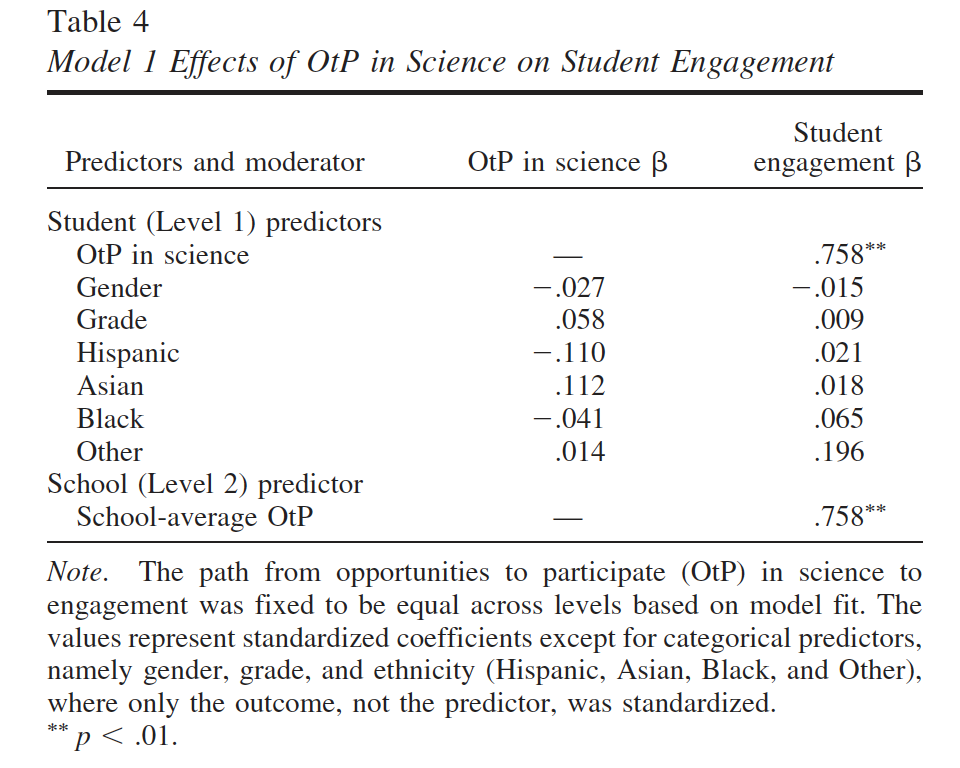
|  |  |
| --- | --- |
|  |  |

1. Argue why the extent to which the IQ score explains variation in the language skills score is best described by Model 3.

|  |
| --- |
| Argumentation |
|  |

## A2. Opportunities to participate in science learning and engagement (2 credits)

In their recently published paper, Bae and Lai (2020) examined which personal (i.e., student-level) and which contextual (i.e., school-level) variables explain variation in the opportunities the students were given to participate in science learning (*OtP in science*) and their engagement in science (*Student engagement*). The authors presented some of their results as follows:



*Note.* Bae & Lai (2020), DOI: 10.1037/edu0000410, p. 1139.

Describe two results of their study on the basis of this table.

|  |
| --- |
| Results |
|  |

## A3. Random-intercept models with an L2 explanatory variable (8 credits)

In a typical random-intercept model, the outcome variable varies across the L2 clusters. In other words, its variance between clusters is quantified at L2. Besides “only quantifying” this variance, researchers may also introduce variables at L2 that explain this variation (so-called “L2 explanatory variables”).

A researcher is interested in the relations between students’ socioeconomic status (), their perceptions of school climate (), and science achievement (). He or she specifies a random-intercept model with and as level-1 predictors and school size () as an L2 explanatory variable that explains at the school level.

The model is specified as follows:

|  |
| --- |
| Model specification |
| *Level 1 (students):*  with  *Level 2 (schools):*  with |

Draw the corresponding path diagram.

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

# D. Data-analytic problem

## The Tennessee Student Teacher Achievement Ratio (STAR) project (8 credits)

In the following, we will analyze the star data stored in the R package mlmRev. The data are comprised of 26796 students from 80 schools in Tennessee, USA. Two variables describe students’ educational achievement: the performance score of a math test (math) and the performance score of a reading test (read). The dataset contains the following variables that are of interest for the subsequent analyses:

* Student ID (id)
* School ID (sch)
* Total score on the mathematics test (math)
* Total score on the reading test (read)
* Student socioeconomic status (ses) coded as *F = eligible for free lunches (low SES)* and *N = Not eligible for free lunches (high SES)*

**Our main question is:**

To what extent does performance on the reading test predict students’ performance on the mathematics test after controlling for socioeconomic status?

1. Install the R package mlmRev and load the dataset.

*Note.* You may use the command data(star, package = "mlmRev") and then attach the dataset. In all models, use the REML estimation.

1. Prepare the data set before the analyses as follows:

* *Handling missing values:* The variables math, read, and ses have missing values. For the ease of use, remove all cases, which show missing data in these three variables. You have successfully removed these cases if your new, reduced data set contains 23588 cases.
* *Recoding the variable* ses*:* Create a new, numeric variable that is coded as 1 for eligibility for free lunches (former “F”) and 0 for non-eligibility for free lunches (former “N”).

1. **Model 0:** Specify and estimate the two-level null model for math. Calculate the intraclass correlation ICC[1].

|  |
| --- |
| R code for the null model |
|  |

|  |
| --- |
| Intraclass correlation, |
|  |

1. **Model 1:** Specify and estimate a two-level random-intercept model with math as the outcome variable and read as the predictor variable.

|  |
| --- |
| R code for the random-intercept model |
|  |

1. **Model 2:** Specify and estimate a two-level random-intercept model with math as the outcome variable and read as the predictor variable, controlling for ses.

|  |
| --- |
| R code for the random-intercept model |
|  |

1. Complete the following table.

*Note.* You may choose any form of L1 variance explanation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 0 | Model 1 | Model 2 |
| Parameter estimate | Coefficient () | Coefficient () | Coefficient () |
| Intercept () |  |  |  |
| Regression coefficient of () | - |  |  |
| Regression coefficient of () | - | - |  |
| Variance component | Coefficient | Coefficient | Coefficient |
| L2 variance () |  |  |  |
| L1 variance () |  |  |  |
| Deviance () |  |  |  |
| Number of parameters () |  |  |  |
| AIC |  |  |  |
| BIC |  |  |  |
|  | - |  |  |

1. Respond to the main question.

|  |
| --- |
| Your response |
|  |

# Notes