Assignment 2 is due no later than 5pm Friday, 15th of May, 2020. In submitting your work, you are consenting that it may be copied and transmitted by the University for the detection of plagiarism. Submission is your guarantee that the below statement of originality is correct.

"This is my own work. I have not copied any of it from anyone else."

NAME: **Replace this text with your name.**
STUDENT NUMBER: **Replace this text with your student number.**

Instructions for assignment: You will need to submit two documents for this assignment. The first document is a pdf document named `Assign2_StNo.pdf` which will provide all your analysis and solutions for this assignment. To produce this pdf document you will need to use LaTeX. The LaTeX document which was used to produce this assignment is named `Assign2_StNo.tex` and is located in the Assignment 2 folder in the Topic 7 section of LMS. You can use LaTeX online via Overleaf which is a website dedicated to producing documents from LaTeX. To use LaTeX, follow the instructions in the `Overleaf.pdf` document located in the Assignment 2 folder. The second document that you will need to submit is an R document named `Assign2_R_StNo.R` which is located in the Assignment 2 folder. This document should provide the R code you used to perform all your data manipulation and analysis.

Assessment information for assignment: There are a total of 100 marks for this assignment.

Description of assignment: The data and information presented in this assignment is adapted from Anderson et al (2009)[1] and is stored in the file named `Classroom.csv` located in the Assignment 2 folder in the Topic 7 section of LMS. Two hundred and eighty five first-grade classrooms were selected at random and a number of variables were measured on the teacher and some students in the classroom. The variables of interest for Assignment 2 are:

- *Classid*: This is a factor variable that identifies the classroom.
- *Chilid*: This is a factor variable that identifies the child.
- MG: This is a continuous variable that measures the child's gain in mathematics score from the spring of kindergarten to the spring of the first grade.
- YE: This is a continuous variable that measures the number of years of teaching experience of the teacher.
- MK: This is a continuous variable that measures the teacher's mathematical knowledge (higher values indicate greater knowledge).
- MP: This is a continuous variable that measures the teacher's mathematics preparation.
- HP: This is a continuous variable that measures the percentage of households that are below the poverty level in the neighbourhood of the classroom.

2 marks are allocated for each question that requires the use of the R computer package. These marks are awarded using the following criterion:

1. **R code that accurately produces the analysis/output required in the question.**

Answer the following questions.

[1] Anderson, D., Oti, R., Lord, C., and Welch, K. (2009). Patterns of growth in adaptive social abilities among children with autism spectrum disorders. *Journal of Abnormal Child Psychology*, **37(7)**: 1019–1034.

1 Describing the model

The researchers in the study set up the following linear mixed model to analyze their research questions.

$$\begin{aligned} MG_{ij} &= \beta_0 + \beta_1\, YE_j + \beta_2\, MK_j + \beta_3\, HP_j \\ &\quad + \beta_4\, YE_j \times MK_j + \beta_5\, YE_j \times HP_j + \beta_6\, MK_j \times HP_j \\ &\quad + \mu_{0j} + \varepsilon_{ij}, \end{aligned} \tag{1}$$

- where MG_{ij} is the gain in mathematics score for child i ($i = 1, \ldots, n_j$) in classroom j ($j = 1, \ldots, 285$),
- YE_j is the number of years of experience of the teacher in classroom j.
- MK_j is the mathematical knowledge score of the teacher in classroom j.
- HP_j is the percentage of households that are below the poverty level in the neighbourhood of classroom j.
- β_0 is the fixed intercept,
- β_1, β_2 and β_3 are the fixed simple effects of YE, MK and HP respectively,
- β_4, β_5 and β_6 are the fixed two-way interaction effects of $YE \times MK$, $YE \times HP$ and $MK \times HP$ respectively,
- μ_{0j} is the random intercept specific to classroom j,
- ε_{ij} is the random error associated with measuring MG for child i, in classroom j.

1. The researchers would like to express model (1) in matrix form, $\boldsymbol{Y}_j = \boldsymbol{X}_j\,\boldsymbol{\beta} + \boldsymbol{Z}_j\,\boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_j$, where $\boldsymbol{Y}_j$ represents the response vector for classroom j, $\boldsymbol{X}_j$ represents a matrix, for classroom j, that contains the values of the predictors associated with the fixed effects of model (1), $\boldsymbol{\beta}$ is the fixed effect vector, $\boldsymbol{Z}_j$ is a matrix, for classroom j, that contains the values of the predictors associated with the random effects of model (1), $\boldsymbol{\mu}_j$ is the random effect vector for classroom j and $\boldsymbol{\varepsilon}_j$ is the random error vector for classroom j. Answer the following questions.
   (a) Write down the response vector, $\boldsymbol{Y}_j$, of model (1), for classroom $j = 3$ ($Classid = 211$). **(2 marks)**
   (b) Write down the matrix, $\boldsymbol{X}_j$, of model (1), for classroom $j = 3$ ($Classid = 211$). **(7 marks)**
   (c) Write down the fixed effect vector, $\boldsymbol{\beta}$, of model (1). **(1 mark)**
   (d) Write down the matrix, $\boldsymbol{Z}_j$, of model (1), for classroom $j = 3$ ($Classid = 211$). **(3 marks)**
   (e) Write down the random effect vector, $\boldsymbol{\mu}_j$, of model (1), for classroom $j = 3$ ($Classid = 211$). **(2 marks)**
   (f) Write down the random error vector, $\boldsymbol{\varepsilon}_j$, of model (1), for classroom $j = 3$ ($Classid = 211$). **(2 marks)**
2. Interpret the intercept β_0. **(3 marks)**
3. Interpret the simple effects β_1, β_2 and β_3. **(9 marks)**
4. Interpret the interaction effects β_4, β_5 and β_6. **(9 marks)**
5. For model (1), the researchers choose an *unstructured* structure for the variance-covariance matrix of the random effect vector, $\boldsymbol{\mu}_j$. That is, the variance-covariance matrix of the random effect vector, $\boldsymbol{\mu}_j$, is

$$\boldsymbol{D} = \left[\ \theta\ \right],$$

   - where θ denotes the variance of the random intercept μ_{0j}.

Also for model (1), the researchers choose the diagonal structure for the variance-covariance matrix of the random error vector, $\boldsymbol{\varepsilon}_j$. That is, for classroom $j = 3$ ($Classid = 211$) the variance-covariance matrix of the random error vector, $\boldsymbol{\varepsilon}_j$, is

$$\boldsymbol{R}_3 = \begin{bmatrix} \tau & 0 & 0 & 0 \\ 0 & \tau & 0 & 0 \\ 0 & 0 & \tau & 0 \\ 0 & 0 & 0 & \tau \end{bmatrix}$$

- where τ denotes the constant variance of the random errors.

For model (1), derive the variance-covariance matrix of the response vector, $\boldsymbol{Y}_j$, for classroom $j = 3$. (Use the same notation as above and show all workings) **(6 marks)**. What is the structure of the variance-covariance matrix of the response vector? **(1 mark)**

2 Testing for the random intercept

6. The researchers would like to test whether the random intercept, μ_{0j}, should be included in model (1). They decide to test, at the 5% significance level, the null hypothesis $H_0 : \theta = 0$ vs the alternative hypothesis $H_1 : \theta > 0$ using the REML-based likelihood ratio test p-value.
   (a) Write down the reference model for this test. **(1 mark)**
   (b) Write down the nested model for this test. **(1 mark)**
   (c) Use the R computer package to perform this test. What is the p-value for this test? **(1 mark)**
   (d) Which model would you choose (reference or nested) to continue your analysis? Explain. **(2 marks)**

3 Variance-covariance estimates of the final linear mixed model

As their final linear mixed model the researchers choose model (1) which has variance-covariance matrices, $\boldsymbol{D}$ and $\boldsymbol{R}_j$, defined in section 1. Use this model to answer the questions in this section and in section 4.

7. Use the R computer package to calculate the estimate of the $\boldsymbol{D}$ matrix of the final linear mixed model. Present this estimate below. **(3 marks)**

8. Use the R computer package to calculate the estimate of the $\boldsymbol{R}_3$ matrix of the final linear mixed model. Present this estimate below. **(3 marks)**

9. Use the R computer package to calculate the estimate of the variance-covariance matrix of the response vector of the final linear mixed model, for classroom $j = 3$. Present this estimate below. **(3 marks)**

10. Use the information in questions 7 and 8 to calculate, by hand, the intraclass correlation coefficient (**3 marks**). Interpret this intraclass correlation coefficient (**2 marks**).

4 Predicted values and residuals of the final linear mixed model

11. Use the R computer package to produce a table that lists the estimates of the fixed effects in the final linear mixed model, together with their corresponding standard errors, degrees of freedom, observed test statistics and p-values. Present this table below. **(2 marks)**

12. Figure 1 presents the random intercept predictions of the final linear mixed model, for the first 10 classrooms in the study. These predictions were obtained by using the `ranef()` command in R. Calculate

Figure 1: Random Intercept Predictions

```
(Intercept)
    -4.4535
     3.2520
   -16.1744
    18.2063
    20.8763
    -8.9031
   -15.4906
   -24.7702
    14.3620
    11.1952
```

by hand the predicted conditional value of MG for classroom $j = 6$. Show all your workings. (**Note:** To answer this question, you will need to use the appropriate information presented in Figure 1, the fixed effect estimates you computed in question 11 and the raw data presented in the `Classroom.csv` file). **(6 marks)**

13. Calculate by hand the conditional residual for child $i = 5$ in classroom $j = 1$. Show all your workings. (**Note:** To answer this question, you will need to use the appropriate information presented in Figure 1, the fixed effect estimates you computed in question 11 and the raw data presented in the `Classroom.csv` file). **(6 marks)**

14. Calculate by hand the marginal residual for child $i = 2$ in classroom $j = 15$. Show all your workings. (**Note:** To answer this question, you will need to use the appropriate fixed effect estimates you computed in question 11 and the raw data presented in the `Classroom.csv` file). **(6 marks)**

15. Calculate by hand the predicted marginal value for children whose first-grade classroom teacher has 14 years of experience with a mathematical knowledge score of 1.4, and where 20% of households are below the poverty level in the neighbourhood of the children's classroom. Show all your workings. (**Note:** To answer this question, you will need to use the appropriate fixed effect estimates you computed in question 11). **(6 marks)**