Model card:

1. Model details:

- Jia Jia Ji developed this linear mixed model in April 26, 2021.
- This model is the first version.
- This is a linear mixed model that includes both the fixed effect and random effect and uses a continuous variable as the response variable.
- The future users can contact the developer by email: jiajia.ji@mail.utoronto.ca.

2. Intended Use:

- This model was developed for a specific task: to analyze the effect of DLL program (a one-on-one literacy program) on the students' Spanish literacy skills measured by the posttest scores of Logramos literacy test. It was developed for performing a statistical analysis on the collected data.
- This model intended for the users that were interested in analyzing this research question or reproducing this project analysis.
- This model could be used to analyze the effect of a similar literacy program on the outcome (usually test scores) based on the randomized controlled trial where a control group and a treatment group (that received the treatment) are generated, so the test score was treated as the response variable and the random assignment was treated as the main fixed effect. Also, there should be repeated measurements in the data, so the random effect should be included.

3. Factors:

- Since there was a randomized controlled trial, this model involved 2 groups that received different forms of education. The participants were randomly assigned to the control group that received the regular education and treatment group that enrolled in the DLL program. So, these 2 groups (or the variable representing the random assignment: T_assignment) were used to identify the respective students.
- The data contained the school information, test scores and group assignment result of 152 students, the input to the model (i.e. the data) was manually recorded.
- This model worked well with the data collected by the randomized controlled trial where students were randomly assigned to groups and students were sampled from many schools (so repeated measurements), and may not work well for the observational data.
- The predicted posttest score for a student that was in a specific group and had a specific pretest score (also adjusted for the school differences) was reported.

4. Metrics:

- The estimated coefficients (i.e. MLEs), p-values, estimated marginal means and 95% confidence interval were used as the metrics to check whether the effect of an explanatory variable (i.e. DLL program) was statistically significant on the posttest score.

- I checked whether the estimates were positive or negative, whether the p-value was less than 0.05 (a threshold for p-value, I reject the null hypothesis (i.e. H0: no effect of DLL on the outcome) if p-value was less than 0.05, and I fail to reject the null hypothesis if p-values was more than 0.05), to see whether this variable was positively/negatively affected the outcome.
- Also, I checked the estimated marginal means and 95% confidence intervals for 2 groups to compare their average predicted value and further concluded that whether the students in treatment group had higher posttest scores than those in control group on average.

5. Evaluation Data:

- Since this was a statistical model rather than a machine learning model. I used the whole dataset to fit the model, and I did not split the data into training data and test data to evaluate the fitted model.

6. Training Data:

- Since this was a statistical model rather than a machine learning model. I used the whole dataset to fit the model. I included the variable representing the random assignment T_assignment and the pretest score into the model as the fixed effects, and also included the school_id in the model as the random effect. I predicted each posttest score based on the group and the corresponding pretest score.

7. Quantitative Analyses:

- For each model, the estimate for T_assignment was positive and the p-values for this variable was less than 0.05 (except that Language model had a p-value slightly larger than 0.05, but based on the context, I could still interpret the DLL positively affected the posttest Language score), which revealed that DLL statistically significantly affected the posttest score positively.
- Also, the estimated marginal means for treatment group was higher than that for control group, and the confidence interval for treatment group had a higher upper bound and a higher lower bound than that for control group, which revealed that the students in treatment group had higher posttest scores on average than those in control group. So, DLL could improve the Spanish literacy skills (that was measured by posttest scores).

8. Ethical Considerations:

- The dataset did not contain the demographical information and each student was identified using an id number, so no sensitive data in the data.