Assignment 2 part 2

The questions to be answered (in a separate document) are:

**1- Discuss the differences in performance of your model in training and testing data**

The RMSE in the test data is higher meaning they are worse performing than the training data.

This means there is a bigger difference between the predicted values and the actual values in the testing data. As the RMSE is twice as big in the test data, there might be an indication that we slightly overfitted the model.

**2- Which individual differences should be included in a model that maximizes your ability to explain/predict new data?**

We used the following model as the basic model:

Mean Length of Utterances for the child ~ Visit \* Diagnosis + (1 + Child ID | Visit).

We added Verbal IQ to the model, which gave us a better model than the initial:

Mean Length of Utterances for the child ~ Visit \* Diagnosis \* VerbalIQ1 (1 + Child ID | Visit).

We measured the performances of the two models:

Mean RMSE of the basic model 0.75

Mean RMSE of the model with verbal IQ 0.56

We did find that the model had a little better RMSE when adding Ethnicity and Mother MLU, however we took into consideration that Ethnicity did not seem like a good factor, due to the fact that there might be one person from Brazil, which is rather specific, but a lot of participants who were described as “white” which is very general. Mother MLU might be correlated with child MLU.

**3- Predict a new kid's performance (Bernie) and discuss it against the expected performance of the two groups**

We found that Bernie is 0.47 below the predicted language performance measured in mean length of utterances of a typically developed child, and he is 0.17 above the predicted performance of a child with an autism spectrum disorder.