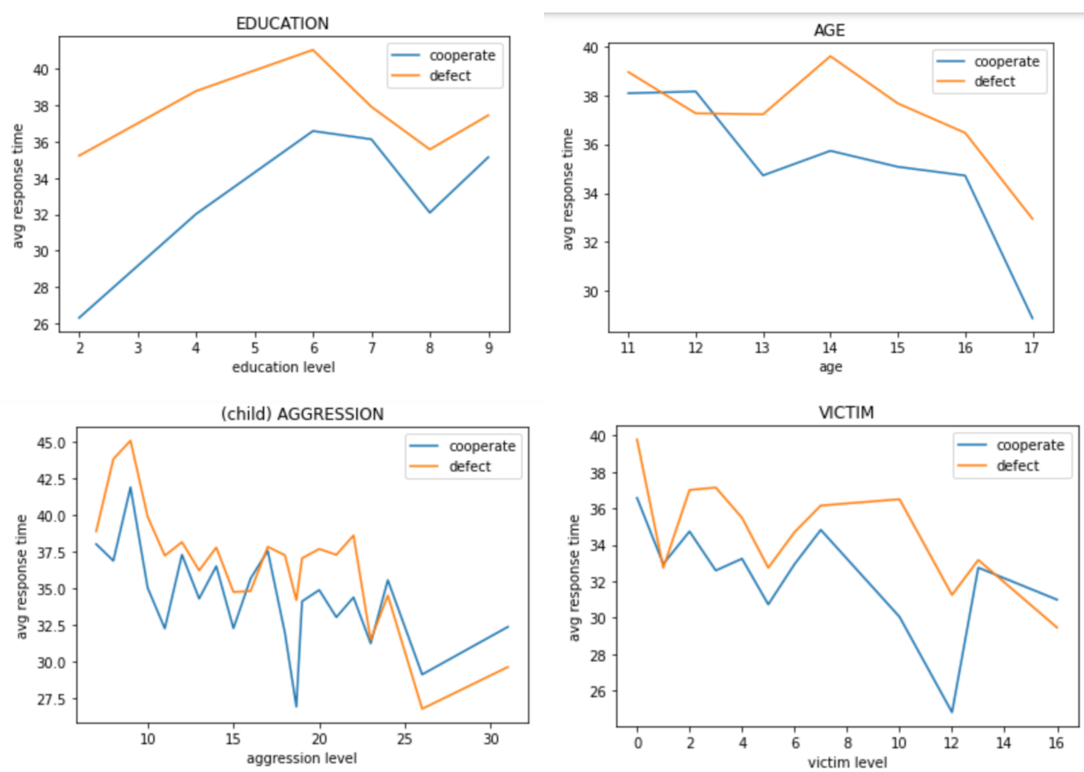
Drift Diffusion Models of children's interactions in a Repeated Prisoner's Dilemma Game (Team 2)

Dakota Clark-Singh
Audrey Tucker
Ryan Shnitman

Deliverable 2

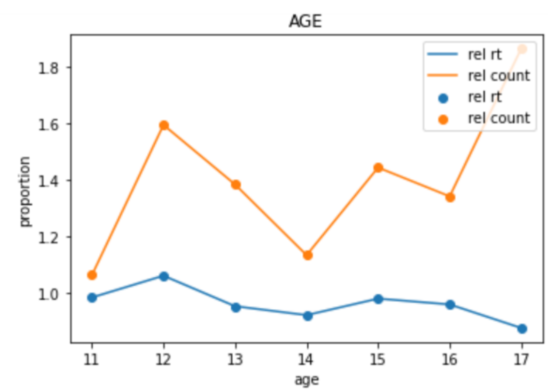
1. Refinement of Initial Analysis from PD1:

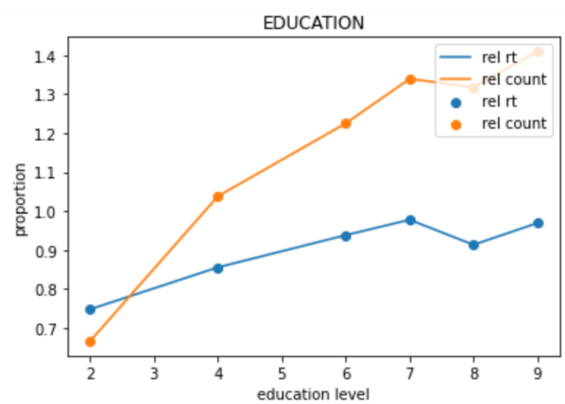
While our initial analysis focused on predicting cooperation percentage from aggression scores, the results did not show clear patterns, and suggested that we should consider a new response variable. We decided to analyze the way reaction times for defection and cooperation decisions differed across various characteristics. We plotted a few interesting patterns for average response times for defection and cooperation at each level of a characteristic below:

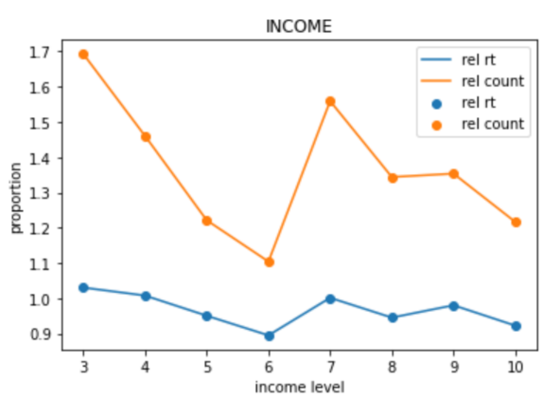


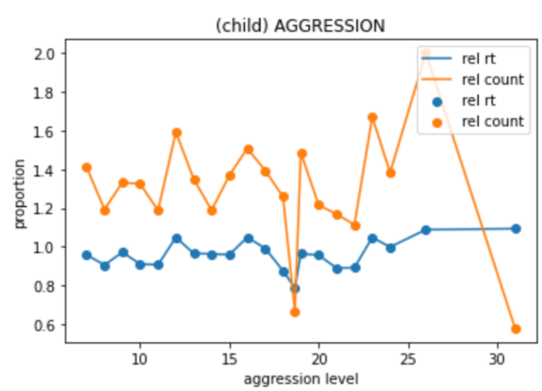
We notice that as education level of the parent rose, avg rt for both cooperation and defection rose, and the large gap between defection and cooperation at the lower levels closed a good amount at the higher level. As age increased, avg response times for both decreased. For self reported aggression composite scores, we also observe a downward trend, with defection time above cooperation time for most levels except for the highest scores, which suggests that these hyper aggressive children are taking less time to decide to defect relative to cooperation. For victimization we notice a similar pattern to aggression but with a higher gap separating defection and cooperation times.

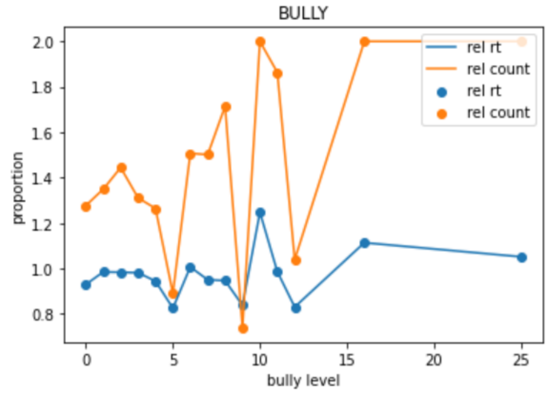
A natural problem with these plots is that a child may take a long time to decide to both cooperate and defect when really we want to understand the difference between how long it takes them to defect and cooperate. Our solution is to analyze relative reaction times (avg cooperation time / avg defection time), and to compare against relative decision counts (# of cooperation decisions / # of defection decisions), hence combining insight from the analysis we did above with the analysis we did in the first deliverable. Since these plots contain a lot of information, we've decided to include several, but only note a few interesting patterns for the sake of concision. Interpreting the proportion: a relative rt > 1 means it took more time to decide to cooperate than time to defect, a relative count > 1 means there were more cooperation decisions relative to defection decisions.

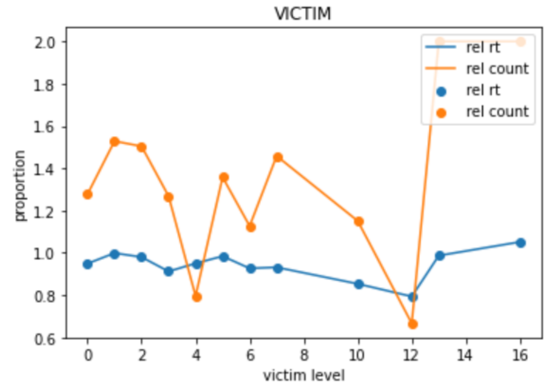


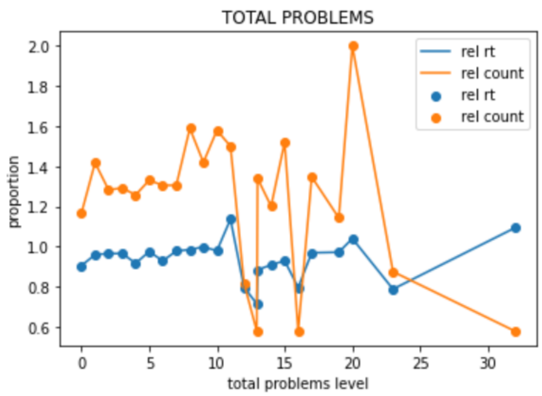






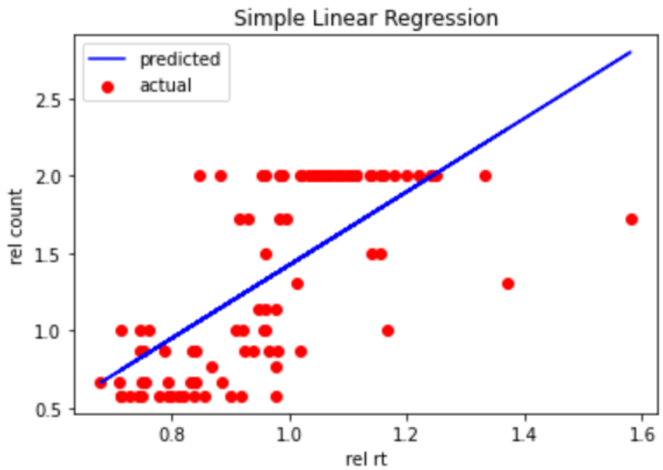






The relative reaction time is relatively stable except in the case of education, where cooperation time clearly rises above defection time as education level increases (as well as the cooperation count). Counts appear to be quite erratic in other cases. In aggression, we notice a slight upward trend in relative rt at the higher aggression levels which coincides with a decrease in relative count; at the higher aggression levels, it takes children longer to decide to cooperate, and they cooperate less. We notice a similar “flip” pattern in the last level of total problems.

We now analyze the relationship between cooperation/defection counts and response times more concretely, in order to decide whether the client’s desired model, a drift diffusion model (DDM), would be appropriate. A core assumption of this model is that a faster response time indicates a true preference; in other words, if the child chooses to cooperate faster on average than defect, their “true” preference is cooperation. To test this, we look at a simple linear regression model explaining variation in relative decision count by relative reaction time. We first split the twins up into separate sets, fit the model on the first set, and test on the second. Below we can see the best fit line (i.e. predicted relative counts) determined by the regression in blue, and the actual test twin relative counts as the red scatter.



To determine if our linear model is a good representation of the data, we expect the data to vary with similar randomness across the length of our best fit line. It appears that the variation may be heavier below the line, but it does not appear there would be a better fit with a quadratic or logistic linear model. We do also note that the variance seems to increase in the higher rt's (heteroskedasticity), which may be due to outliers, and there are two modes at the lower counts and the highest counts, so we will keep this in mind. Furthermore, there does appear to be a linear relationship here, but it's positive rather than negative, meaning that as it took twins longer to decide to cooperate rather than defect, they actually cooperated more times than defected. This observation unfortunately implies that a DDM would not make accurate predictions about preferences, and can be explained by a few things: it's possible that to the contrary, when the kids spend more time thinking about what to do, they end up making the "correct" choice, in a sense, of cooperation. This is made more likely because the DDM is supposed to be applied in scenarios where the participants make quick decisions under a relatively tight time constraint, which was not the case for the twins experiment. The added complication of predicting partner behavior and not just acting on one's own true preference presents another challenge. Moreover, we conclude that a regression model is more appropriate at this time.

Investigating one of our key questions:

Question:

Can a child's decision to cooperate or defect be predicted using their reported aggression as well as their previous decisions in the game?

- **Progress made so far and current inferences:**

We have started to investigate this relationship, however so far we have not found a clear correlation between reported aggression and decision making. However (as explained below) there are still several paths we plan on taking towards answering this question. We have not yet used their previous decisions in the game as a parameter, and we plan on using this along with aggression to try and predict cooperation or defection. So at the moment there does not seem to be a clear way to predict decision making, although we are hopeful that through more analysis and a different approach we will be able to do this.

- ### Our plans for getting closer to answering this question:

- **Look at change in state, using a LSTM neural network:**

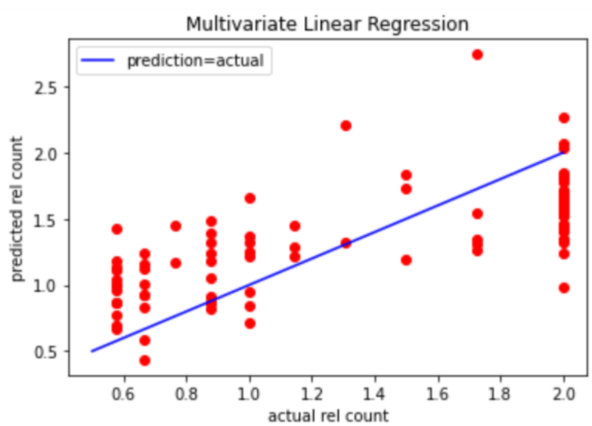
We want to explore more deeply the concept of a child's "state", or general inclination to cooperate or defect. This state can change over the course of the trials as a result of the partner's behavior and the child's aggressive traits. A change of state from cooperation to defection can take the form of repeated defections regardless of the partner's decisions, and shorter reaction time as the child defects without thinking or reasoning as much. This behaviour signals that a child has switched into a state where they automatically defect.

To model this, we will use an LSTM (Long short-term memory) neural network. An LSTM uses feedback connections to process sequences of data. At a high level, we hope this can simulate a child's mind by making decisions based on previous trials with more weight given to the trials immediately preceding the current trial. Using the reaction time and aggressiveness of each child as features as well as the sequence of decisions, the neural network could potentially learn how to mimic a change of state.

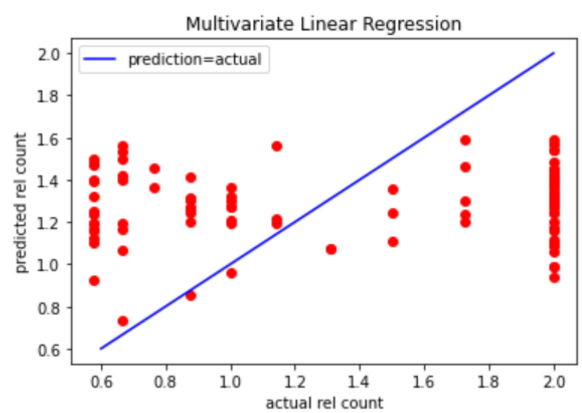
However, one of the biggest challenges facing this model is the lack of training data. Since we only have data from 215 children, with 30 trials for each, our model may suffer from overfitting and learn only how to make predictions for these experiments rather than generalizing the thought process behind the data.

- **Change the model we are using to predict cooperation:**

As we will discuss below, there are several other models we could try, some of which may better fit our data (for example we were thinking of using a logistic regression instead of a linear regression). We also plan on creating a correlation matrix to look at the relationship between variables, and include heavily correlated variables in our regression equation. To evaluate the accuracy of our model in predicting decisions, we are planning on using the root square mean error between the predicted values and the actual values, as well as looking at the plot (like the one below) showing how close our data points are to the line where predicted values equals actual values (ideally our points would be close to this line). Of course to avoid overfitting our data, we should not see our data exactly on this line. Points lying above the line are overpredicted and points lying below are underpredicted.



predictors: rel rt, age, education, income, aggression



predictors: rel rt, age, education, income, aggression, total problems, bully, victim

Looking at the regression on the right, it appears that we have a relatively decent fit, which is likely heavily reliant on the relative reaction time predictor, as the plot looks quite similar to the simple linear regression we performed earlier. The plot on the right, however, seems to be over predicting the lower rel counts and under predicting the high rel counts, perhaps implying that a quadratic model would be a better fit. Likely the inclusion of extra predictors is causing some overfitting as well. We based our variable selection off intuition and the plots we made in the first section, but going forward we would like to employ some more sophisticated techniques like observing a scatter plot matrix or using a lasso regression.

Limitations and potential risks of reaching project goal:

- **Drift Diffusion model does not match data**

  The wait times between decisions in our data does not match the wait times needed to apply the drift diffusion model. Because of this, we concluded that it would be best to move away from the drift diffusion model (our initial model for explaining the data).

- **Small sample size**

One limitation we face is that we do not have many data samples for training a more robust model. This could cause overfitting as we do not have enough data points to expose a model to a super diverse set of data. This may also contribute to our current lack of a clear understanding of the relationship between variables (as explained below).

- **Lack of a clear relationship between variables**

When plotting the general relationship between percentage of cooperation against different parameters such as age, behavioral problems, and aggression, we did not observe clear trends in the data. We are hoping to find some relationship between the child's decision making in the game (as well as the time it takes them to make these decisions), and their level of aggression or other reported behaviors. A potential limitation to us finding this correlation, is that our data does not indicate a clear correlation between these variables. However, we are hopeful that some correlation exists, as there are several more ways in which we can split up our data (rather than just looking at cooperation percentage in general). For example, we have not yet differentiated between defection after their partner defects (which may be thought of as retaliatory or impulsive), and defection after their partner cooperates (which may be thought of as more premeditated or strategic). We can also look at the change in a child's trend - so for example if they cooperate for several rounds and then suddenly defect, or vice versa. It is possible that trends in the data do exist, we just need to look at it a bit differently to more clearly see these trends.