Manipulating Data and Run Regression

Qinghua Feng

March 2023

1 **To manipulate the data using count, summarize, group by, etc. Specifically, I use distinct function to remove any duplicate rows from the data frame; Then I use group by function to group the data by the ID variable; Summarize function is used to calculate the number of unique treat variables within the group; After this, to un group the data frame; Finally, to count the number of groups in the resulting data that has unique value.**

2 **Using lm function to estimate regression.**

3 **To plot the data. The question we might be interested in is whether attending the science camp improves test performance. Visualization could help readers to know the change trend of scores at different grade. It's very clear to know whether and when the students improved their performance.**