Anthony Huang
INF2178
March 18th

Assginment 3

Introduction:

The development of academic skills in early childhood is crucial for long-term educational outcomes. Among these skills, proficiency in reading and mathematics forms the cornerstone of future academic success. These foundational skills are influenced by a lot of factors, such as educational environment, social status and family income. Understanding how these factors affect academic growth over time is important for the academic development of children. This report focuses on studying the effects of family income on kindergarten students. Specifically, we examine the changes in reading and math scores from the fall to the spring semester, aiming to uncover how these changes are influenced by students' socioeconomic backgrounds.

This report's research question is: "How do kindergarten students' reading and math scores change from fall to spring, and does the change in these scores vary by income group when controlling for their baseline general knowledge scores?" To answer this question, several hypotheses are made:

Hypothesis for Reading Scores:

- H0: There is no significant difference in the change in reading scores from fall to spring among kindergarten students across different income groups when controlling for their baseline general knowledge scores.

- H1: There is a significant difference in the change in reading scores from fall to spring among kindergarten students across different income groups when controlling for their baseline general knowledge scores.

Hypothesis for Math Scores:

- H0: There is no significant difference in the change in math scores from fall to spring among kindergarten students across different income groups when controlling for their baseline general knowledge scores.

- H1 : There is a significant difference in the change in math scores from fall to spring among kindergarten students across different income groups when controlling for their baseline general knowledge scores

This report will use exploratory data analysis, data visualization and multiple one-way ANCOVAs to examine these hypotheses.

Exploratory Data Analysis:

The first step of EDA is to check for missing value in the dataset, and no missing data is found.

To directly address the study's objectives, new variables 'change_in_reading_score' and 'change_in_reading_score' were created to represent the difference between students' academic performance at the beginning (fall) and end (spring) of the academic term. These metrics were chosen as a tangible indicator of academic progress or regression, providing a clear focus for assessing the impact of income groups.

In order to better understand the relationship between the dependent variable changes in score, independent variable income group and the baseline fall general knowledge, two side by side graphs are made (Figure 1 & 2).

Figure 1- Relationship between general knowledge, income and reading score

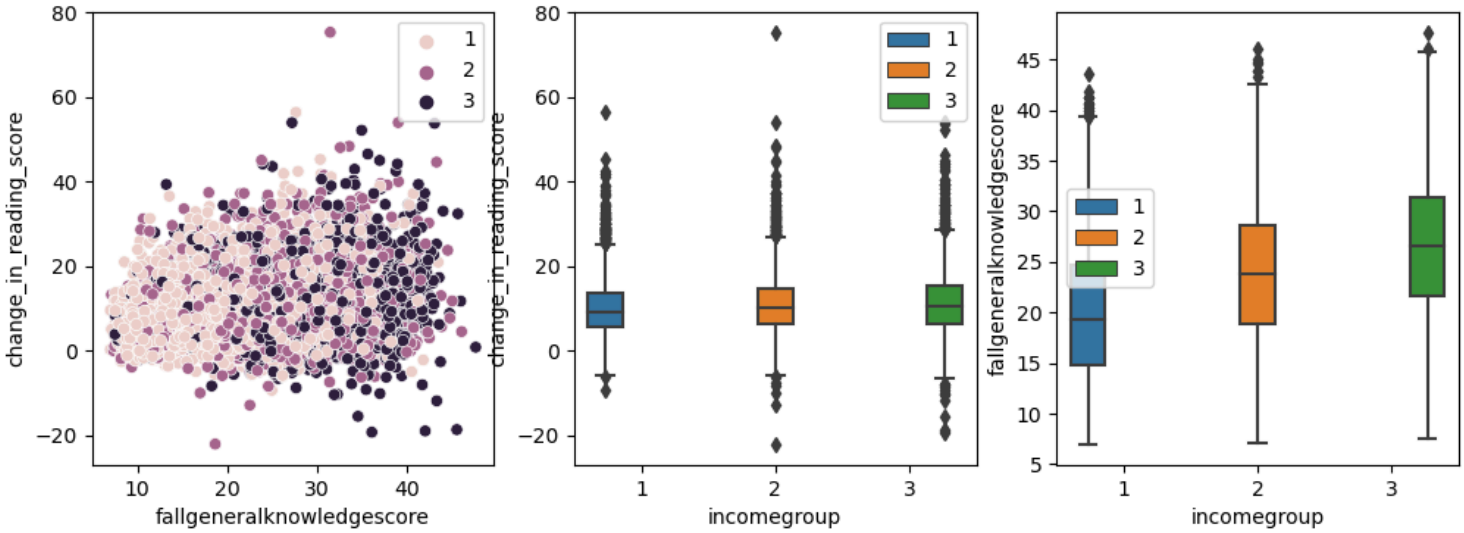


Figure 1 illustrates a potential trend that income groups may be related to both the change in reading scores and the baseline general knowledge scores of kindergarten students. The positive association between baseline general knowledge and changes in reading scores suggests that the starting level of knowledge could play a role in the students' academic development over the school year. The box plots reinforce this trend, showing that median reading scores and general knowledge scores both seem to be higher in the upper income groups. This might indicate that socioeconomic status is correlated with educational outcomes. For the ANCOVA, these plots suggest that baseline general knowledge could be a significant covariate when looking at the change in reading scores across different income groups. However, before proceeding with ANCOVA, further statistical tests would need to be conducted to ensure that all assumptions for ANCOVA are met, including the homogeneity of variances, homogeneity of regression slopes between the covariate and the independent variable.

Figure 2 - Relationship between general knowledge, income and math score

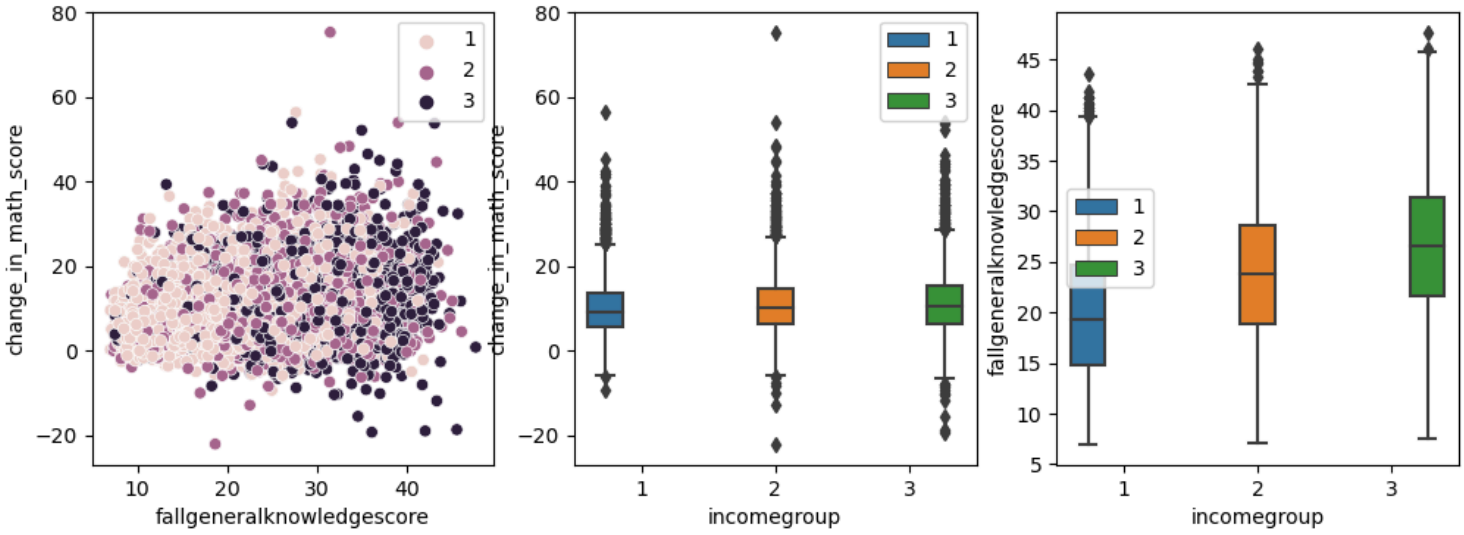


Figure 2 suggests that there are differences in math score changes and baseline knowledge across income groups. The positive correlation between baseline knowledge and math score change could indicate that students with a strong foundational knowledge at the beginning of the school year are more likely to improve in math. Moreover, the median baseline knowledge and changes in math scores seem to be higher for students from higher income groups, hinting at a socioeconomic influence on academic progress. For the purposes of ANCOVA, the plots offer preliminary support for the assumption of linearity between the covariate and the dependent variable, which in this case is the change in math scores. Same as the above graph, further testing is needed to perform ANCOVA.

Assumption Test:

To test the assumption for performing ANCOVA, this report used Levene's test, test for homogeneity of regression slopes, and shapiro test for change in math score and change in reading score. Unexpectedly, all the data failed to pass the assumption tests for performing one-way ANCOVA. The p-value of all the tests are lower than 0.01, suggesting that ANCOVA may not be the best test to use for this dataset. Moreover, both interaction plots (figure 3 & 4) suggest that the interaction effect may not be significant enough. This is likely due to the fact that real world data sometimes is too complex for using ANCOVA. However, ANCOVA tests will still be performed due to the limited scope of this report. Due to the fact that assumption tests did not pass, the result of ANCOVA may not be very accurate, but it can still give us a comprehensive understanding of the relationship between income and academic performance.

Figure 3 - Interaction Plot for Reading Score

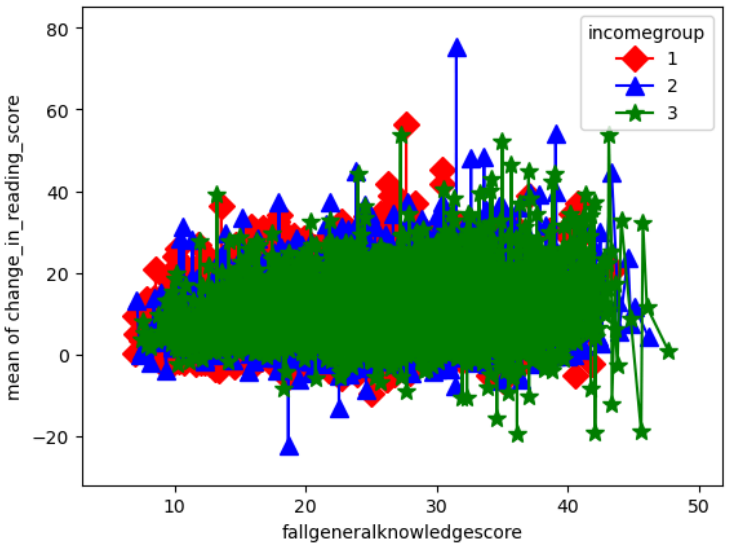
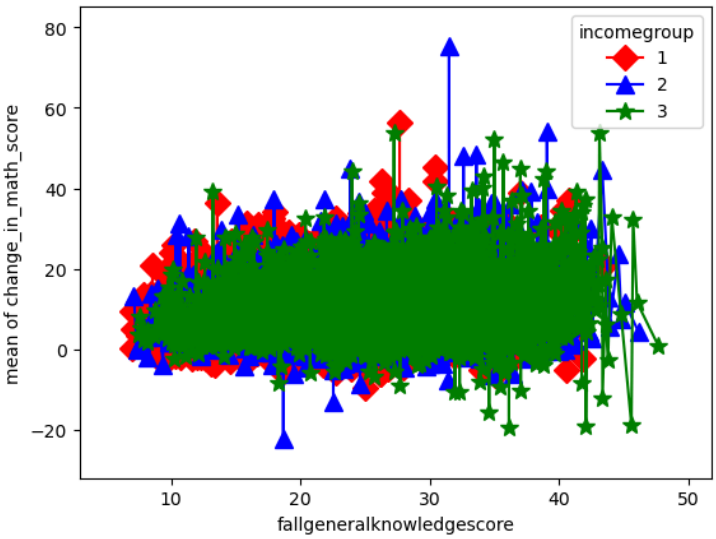


Figure 4 - Interaction Plot for Math Score



ANCOVA Result:

For reading score, The p-value for the income group is smaller than 0.01, which is greater than the common alpha level of 0.05, indicating that the differences between income groups are not statistically significant. The F-statistic for fall general knowledge score is 501.08, which is very high and indicates that the covariate is a significant predictor of the dependent variable. The p-value for fall general knowledge score is smaller than 0.01, suggesting that the effect of general knowledge scores on the dependent variable is highly significant. The lack of significance for the income group could be due to several factors, including a lack of a true effect, insufficient sample size within groups, or variability within income groups that is not accounted for by the model. The significant effect of the fall general knowledge score suggests that this covariate is an important predictor of the dependent variable.

For math scores, the p-value for the income group is smaller than 0.01, suggesting the result is not statistically significant, suggesting no evidence that different income groups have different effects on the dependent variable. The F-statistic is 501.08 for fall general knowledge, indicating that the covariate has a significant relationship with the dependent variable. The p-value is smaller than 0.01, suggesting that the relationship between the fall general knowledge score and the dependent variable is highly statistically significant.

In both cases, we failed to reject the null hypothesis.

Conclusion:

In conclusion, our analysis found no significant difference in the changes in reading and math scores among kindergarten students from different income groups when controlling for baseline general knowledge. The general knowledge scores, however, were a significant predictor of the changes in both reading and math scores, highlighting the importance of early general knowledge in academic development. These results suggest educational strategies focused on enhancing general knowledge could be key in supporting early academic growth.