Student Name: Jiajun Yang
Student Number: 1010128862

1. Introduction

In this analysis, I delve into a dataset drawn from a comprehensive study on early childhood education conducted over the academic year of 1998-1999. This particular dataset, referred to as "INF2178 A3 data.csv", captures a snapshot of kindergarten students' abilities across three core areas: reading, mathematics, and general knowledge. These abilities were measured twice—once at the outset of the fall semester and again in the spring, allowing for an assessment of academic growth over a critical period in early education. Additionally, the dataset includes information on household income, categorized to examine the potential influence of socioeconomic status on educational outcomes. The primary focus of this study is to analyze changes in reading and math scores relative to students' income groups, using their general knowledge scores as a reference point.

Research Questions

1. How do kindergarten students' reading and math scores change from fall to spring across different income groups?
2. Is there a significant interaction between students' income groups and their baseline general knowledge scores in predicting changes in reading and math scores over the academic year?

To explore the research questions, this study employs a quantitative methodology, focusing on the use of Analysis of Covariance (ANCOVA) to examine score changes over time, controlling for baseline general knowledge scores. This approach enables an assessment of the differential impact of income groups on the academic progress of kindergarten students in reading and math. Initial exploratory data analysis (EDA) provides insights into the distribution and underlying patterns of the data, laying the groundwork for more detailed statistical analysis.

2. Data Cleaning

Building on the foundation laid out by my research questions and statistic approach, the next critical step involves preparing our dataset for in-depth analysis. This dataset has a total of 9 column with 11933 records. To align the dataset with the analytical needs of my study, I calculated the changes in reading, math, and general knowledge scores from fall to spring for each student. Then I get the following column use to analysis the research question.

- reading_score_change: The calculated difference in reading scores from fall to spring.
- math_score_change: The calculated difference in math scores from fall to spring.
- general_knowledge_score_change: The calculated difference in general knowledge scores from fall to spring.
- Incomegroup: Divides students into three group by household income

3. Exploratory data analysis (EDA)

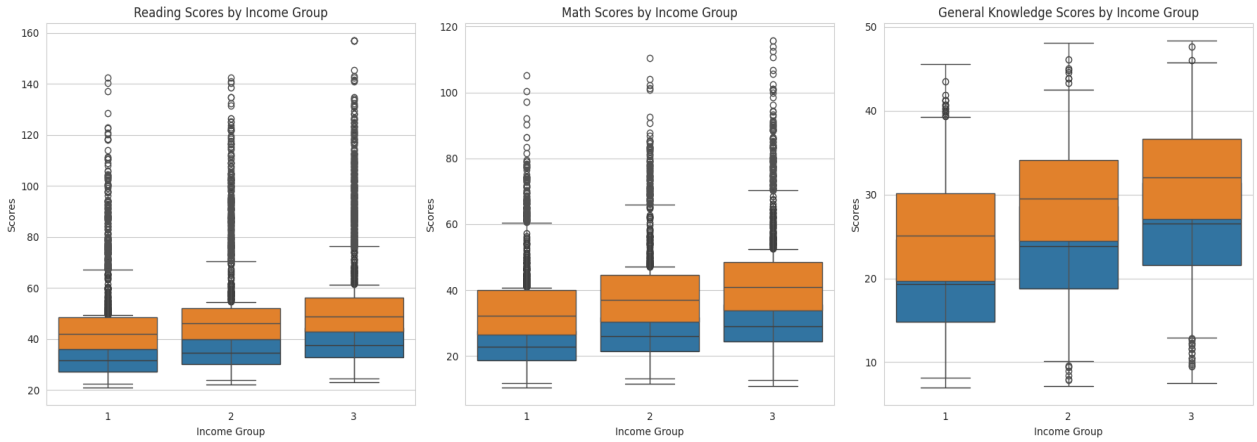

Figure 1. Boxplot for Score Change by Income Group

In the Exploratory Data Analysis (EDA) phase, I produced box plots for reading, math, and general knowledge scores to examine their distribution across different income groups. The box plots for reading and math scores reveal a notable range of scores within each income group, with a considerable number of outliers, especially in the higher scoring regions.

In reading scores, the fall (blue) and spring (orange) scores exhibit consistency in median levels across all income groups, with the spring scores showing a slight upward shift, indicative of overall academic growth. Interestingly, the trend of outliers persists across both terms, suggesting that exceptional literacy skills are not confined to any single income group.
In contrast, math scores reveal a more nuanced relationship with income levels. While all groups show improvements from fall to spring, it's apparent that students from higher-income families tend to have greater gains. Similarly, general knowledge scores exhibit growth, particularly in higher-income brackets, where the breadth of scores widens in the spring. This visible springtime elevation in scores across math and general knowledge for wealthier families points toward a socioeconomically linked compounding effect in educational attainment. In conclusion, the spring observations suggest that math and general knowledge scores improved, especially among wealthier families, which may indicate that socioeconomic that could intersect with other factors. This lays the groundwork for examining whether interactions such as those between income levels and basic knowledge can further uncover differences in educational progress.

4. Interaction plot

Following the boxplot visualization, I pivot to a more focused analysis directly tied to our research questions. To investigate whether socioeconomic status interacts with students' baseline academic abilities to influence educational growth, I analyzed score changes from fall to spring across different income groups using interaction plots.

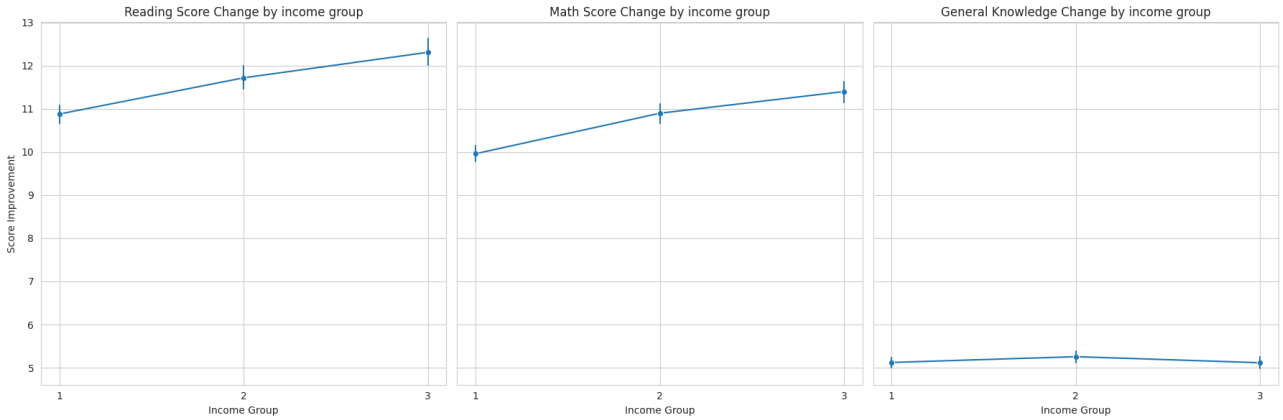

Figure 2. Interaction Plot for Score Change by Income Group

From the Figure 2, it's observed that reading scores tend to rise with higher income levels, suggesting a potential synergy between socioeconomic status and literacy development. Math score improvements also escalate with income, reinforcing the notion that financial resources may play a role in facilitating mathematical learning. In contrast, general knowledge improvements do not display a marked variance across income groups, implying a more complex set of factors at play beyond just socioeconomic status.

These patterns of interaction directly inform our understanding of how the socioeconomic environment may serve as a lever or a barrier to academic development, specifically in reading and math. It illuminates areas where income disparities could be contributing to educational inequality, thereby addressing our core research questions and setting a path for targeted interventions.

5. One-Way ANCOVA

Moving forward in our analytical journey, I turn to One-Way Analysis of Covariance (ANCOVA) to deepen our understanding of the academic growth among kindergarten students across income group. In conducting our analyses, I developed two model that included the change in reading and math score from fall to spring as the two dependent variable, income group as the independent variable and fall general knowledge as a covariate. This approach allowed us to scrutinize the direct impact of income while considering baseline academic achievement.

1. Reading Score Change, Income Group, and Fall General Knowledge Interaction

Variable	Coefficient	Std. Error	t-value	P-value	95% Confidence Interval
Intercept	7.73	0.24	31.96	<0.001	[7.257, 8.205]
Income Group 2 (vs. Group 1)	0.21	0.18	1.20	0.22	[-0.136, 0.570]
Income Group 3 (vs. Group 1)	0.40	0.19	2.11	0.03	[0.029, 0.779]
Fall General Knowledge Score	0.15	0.01	14.83	<0.01	[0.137, 0.179]

Table 1: Reading Score Change, Income Group & Fall General Knowledge Interaction

The regression analysis highlights that students from the highest income group (Group 3) significantly outperform those in the lowest (Group 1) in reading score improvement, with an increase of approximately 0.40 points ($p=0.03$). No significant difference is observed for Group 2 compared to Group 1. Moreover, initial general knowledge strongly predicts reading progress, indicating an average increase of 0.15 points in reading score change for every additional point in fall general knowledge scores ($p<0.001$). These results underline the influence of income group and early knowledge on reading development among kindergarten students.

2. Math Score Change, Income Group, and Fall General Knowledge Interaction

Variable	Coefficient	Std. Error	t-value	P-value	95% Confidence Interval
Intercept	5.98	0.20	29.54	<0.001	[5.586, 6.380]
Income Group 2 (vs. Group 1)	0.15	0.15	1.01	0.31	[-0.143, 0.448]
Income Group 3 (vs. Group 1)	0.14	0.16	0.90	0.03	[-0.170, 0.458]
Fall General Knowledge Score	0.19	0.01	22.38	<0.01	[0.182, 0.217]

Table 2: Math Score Change, Income Group & Fall General Knowledge Interaction

This regression analysis reveals that while both income Group 2 and Group 3 show positive coefficients for math score changes, neither group's change is statistically significant ($p=0.312$ for Group 2 and $p=0.36$ for Group 3), suggesting that income group may not have a discernible impact on math score improvements. In contrast, the fall general knowledge score presents a significant positive coefficient of 0.19 ($p<0.001$), indicating a strong and statistically significant relationship with math score improvements. Every point increase in fall general knowledge score is associated with an average increase of approximately 0.19 points in math score change. This suggests that initial general knowledge plays a crucial role in math achievement over the academic year, overshadowing the influence of socioeconomic status as delineated by income groups.

6. Assumption Check

Assumption 1: Normality of Residuals

The Shapiro-Wilk tests conducted on the residuals of both reading and math models reveal significant deviations from normality, as indicated by p-values<0.001 for both tests. Specifically, the statistic for reading model residuals is 0.89, and for math model residuals, it is 0.96. Although the math scores show a closer fit to normal distribution compared to reading scores, the p-values suggest that the assumption of normality is violated in both cases. This could potentially impact the validity of subsequent ANCOVA results, necessitating a careful interpretation or consideration of alternative analytical methods.

Assumption 2: Homogeneity of Variances

The Levene's test results for both reading and math score changes indicate significant differences in variances across income groups, with p-values far below the typical alpha level of 0.05 ($p=2.79e-09$ for reading and $p=2.34e-10$ for math). The statistics of 19.73 for reading and 22.22

for math highlight notable disparities in the spread of score changes among the different income groups. This violation of the homogeneity of variances assumption suggests that the variability in academic improvements is not consistent across socioeconomic categories, which could influence the interpretability and applicability of the ANCOVA findings.

Assumption 3: Homogeneity of Regression Slopes
The analysis of the homogeneity of regression slopes through interaction terms in our models reveals significant interaction effects between income groups and fall general knowledge scores on both reading and math score changes. Specifically, the interaction terms for income groups 2 and 3 with fall general knowledge scores are statistically significant in predicting reading score changes ($p=0.002$ for Group 2, $p<0.000$ for Group 3) and math score changes ($p=0.002$ for Group 2, $p<0.000$ for Group 3), indicating that the relationship between baseline knowledge and score improvements varies across income levels. This difference suggests that socioeconomic status not only directly affects academic achievement, but also contributes the impact of initial general knowledge on learning progress, highlighting the interplay between economic factors and educational development.

7. Conclusion

This study explored how kindergarten students' reading and math scores change across different income groups and whether there's a significant interaction between their income and initial general knowledge, focusing specifically on the variations in reading and math scores over an academic year. In response to my main research question, I sought to understand the patterns of score changes across income groups and to investigate whether the interplay between socioeconomic context and baseline common sense played an important role in these changes. Analyses based on ANOVA and hypothesis testing suggest that students from higher-income families typically show greater gains in reading and math scores. In addition, my investigation of interaction effects points to a subtle mechanism by which socioeconomic status not only directly affects academic achievement, but also contributes the impact of initial general knowledge on educational progress.

These findings emphasize the profound impact of socioeconomic disparities on equity in education from kindergarten onwards. Therefore, I consider that educational policies and interventions need to be sensitive to the socioeconomic backgrounds of students and can provide targeted support to close the initial knowledge gap and promote equitable academic development.