Student Name: Tony Yang
Student Number: 1003289229

Introduction

The development of academic skills in early childhood is crucial to long-term educational outcomes, with proficiency in foundational areas like reading and mathematics serving as the step-stone of future academic success. These skills are influenced by various factors, including the educational environment, social status, and family income. Understanding the interplay among these factors and their collective impact on academic growth is important for promoting children's educational development. This report focuses on the influence of family income on kindergarten students, specifically scrutinizing changes in reading and math scores from the fall semester to the spring semester. By delving into how these changes are shaped by students' socioeconomic backgrounds, the goal is to see insights that can inform strategies for fostering equitable academic development across diverse socioeconomic contexts. Our exploration will address two fundamental research questions, serving as guiding principles in uncovering the patterns of early childhood education across various income scenarios:

1.Research Questions 1: Does income group have a significant effect on the improvement in fall and spring reading scores, while controlling for the initial fall reading score?

2. Research Questions 2: Is there a statistically significant impact of income group on the change in fall and spring math scores, when adjusting for the initial fall reading score?

Data Preparation

Upon initial examination, the raw dataset contains 9 columns and 11933 rows. Following my review, I determined that no data cleaning is necessary for the scope of my analysis. Since my analysis is quantitative, I have reduced our working dataset to the following columns from the raw dataset. Below is a short description of each column:

- fallreadingscore: Fall reading score of Kindergarten students
- fallmathscore: Fall math score of Kindergarten students.
- fallgeneralknowledgescore: Fall general knowledge score of Kindergarten students.
- springreadingscore: Spring reading score of Kindergarten students.
- springmathscore: Spring math score of Kindergarten students.
- springgeneralknowledgescore: Spring general knowledge score of Kindergarten students.
- incomegroup: Income group classified on total household

Exploratory Data Analysis (EDA)

After data preparation, I delved into a comprehensive Exploratory Data Analysis (EDA) to uncover insights from the dataset. I began by describing our numerical data, as shown in Table 1, which presents a summary of these variables. Additionally, I used box plots and histogram plots for all fall and spring scores to visually represent how these features are distributed in Figure 1 and 2. This detailed analysis provided a clear understanding of the general trends within each feature.

	Fall reading score	Fall math score	Fall General Knowledge score	Spring Reading score	Spring Math score	Spring General Knowledge score
count	11933.00	11933.00	11933.00	11933.00	11933.00	11933.00
mean	35.95	27.13	23.07	47.5118	37.80	28.24
std	10.47	9.12	7.40	14.33	12.03	7.58
min	21.01	10.51	6.99	22.35	11.90	7.86
25%	29.34	20.68	17.39	38.95	29.2	22.80
50%	34.06	25.68	22.95	45.32	36.41	28.58
75%	39.89	31.59	28.31	51.77	44.22	33.78

max	138.51	115.65	47.69	156.85	113.80	48.35

Table 1 : Quantitative Data Statistics

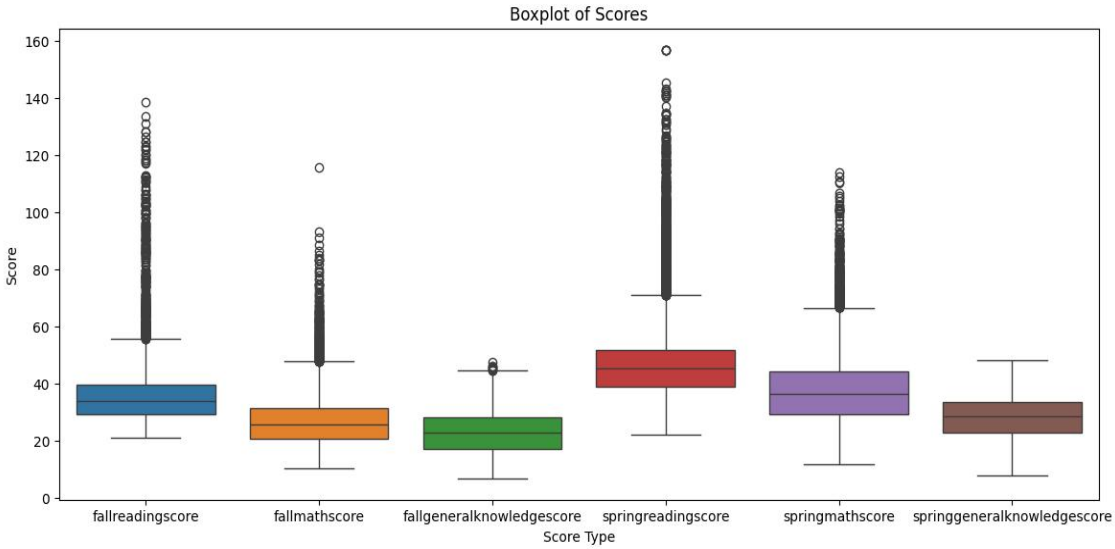

Figure 1: Box Plot Of Fall And Spring Scores

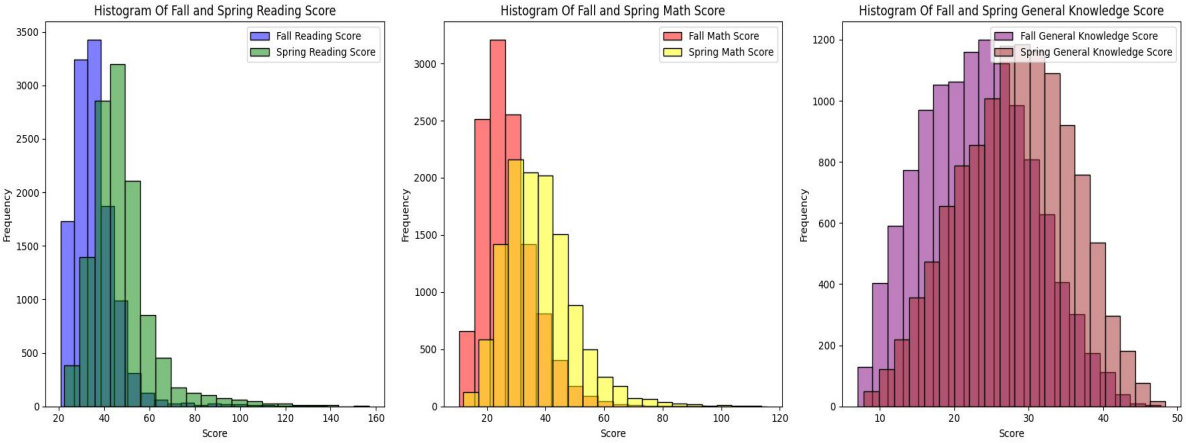

Figure 2: Histogram Of Fall And Spring Scores

Based on the data and plots presented in Table 1 and Figure1 , it is evident that there are variations in average scores and standard deviations across different subjects and between the Fall and Spring assessments. In Fall, the average scores for reading, math, and general knowledge are approximately 35.95, 27.13, and 23.07 respectively, with corresponding standard deviations of 10.47, 9.12, and 7.40. Moving to the Spring assessments, there is an increase in average scores, with reading averaging around 47.51, math approximately 37.80, and general knowledge about 28.24, accompanied by wider standard deviations of 14.33, 12.03, and 7.58 respectively. In Figure 2, the histograms of all reading, math, and general knowledge scores show the distributions for Spring exhibit a noticeable rightward shift compared to Fall, indicating an overall enhancement in all scores. This shift holds significance for my research questions, presenting potential insights. Overlaying the income category onto these distributions enables an examination of whether the mean score shift differs across income groups, while controlling for the initial fall score categories.

One-Way ANCOVA

Research Questions 1: Does income group have a significant effect on the improvement in fall and spring reading scores, while controlling for the initial fall reading score?

For the research question, I aim to determine whether income group has a significant influence on the change in reading scores over time, considering the initial reading level. To explore this, I conducted a one-way ANCOVA analysis. Before delving into the analysis, I created a Interaction Plot, Scatter Plot and Box Plot to visualize the relationship among various income groups.

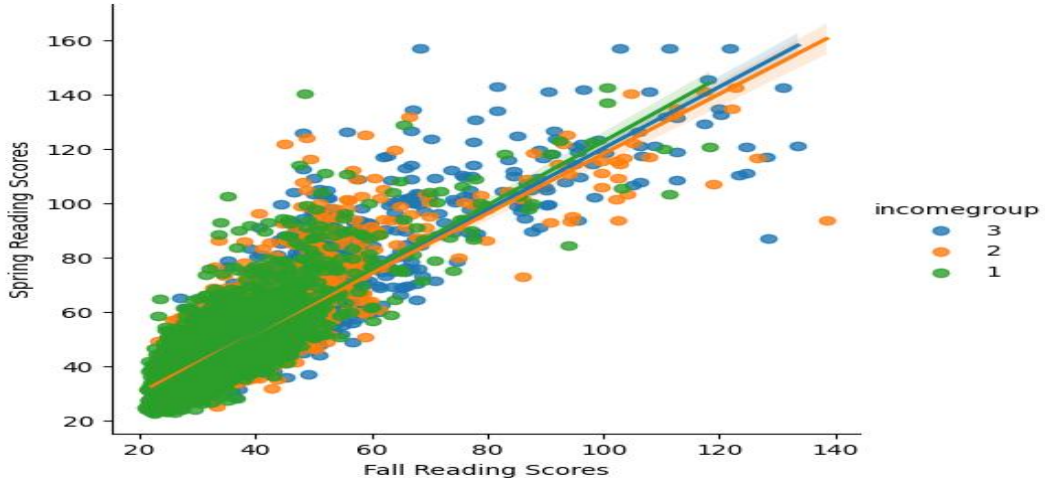

Figure 3: Interaction Plot Of Fall And Spring Reading Scores Among income groups

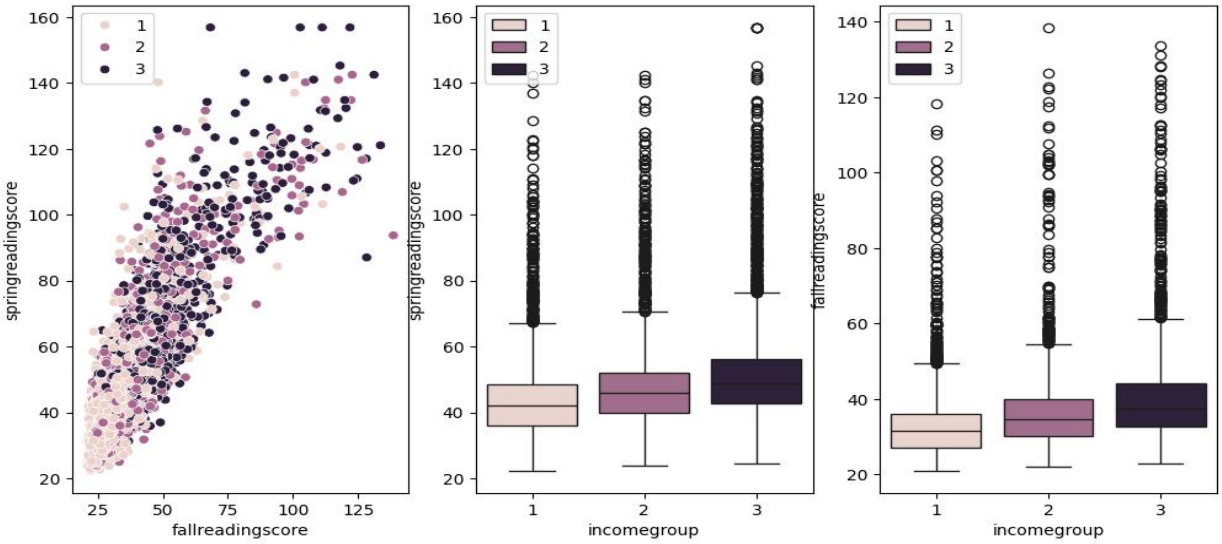

Figure 4: Scatter Plot and Box Plot Of Fall And Spring Scores Among income groups

From Figure 3, it suggests a positive correlation between fall and spring scores, indicating that higher scores in the fall are linked to higher scores in the spring across all income groups. The upward slope of the trend line further illustrates this positive correlation, emphasizing the overall trend towards higher scores in the spring with increasing fall scores. From the Figure 4, it's evident that higher income group 3 exhibits a more substantial improvement in reading scores from fall to spring compared to lower income groups. Following closely is income group 2, showing moderate improvement, while the least improvement is observed in the lowest income group 1.

Table 2: ANCOVA Table results

R-squared	0.692
F-statistic	1.339e+04

Prob (F-statistic)		<0.01
	coef	**P>|t|**
Intercept	6.3272	<0.01
fallreadingscore	1.1322	<0.01
incomegroup	0.2512	0.006 <0.01

From Table 2, it provided valuable insights into the relationship between income groups associated with reading score. It yielded a strong fit (R-squared: 0.692). Both fall reading scores and income group significantly predicted spring reading scores ($p < 0.001$). For each one-unit increase in fall reading score, spring reading score is expected to rise by approximately 1.1322 units. Higher income groups were associated with slightly higher spring reading scores (coefficient: 0.2512). Therefore,income group have a significant effect on the improvement in fall and spring reading scores, while controlling for the initial fall reading score

To validate the one-way ANOVA, certain assumptions need to be met.
Assumption 1: residuals are normally distributed: Shapiro Wilk test.
Assumption 2: homogeneity of variance - normality is not met : Levene's test

	statistic	**p-value**
Shapiro Wilk test	0.91	<0.01
Levene's test	39.55	7.57e-18

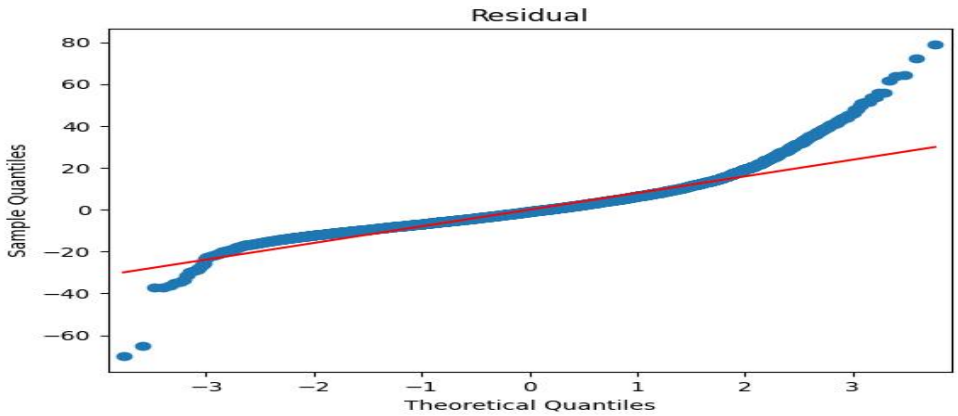

Figure 5: QQ-plot Of Residuals

The Shapiro-Wilk test yielded a statistic of 0.91 with a p-value of less than 0.01. This suggests that the data significantly deviates from a normal distribution, indicating non-normality as shown in Figure 5. Levene's test resulted in a statistic of 39.55 with a very small p-value of approximately 7.57e-18. This indicates a significant difference in variances between groups. Based on the results of the assumption tests conducted above, it suggests that the reliability of the one-way ANCOVA result may be compromised.

Research Questions 2: Is there a statistically significant impact of income group on the change in fall and spring math scores, when adjusting for the initial fall reading score?

To address this question, I also conducted a one-way ANCOVA analysis. Before delving into the analysis, I created a Interaction Plot, Scatter Plot and Box Plot to visualize the relationship among various income groups.

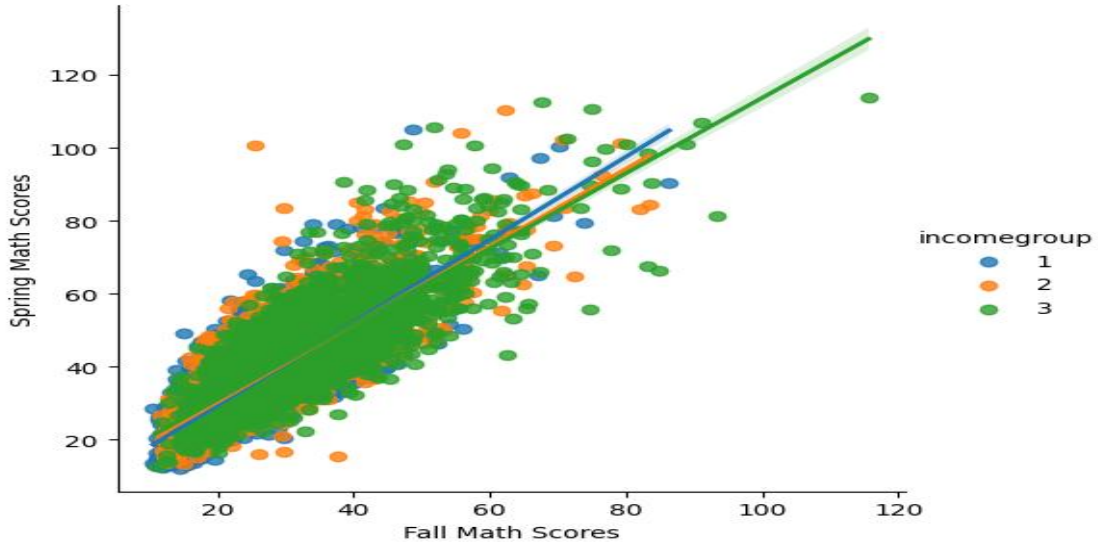

Figure 6: Interaction Plot Of Fall And Spring Math Scores Among income groups

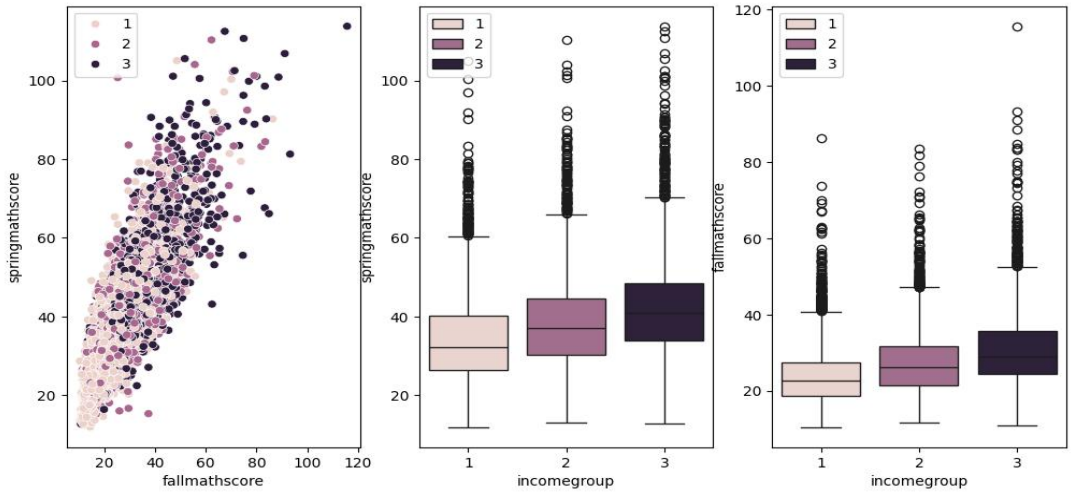

Figure 7: Scatter Plot and Box Plot Of Fall And Spring Scores Among income groups

From the Figure 6, it demonstrate a positive correlation between fall and spring math scores, indicating that students who scored higher in the fall also tended to score higher in the spring. This relationship is depicted by an upward-sloping trend line, suggesting that an improvement in fall scores predicts better performance in the spring. From Figure 7, it's clear that higher income group 3 demonstrates a more significant improvement in math scores from fall to spring compared to lower income groups. Income group 2 shows a moderate improvement, while the lowest income group 1 exhibits the least improvement. Overall, these findings support the idea that higher scores are associated with higher income levels.

Table 3: ANCOVA Table results

R-squared		0.680
F-statistic		1.270e+04
Prob (F-statistic)		<0.01
	coef	**P>|t|**
Intercept	7.7862	<0.01
fallmathscore	1.0735	<0.01
incomegroup	0.4699	<0.01

From Table 3, the regression model yielded a highly significant result (F-statistic = 1.270e+04, $p < 0.01$), with an R-squared value of 0.680, indicating that approximately 68.0% of the variance in spring math scores can be explained by fall math scores and income group. Both fall math scores (coef = 1.0735, $p < 0.01$) and income group (coef = 0.4699, $p < 0.01$) were significant predictors of spring math scores, suggesting that higher fall math scores and income groups were associated with higher spring math scores. Therefore, there is a statistically significant impact of income group on the change in fall and spring math scores.

To validate the one-way ANOVA, certain assumptions need to be met.
Assumption 1: residuals are normally distributed: Shapiro Wilk test.
Assumption 2: homogeneity of variance - normality is not met : Levene's test

	statistic	**p-value**
Shapiro Wilk test	0.96	<0.01
Levene's test	18.90	6.38e-09

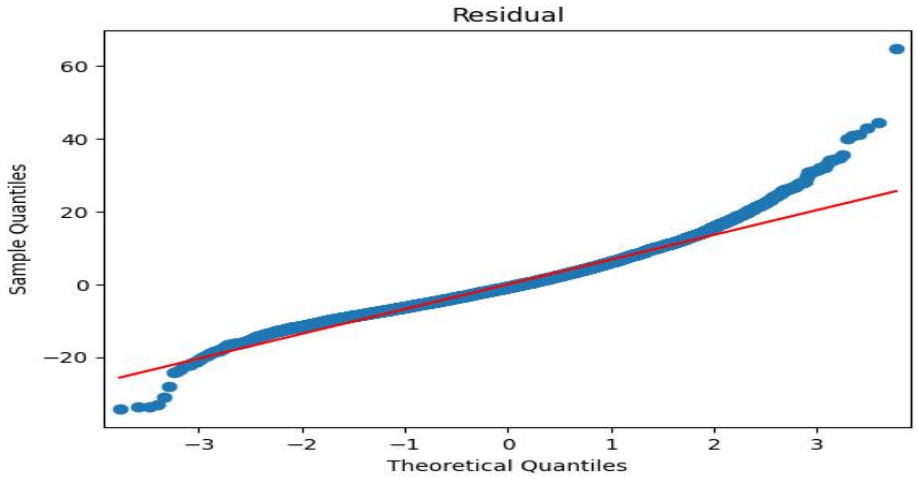

Figure 8: QQ-plot Of Residuals

The Shapiro-Wilk test yielded a statistic of 0.96 with a p-value of less than 0.01. This suggests that the data significantly deviates from a normal distribution, indicating non-normality as shown in Figure 8. Levene's test resulted in a statistic of 18.90 with a very small p-value of approximately 6.38e-09. This indicates a significant difference in variances between groups. Based on the results of the assumption tests conducted above, it suggests that the reliability of the one-way ANCOVA result may be compromised.

Conclusion

The One-way ANCOVA analyses offered valuable insights of the early childhood study among different income groups. While these analyses provided valuable information, we faced challenges as the tests did not meet all the expected assumptions, underscoring the complexity of utilizing ANCOVA in this context. Addressing these limitations through robust sensitivity analyses and exploring alternative modeling techniques will improve the validity and reliability of the results