Examining Factors Influencing Kindergarten Academic Progress in Reading and Math

Xiaojing Liu (#1004263888)

1. Introduction

The role that schooling plays in influencing a child's academic path has come to light more and more in recent years. Early reading and numeracy abilities provide the groundwork for further learning. It is crucial to understand factors that impacting academic advancement. Notably, a child's knowledge is one of the most important factors influencing the academic progress.

The report provides a thorough quantitative analysis of kindergarten educational improvement, particularly in reading and math scores, from 1998 fall to 1999 spring using several one-way ANCOVAs. I examine the relationship between beginning knowledge and the change in reading and math scores from 1998 to 1999 using the dataset named "INF2178_A3_data.csv".

In order to untangle the academic advancement in reading and math linkages, our investigation will focus on two core research questions.
1. How does the income group influence the kindergarten children's reading score progress from fall to spring, controlling for their initial knowledge score in fall 1998?
2. How does the income group influence the kindergarten children's math score progress from fall to spring, controlling for their initial knowledge score in fall 1998?

2. Data cleaning and Data Wrangling

As part of the preparation steps for the analysis, I performed data cleaning and data processing. First, I examined the data for missing values or outliers that could affect the quality of analyses. There were no missing values in the key fields, so score changes could be calculated directly.

I added two new variables to capture students' academic progress.
Change in Reading Scores: It is the difference between the reading scores of spring and fall.
Change in Math Scores: It is the difference between the math scores of spring and fall.

3. Exploratory Data Analysis

The exploratory data analysis focused on uncovering initial patterns and insights from the data related to progress in reading and maths attainment from the Autumn term to the Spring term.

1) Summary statistics for dependent variable change in reading scores by income group

income group	n	mean	std
1	4729	10.8783	7.4622
2	3726	11.7169	7.9061
3	3478	12.3083	8.9676

n: number of observations
mean: the average of the change in reading scores
std: the standard deviation of the change in reading scores

Interpretation:

- The mean change in reading scores tends to increase as the income group increases, suggesting that children from higher income families may make greater progress in reading from fall to spring.
- Standard deviation (change in reading score) for the higher income groups is higher, suggesting higher variability in the higher income group.

2) Summary statistics for dependent variable change in math scores by income group

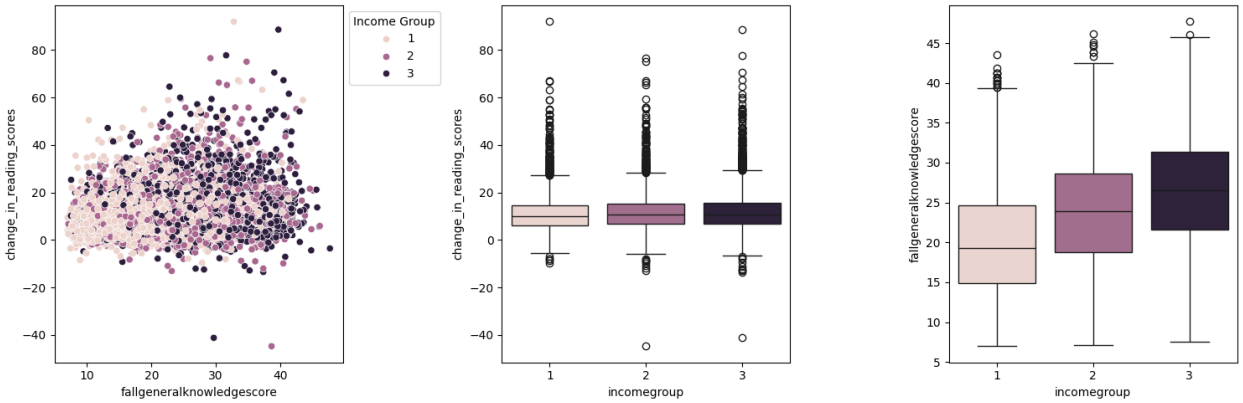
income group	n	mean	std

1	4729	9.9585	6.4065
2	3726	10.8962	6.8537
3	3478	11.3992	7.3390

Interpretation:

- The mean of maths score variability increased with higher income groups.
- The standard deviation is higher with higher income groups, suggesting greater variability in maths score change for students from wealthier backgrounds.

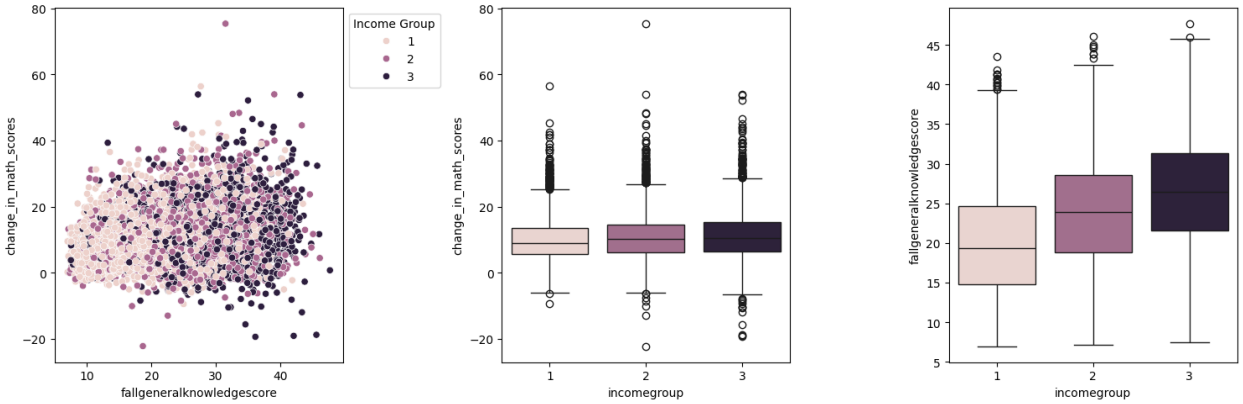
3) Data visualization – Change in Reading Score Analysis



Scatterplot interpretation: The scatterplot shows the relationship between children's autumn general knowledge scores and the change in their reading scores. For children from all income groups, the dots seem to move slightly upwards as they move to the right, which mean that children who learnt more in the fall tended to see a greater increase in their scores.

Box plots provide a visual summary of the same descriptive statistics as in the tables. The detailed explanation is below.

4) Data visualization – Change in Math Score Analysis



Scatterplot interpretation: The scatterplot shows the relationship between a child's fall general knowledge score and the change in their math score. For children from income group 1 and 2, the dots seem to move slightly upwards as they move to the right, which mean that children who learnt more in the fall tended to see a greater increase in their scores. For Income Group 3, the dots seem to move slightly downward as they move to the right, which may mean that students who know more in the fall tend to have smaller increases in their scores.

The boxplots provide a visual summary of the same descriptive statistics shown in the table. The descriptive statistics information has explained in the below.

5) a. Descriptive statistic table (change in reading score by income group)

Income Group	Q1	Median	Q3	IQR	Outliers
1	5.98	9.990	14.45	8.47	134
2	6.8	10.695	15.4	8.6	127
3	6.7125	10.75	15.77	9.0575	178

Interpretation:

- The median change in reading scores tends to be higher for higher income groups, suggesting that income may be positively associated with improved reading scores.
- The number of outliers increases with income group, which may mean that while most high-income students' reading scores improved, a significant number of students had unusually large changes in reading scores.
- the IQR increased with income group suggests that there was more variability in reading score changes among the higher income groups.

b. Descriptive statistic table (fall general knowledge score by income group)

Income Group	Q1	Median	Q3	IQR	Outliers
1	14.84	19.298	24.643	8.47	15
2	18.81	23.869	28.5928	8.6	7
3	21.6065	26.5055	31.367	9.0575	2

Interpretation:

- The higher the income, the higher the knowledge score. Each of the higher income groups had higher Q1, median and Q3 values, indicating an overall shift towards higher scores.
- The IQR values are relatively similar across groups, indicating a consistent distribution of scores around the median across income groups.
- The number of outliers decreases with higher income groups, which may imply greater consistency of scores within these groups.

c. Descriptive statistic table (change in math score by income group)

income group	Q1	Median	Q3	IQR	Outliers
1	5.57	9.04	13.47	7.9	107
2	6.31	10.22	14.5475	8.2375	97
3	6.43	10.60	15.32	8.89	80

Interpretation:

- Median change in maths scores is higher for higher income group, suggesting a positive correlation between income level and improved maths scores.
- The number of outliers decreases as income groups increase, which may indicate that higher income groups have less variability in score change.
- The IQR increases slightly with each income group, suggesting that the higher income groups have more variability in the middle 50% portion of the score change.

4. One-Way ANCOVAs Analysis

a. One-Way ANCOVAs (dependent variable: change in reading score, independent variable: income group, covariates: fall general knowledge score

Source	SS	DF	F	p-unc	np^2
income group	287.4859	2	2.2512	1.053126e-01	0.000377
fall general knowledge score	14054.1247	1	220.1103	2.354473E-49	0.018117
Residual	761671.0364	11929	NaN	NaN	NaN

Interpretation:
SS (Sum of Squares): This indicates the variance attributed to each source. For the income group, the SS was 287.4859 and for the fall total knowledge score, the SS was 14054.1247, which indicates that the covariates explained more variance in the dependent variable than the independent variables.
F (F-Statistic): This is the ratio of the model mean square to the error mean square. It tests whether the explained variance of the model is significantly greater than the unexplained variance. The F-statistic for the income group is 2.2512, while the F-statistic for the fall total knowledge score is significantly higher at 220.1103.
p-unc (p-value): P-value used to test significance of each source. For income group, the p-value is 0.1053126e-01 (or 0.01053), which is statistically significant. For fall general knowledge score, the p-value is exceedingly small (2.354473e-49), indicating that the covariate is highly significant.
np^2(Partial Eta Squared): This indicates the effect size, which shows the proportion of total variation attributable to a variable, controlling for other variables. The np2 for the income group is 0.000377, which is very small, indicating a negligible effect. For fall general knowledge score, np2 is 0.018117, indicating a small to medium effect size based on common benchmarks.

b. One-Way ANCOVAs (dependent variable: change in math score, independent variable: income group, covariates: fall general knowledge score

Source	SS	DF	F	p-unc	np^2
income group	55.8796	2	0.6243	5.356614e-01	0.000105
fall general knowledge score	22425.9329	1	501.0839	9.425259e-109	0.040312
Residual	533880.4998	11929	NaN	NaN	NaN

Interpretation:
SS (Sum of Squares): The SS for income group is 55.8796 which indicates that the variation in the dependent variable is related to income group. The SS for the general knowledge score is 22425.9329, which is much higher, indicating that it accounts for a greater amount of variation in the dependent variable.
F (F-Statistic): For the income group, the F-statistic is 0.6243, while for the non-general knowledge scores, the F-statistic is 501.0839. The larger the F-statistic, the greater the effect on the dependent variable.
p-unc (p-value): The p-value of 0.5356614e-01 (or 0.05356614) for the income group is slightly higher than the traditional 0.05 alpha level, indicating that the income group does not significantly predict change in math achievement. Fall general knowledge score has a very small p-value (9.425259e-109), suggesting that this covariate significantly predicts change in math achievement.
np^2(Partial Eta Squared): For the income group, np2 is 0.000105, which is negligible. For fall general knowledge score, np2 is 0.040312, which is a small to moderate effect size.

5. Interaction Plot

a.

Interpretation:

- Across all income groups, the change in reading scores increased as general knowledge scores increased in the fall. This suggests a positive correlation between general knowledge and improved reading skills.
- The slope of the straight line (blue line) is steeper for income group 1 than for income groups 2 and 3. This suggests that the relationship between general knowledge and improved reading scores is stronger for the low-income group.

b.

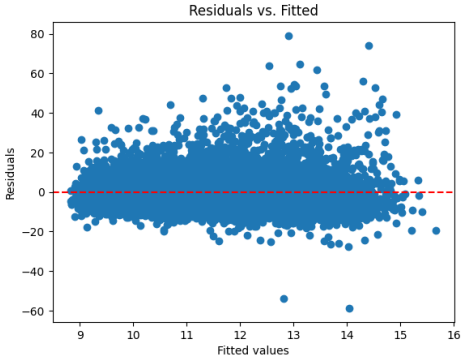
Interpretation:

- All three lines have positive slopes, indicating an overall trend of higher math scores for all income groups as general knowledge scores increase in the fall.
- The slope of the straight line (green line) is steeper for income group 1 than for income groups 2 and 3. For the low-income group, the relationship between general knowledge and improved math achievement is stronger.

6. Check for ANCOVAs assumptions

a. ANCOVAs (DV: change in reading score, covariant: fall general knowledge score independent variable: income group (categorical variable)

Assumption 1: Linearity of variance



Interpretation: This is Residual vs fitted scatter plot. There is a clear pattern that suggests that the relationship between the fitted values (the predicted values of the model) and the residuals (the difference between the observed and predicted values) is not linear. The residuals increase as the fitted values increase, which is a sign that the variance is not constant (heteroskedastic) and indicates a potential problem with the assumption of linearity.

Assumption 2: Homogeneity of Variance

Levene's test

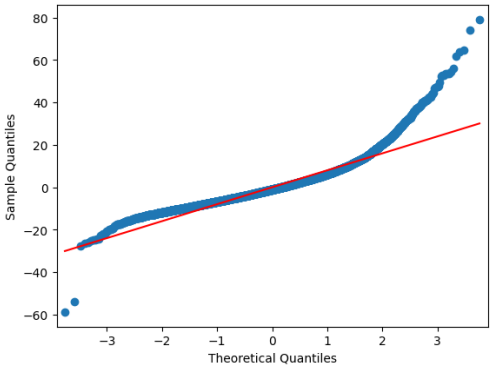
Levene Result	statistics	p-value
	38.2065	6.4672e-10 <0.05

Interpretation:

Since the p-value is much smaller than the standard alpha level of 0.05 (in fact, it is very close to 0) So, there is statistically significant evidence that the variance of "change_in_reading_scores" is not equal across levels of "fall general knowledge score".

Assumption 3 Normality of residuals

Q-Q plot



Shapiro-Wilk test

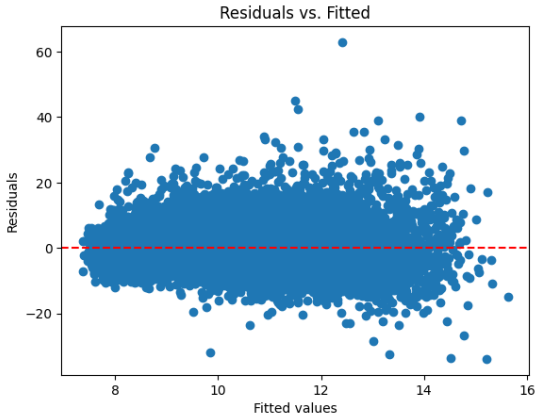
Shapiro-Wilk Result	Statistics	p-value
	0.8996	0.0 <0.05

Interpretation:

- These points do not follow a straight line. The curvature indicates that the tails of the residual distribution are larger than the tails of the normal distribution, which is an indication of non-normality.
- Test statistic: A value of approximately 0.8996 indicates how well the data fit the normal distribution. The closer the value is to 1, the more it fits the normal distribution.
- P-value: A P-value of 0 indicates that there is significant evidence to reject the null hypothesis that the residuals are normally distributed.

b. ANCOVAs (DV: change in math score, covariant: fall general knowledge score independent variable: income group (categorical variable)

Assumption 1: Linearity of variance



Interpretation: The absence of a distinct pattern (e.g., U-shaped or funnel-shaped) suggests that the relationship is linear, and that the variance of the residuals is constant over the range of fitted values.

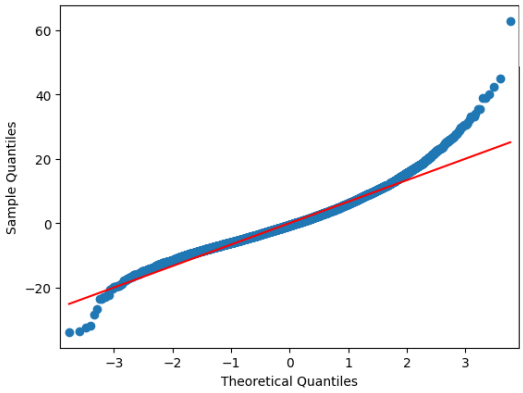
Assumption 2: Homogeneity of Variance

Levene Result	statistics	p-value
	22.2152	2.3442e-10 < 0.05

Interpretation:
Because the p-value is very small, there is very strong evidence against the null hypothesis. It implies that there is statistically significant evidence that the variance of change in math achievement is not equal across income groups.

Assumption 3 Normality of residuals

Q-Q plot



Shapiro-Wilk test

Shapiro-Wilk Result	Statistics	p-value
	0.9964	0.0 <0.05

Interpretation:

- The points deviate slightly from the red line, especially at the ends. This indicates that the residuals are not normally distributed.
- Statistic: The value of 0.9964 is very close to 1, indicating a good fit to the normal distribution.
- P-value: The P-value of 0.0 is less than the commonly used alpha level of 0.05, indicating that there is strong evidence against the null hypothesis, and therefore I conclude that the residuals do not follow a normal distribution.

7. Conclusion

The analysis is conducted with the aim of revealing the impact of income group and initial knowledge on the academic progress of kindergarten children, with a focus on their improvement in reading and math scores. The study employed meticulous methods, including data cleaning, organizing, exploratory analyses, and one-way ANOVA, to paint a clear picture of young children's educational trajectories.

Through our investigation, I found that Initial general knowledge plays a more critical role in children's academic progress than in the income group. These findings have important implications for educational policy and intervention strategies. They suggest the importance of focusing on increasing children's general knowledge in early education to support academic progress. Surprisingly, the effect of income group on academic achievement was not significant, suggesting that economic background is not as critical as one might expect when predicting growth in academic achievement in these areas, which is related to initial knowledge. This may prompt consideration of more targeted educational support in early education that focuses on knowledge enrichment, regardless of a child's socioeconomic background.