INF 2178 – Winter 2024
Technical Assignment 3 – Narrative of findings
Samantha Yin (1004194258)
Mar 23, 2024

Introduction

The connection between a child's background and their learning outcomes has always been a question for educators, psychologists, and social scientists. This topic centres on the impact of socioeconomic factors on educational performance, which may lead to larger societal issues such as income inequality and access to quality education. The kindergarten phase is often seen as the location where foundational cognitive and learning patterns are established. It offers a snapshot of the early academic behaviour that may influence a child's educational journey. Family income is often seen as one of the most important factors that reflect socioeconomic background.

This study will focus on understanding family income levels and how they intersect with children's initial general knowledge in shaping children's academic progress. Due to the complexity of educational development, children's general knowledge is considered a covariate in the study. This helps to isolate the specific contribution of socioeconomic background while holding constant the baseline learning ability reflected in general knowledge scores. The study aims to find deep insights into the connection between child performance and family income level through the following research questions:

Research Question 1: Given the initial general knowledge of kindergarten students, how does family income impact the improvement in reading scores over the year?

Research Question 2: Given the initial general knowledge of kindergarten students, how does family income impact the improvement in math scores over the year?

Data

The dataset we used is from a longitudinal study on the educational progression of kindergarten students from fall 1998 to spring 1999. The data includes a total of 9 columns and 11,933 entries. Based on the research questions, only the most relevant features will be utilized in further studies. The core data are numerical variables including 3 key areas: reading scores, math scores, and general knowledge scores. The scores were assessed twice, once in the fall and once in the spring. This provides a measure of the growth and educational development that occurred within this timeframe. In the analysis, capturing the development and progress of students is crucial. Hence, two new variables: **'change_in_reading'** and **'change_in_math'** (shown in the last 2 columns in Table 1) are created. These variables are calculated by subtracting the fall scores from the spring scores in their respective subjects for each student. This calculation provides a direct measure of educational advancement where negative is decrease, positive is increase in scores.

Table 1. Summary of Statistics

	fallreadingscore	fallmathscore	fallgeneralknowledgescore	springreadingscore	springmathscore	springgeneralknowledgescore	change_in_reading	change_in_math
count	11933.00	11933.00	11933.00	11933.00	11933.00	11933.00	11933.00	11933.00
mean	35.95	27.13	23.07	47.51	37.80	28.24	11.56	10.67
std	10.47	9.12	7.40	14.33	12.03	7.58	8.09	6.86
min	21.01	10.51	6.99	22.35	11.90	7.86	-44.76	-22.16
0.25	29.34	20.68	17.39	38.95	29.27	22.80	6.47	6.01
0.5	34.06	25.68	22.95	45.32	36.41	28.58	10.40	9.86
0.75	39.89	31.59	28.31	51.77	44.22	33.78	15.15	14.33
max	138.51	115.65	47.69	156.85	113.80	48.35	91.94	75.35

Table 1 offers a statistical summary for the numerical variables in our dataset. The data shows a notable increase in the mean scores from fall to spring across all subjects, which indicates an overall academic growth. The range of all scores highlights the diverse levels of academic ability and development. Although some students had a decline in scores, on average, students experienced a positive increase in both their reading and math scores. The average increase in reading scores is approximately 11.56 points, while in math, it is around 10.67 points. In the histograms (Figure 1), the distribution of the change in scores for reading and math shows positive increase in both subjects, as the histograms peaks are on the positive side. The skew towards the lower end for both reading and math changes indicates that while most students improved, a smaller proportion exhibited minimal change. The histograms for general knowledge for both fall and spring present a more symmetric distribution.

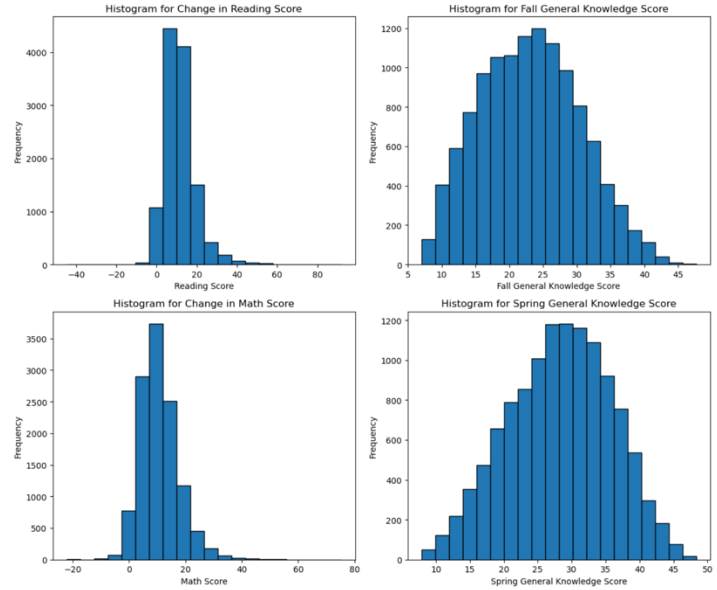

Figure 1

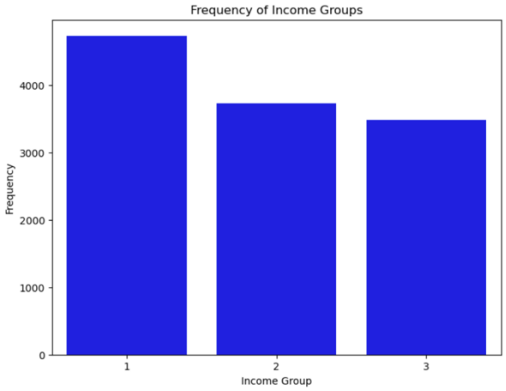

Figure 2

The other important part of the data is a categorical variable: **income group** derived from total income, with group 1 indicating the lowest income group and group 3 representing the highest group. Figure 2 above shows the distribution of students across the three income groups. The frequencies indicate a relatively even distribution, with slightly fewer students in the middle and higher income group (group 2 and 3). This distribution is crucial as it provides a balanced representation of the socioeconomic factors, allowing for more generalized insights into how income levels correlate with academic performance.

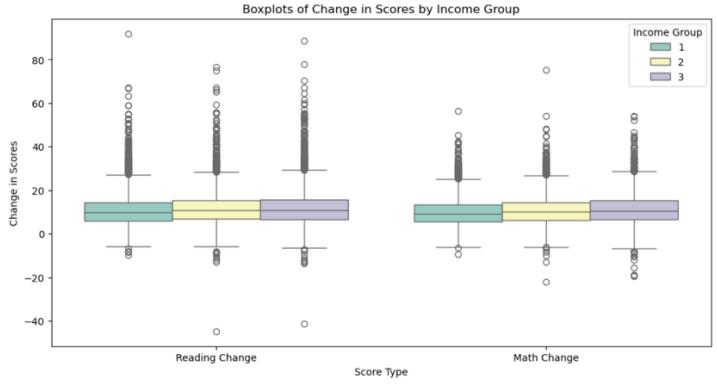

Figure 3

Figure 3 is the boxplot of the changes in reading and math scores divided by income group. There is variability in score changes among students across all income groups. Although the median seems consistent, there are variations in the spread and distribution of score changes across income groups. The variability in score changes among students across all income groups suggests further analysis, such as ANCOVA, to statistically determine if these differences are significant when controlling for other variables like initial general knowledge.

ANCOVA 1 – Improvement of Reading Scores

This section will focus on the first research question: Given the initial general knowledge of kindergarten students, how does family income category impact the improvement in reading scores over the year?

Figure 4 is the interaction plot of general knowledge scores and the change in their reading scores with colours representing the 3 income groups. From the plot, the upward slopes show positive trend where higher fall general knowledge scores lead to an increased trend in the change in reading scores. In terms of income group, the points show an overlapping trend without distinct trends. All groups seem to follow a similar pattern. Further analysis will be needed to analyze how general knowledge and income group impact the improvement in reading scores.

Figure 4

An ANCOVA was conducted with the dependent variable: change_in_reading, independent variable: income group, and covariate as fall general knowledge score. The general knowledge of fall is used as a covariate in this analysis as it serves as a baseline for academic ability. It can help isolate the effect of income group on reading score changes. Table 2 shows the ANCOVA result with only key numbers of the model. Initial general knowledge is the covariate in the analysis. The second row of Table 2 shows a large F value of 220.110 and a significant p-value ($p < 0.001$). This gives us strong evidence to believe that the fall general knowledge is statistically significant, which indicates students with higher general knowledge scores at the beginning of the period showed greater improvements in reading scores. 'Incomegroup' is the independent variable used to predict the reading score improvement. From the first row of Table 2, the F-value for income group is at 2.251 with a p-value of 0.105. The p-value is greater than the significance level of 0.05, hence, we

don't have evidence to reject the null hypothesis that there is no difference in the mean among levels of the income groups. This means the differences in reading score improvements across income groups are not statistically significant after using children's initial general knowledge as covariates.

Table 2. ANCOVA Result

	F-statistics	p-value	np^2
Income Group	2.251	0.105	<0.001
Initial general knowledge	220.110	<0.001	0.018

From this result, it seems that initial general knowledge is a significant predictor of reading score improvements and should be a key focus of educational interventions. On the other hand, family income does not appear to have a significant independent effect on the improvement of reading scores throughout the academic year when controlling for initial general knowledge. However, the np^2 for both fallgeneralknowledgescore and income group is 0.018 and <0.001, respectively. The small and negligible effect size may indicate limited practical applications.

Assumption Testing

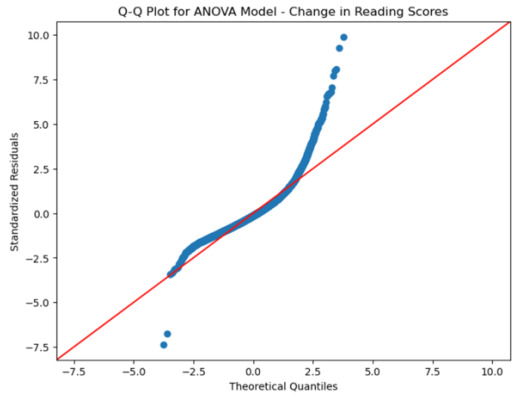

Figure 5

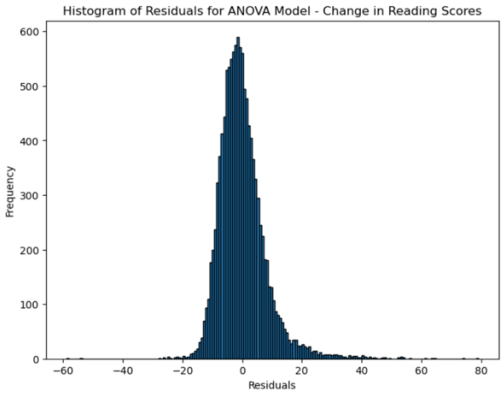

Figure 6

Normality: The QQ plot for the model of change in reading scores (Figure 5) shows that the residuals roughly follow the 45-degree line representing the expected distribution under normality. However, there is a deviation in the heads and tails. Similarly, the histogram of residuals (Figure 6) shows a distribution that is higher and narrower than the normal distribution. We can confirm this visual inspection by the Shapiro-Wilk test where the p-value is <0.001. This suggests that the data might not follow a normal distribution, which violates the assumption of normality.
Homogeneity: Since the data is not normally distributed, a Levene's test was conducted and resulted in a p-value of <0.001. This indicates the variances across the groups are not equal and the homogeneity of variance assumption is violated.

Based on the above, the assumptions for ANCOVA might not hold. The conclusions drawn from the ANCOVA should be approached with caution. Further tests can be conducted to analyze the result.

ANCOVA 2 – Improvement of Math Scores

This section will focus on the second research question: Given the initial general knowledge of kindergarten students, how does family income category impact the improvement in math scores over the year? Figure 7 is the interaction plot for changes in math scores in relation to fall general

knowledge scores and income groups. Similar to the reading scores interaction plot, there is a positive correlation indicating that higher general knowledge scores are associated with greater improvements in math scores, as shown by the inclining slope. There is a slight difference between the slope of the 3 income groups. The line for the highest income group 3 seems to have a less steep slope compared to the other groups. Similarly, an ANCOVA was conducted with the dependent variable as 'change_in_math', independent variable as 'incomegroup', and covariate as fall general knowledge score to analyze how general knowledge and income group impact the improvement in math scores.

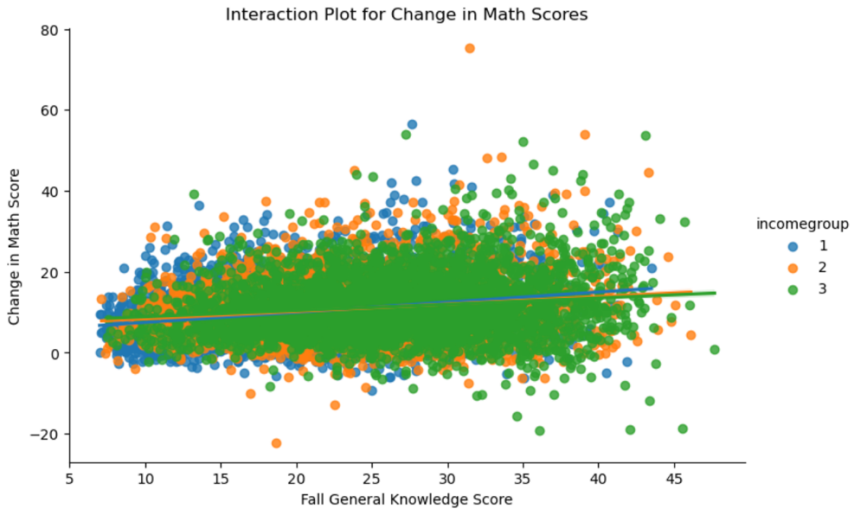

Figure 7

Table 3 is the brief summary of the ANCOVA result for the impact of family income on the improvement of math scores over the academic year, controlling for initial general knowledge. Similar to the first ANCOVA, the covariate initial general knowledge has a p-value <0.001, and an F-statistic of 501.084. The high F-statistic and very low p-value indicate a very low chance that this relationship is due to random chance. This gives us strong evidence that the initial general knowledge is statistically significant in predicting the improvement of math scores. In Table 3, income group has an F-statistic of 0.624, with a p-value of 0.536. The p-value is greater than the significance level of 0.05, which means we do not have sufficient evidence to reject the null hypothesis. Therefore, the family income group does not have a statistical effect on math score improvement when initial general knowledge is controlled. The np^2 less than 0.001 indicates the negligible effect size, which also reinforces that income group does not explain a significant portion of the variance in math score improvement.

Table 3. ANCOVA Result

	F-statistics	p-value	np^2
Income Group	0.624	0.536	<0.001
Initial general knowledge	501.084	<0.001	0.040

Based on the above analysis, we can conclude that income group does not have a statistically significant impact on student's improvement when controlling for initial general knowledge, whereas student's initial general knowledge can directly affect the improvements of math scores.

Assumption Testing

Normality: The QQ plot in Figure 8 shows the residuals near the head and tails deviated significantly from the 45-degree line representing the expected distribution under normality. The histogram of residuals in Figure 9 shows a distribution centred around the middle, however, it is

higher and narrower than the normal distribution bell shaped curve. The Shapiro-Wilk test was conducted, resulting in a p-value <0.001. This suggests that the data does not follow a normal distribution, which indicates the violation of the assumption of normality.
Homogeneity: Again, the Levene test was used due to the non-normal distribution and resulted in a p-value <0.001. This indicates the variances across the groups are not equal and the homogeneity of variance assumption is violated.

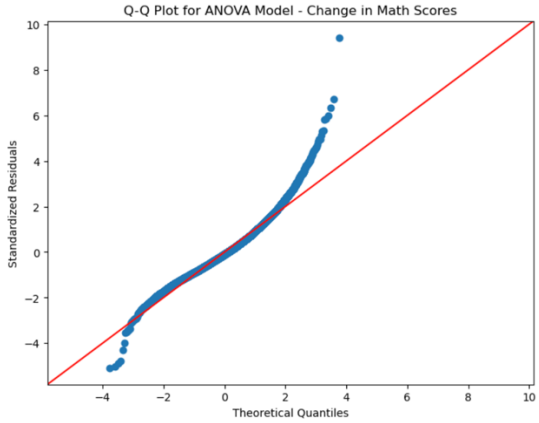

Figure 8

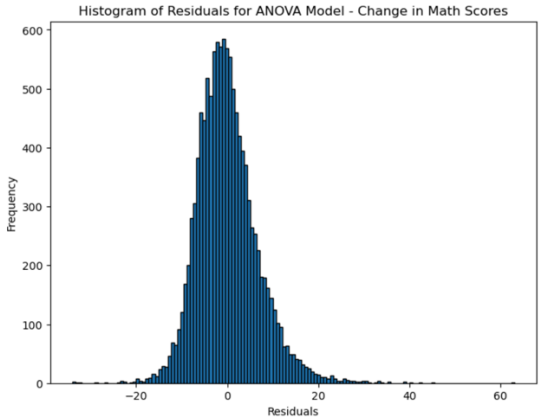

Figure 9

Based on the above, the assumptions for ANCOVA might not hold. The conclusions drawn from the ANCOVA should be approached with caution.

Conclusion

The exploratory data analysis and the ANCOVAs provide valuable insights into the influence of family income on the academic progression of kindergarten students in reading and math, while accounting for the student's initial general knowledge. The results from the ANCOVAs indicated that initial general knowledge significantly predicted both reading and math score improvements, but income groups did not have a statistically significant effect.

Broader implications of the study are the understanding the complex interactions between socioeconomic factors, educational factors, and student performances. Since ANCOVA finds that income group is not a significant factor after controlling for initial knowledge, early interventions could aim at improving initial general knowledge to enhance academic growth rather than solely addressing socioeconomic disparities. From the government's perspective on policy planning, the finding shows the importance of early general knowledge in the academic development of children. The lack of direct effect of income on both reading and math score improvements within the context of this study indicate less effort should be put into socioeconomic status. The result suggests more effort should be put into enhancing general knowledge in the early years, for example, hosting programs that provide diverse learning materials and experiences to enrich student's background knowledge, regardless of their socioeconomic status. Future studies could explore longitudinal data to track these variable's impact over a more extended period, e.g., beyond kindergarten into primary school years. In order to understand the complex relation of socioeconomic factors and education, further research should continue to explore this topic to inform more equitable educational strategies.