Course: INF2178

Assignment: Assignment 3

Name: Jingwen Ji

Student number: 1006128101

Exploring the Impact of Socio-Economic Factors on Educational Outcomes in Early Childhood

Introduction

Early childhood education plays a pivotal role in shaping academic performance and future opportunities for children. Understanding the factors that influence educational outcomes during this critical developmental stage is essential for designing effective interventions and policies. In this study, we investigate the impact of socio-economic factors, specifically income and general knowledge, on academic performance in early childhood. By employing advanced statistical methods such as ANCOVA (Analysis of Covariance), we aim to explore how changes in income and general knowledge scores relate to changes in academic performance, specifically in reading and math. Additionally, we seek to assess whether income impacts academic performance independently or through its association with general knowledge.

Research question:

1. To what extent do changes in math scores from fall to spring vary across income groups in early childhood, while accounting for the impact of general knowledge?
2. To what extent do changes in reading scores from fall to spring vary across income groups in early childhood, while accounting for the impact of general knowledge?

Exploratory Data Analysis

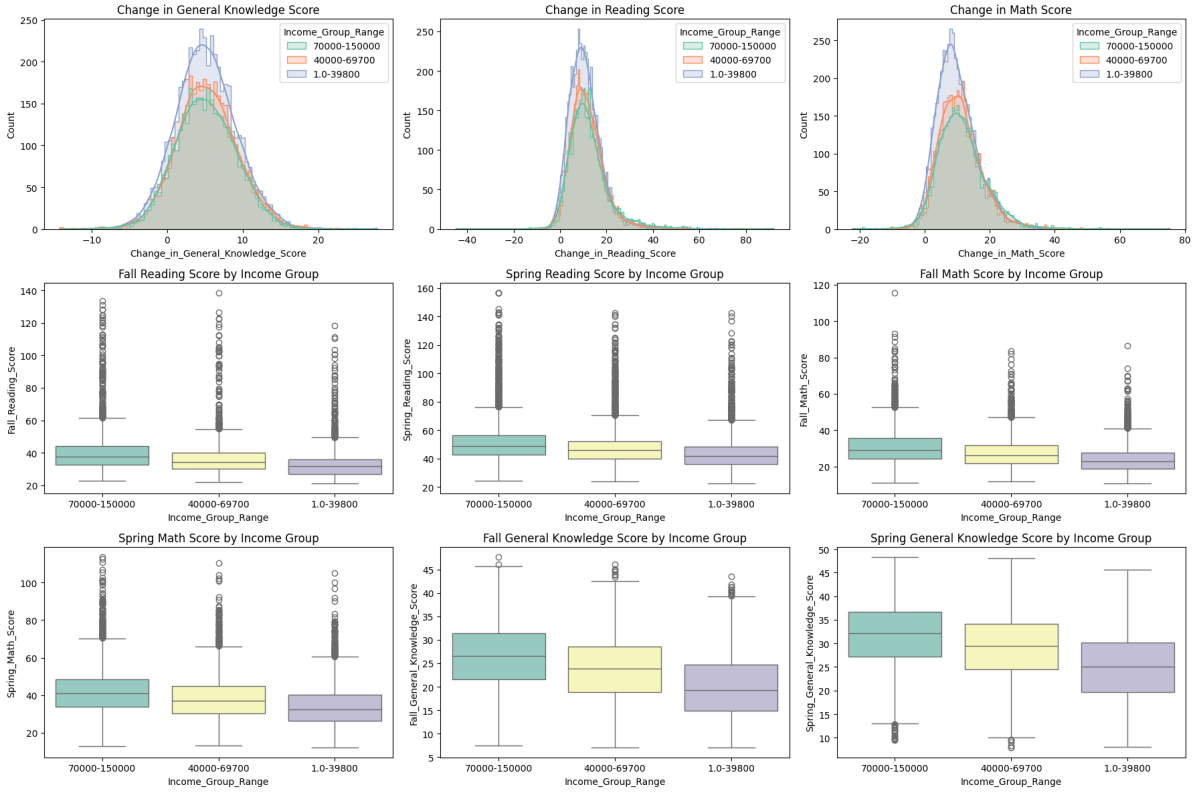
The dataset has a total of 9 columns with 11933 rows. After modifying the column names for better readability, I outlined the data and its' dictionary of what each column means.

Column Name	Description
Fall_Reading_Score	Student's reading score of the fall term
Fall_Math_Score	Student's math score of the fall term
Fall_General_Knowledge_Score	Student's general knowledge score of the fall term
Spring_Reading_Score	Student's reading score of the spring term
Spring_Math_Score	Student's math score of the spring term
Spring_General_Knowledge_Score	Student's general knowledge score of the spring term
Total_Household_Income	Total household income
Income_In_Thousands	Total household income (thousands)
Income_Group	Income group category derived from household income

	Fall_Reading_Score	Fall_Math_Score	Fall_General_Knowledge_Score	Spring_Reading_Score	Spring_Math_Score	Spring_General_Knowledge_Score
mean	35.954	27.128	23.074	47.511	37.799	28.236
std	10.473	9.121	7.397	14.327	12.028	7.577
min	21.01	10.51	6.985	22.35	11.9	7.858
25%	29.34	20.68	17.385	38.95	29.27	22.802
50%	34.06	25.68	22.954	45.32	36.41	28.583
75%	39.89	31.59	28.305	51.77	44.22	33.782
max	138.51	115.65	47.691	156.85	113.8	48.345

	Counts	Percentage
1.0-39800	4729	39.63
40000-69700	3726	31.22
70000-150000	3478	29.15

We observe from the summary of statistics that the means of General Knowledge, Math, and Reading scores increase from Fall to Spring. Additionally, the second table indicates the wide range of household income, spanning from $1 to $150,000. This disparity in income levels suggests that the data was collected from a sample representing significant socioeconomic diversity.



The top row indicates the distribution of changes in scores across different income groups.

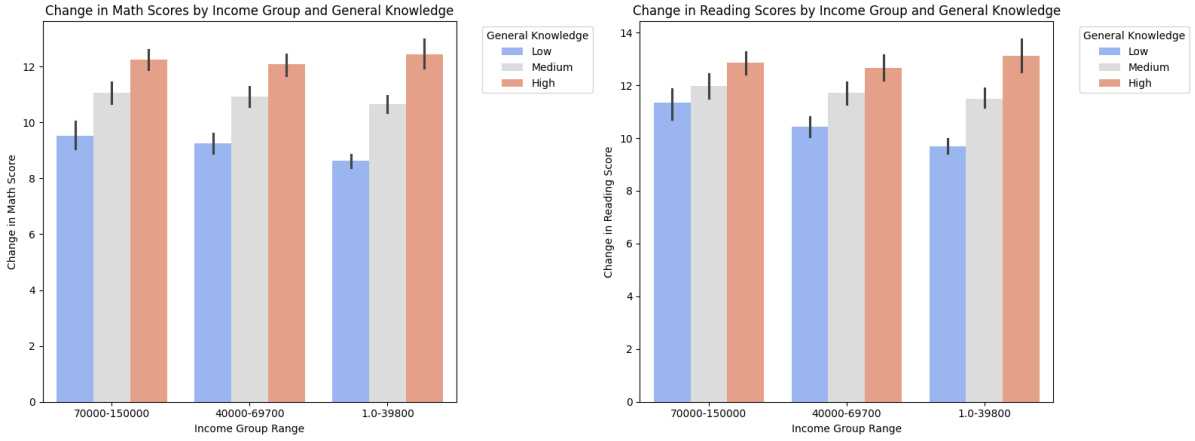
1. **General Knowledge**: The change in General Knowledge scores shows a relatively symmetrical distribution around zero for all income groups, indicating that on average, students across all socioeconomic backgrounds experienced similar changes in their general knowledge scores. However, there's a visible skew in the tails of the distributions.
2. **Reading**: The distribution of changes in Reading scores appears to be slightly right-skewed, especially for the lowest income group, suggesting an overall improvement in reading scores. This improvement is more pronounced and variable in the higher income groups, as indicated by the broader spread and the peak of the distribution being further to the right.

3. **Math**: The changes in Math scores exhibit a right-skewed distribution for all income groups, with the highest income group showing a more significant right skew. This suggests that students from the higher income group generally experienced more substantial improvements in their math scores compared to those from lower income groups.

The middle and bottom rows show the seasonal scores by income group.

1. **Reading and Math Scores**: For both Fall and Spring terms, higher income groups tend to have higher median scores in Reading and Math, indicating a correlation between income and academic performance in these subjects. The spread of scores and outliers also suggest variability within income groups, with higher income groups tending to have a wider range of scores.
2. **General Knowledge Scores**: Similar to Reading and Math, the General Knowledge scores for both Fall and Spring terms show higher median scores for higher income groups. However, the difference in median scores between income groups appears more pronounced in General Knowledge than in Reading or Math, especially in the Spring term.

   There is a clear indication that socioeconomic status, as represented by income group, has an association with academic performance. Higher income groups tend to have higher median scores across all subjects and terms, suggesting that they may have access to more resources or environments that support academic success. The changes in scores suggest that students across all income groups experience improvements in academic performance over time, with variability in the extent of improvement among different subjects and income levels.



Income and Knowledge Impact: Both charts highlight the significant role of income and general knowledge in educational outcomes, with a stronger impact seen in low-income groups. This suggests that interventions aimed at improving general knowledge could have meaningful impacts on academic performance, particularly in math, for students from lower-income backgrounds.

ANCOVA Test

Null Hypothesis (H_0): There is no significant difference in the changes in math scores from fall to spring across different income groups in early childhood, even after accounting for the impact of general knowledge.

Alternative Hypothesis (H_a): There is a significant difference in the changes in math scores from fall to spring among at least two of the income groups in early childhood, after accounting for the impact of general knowledge.

Metric	Value
Dependent Variable	Change_in_Math_Score

Metric	Value
F-statistic	200.0
Prob (F-statistic)	1.36e-126
Intercept (coef)	5.9826
ANCOVA - C(Income_Group)	PR(>F) = 5.356614e-01
ANCOVA - Fall_General_Knowledge_Score	PR(>F) = 9.425259e-109

The average change in Math score when all independent variables are 0 is approximately 5.98 points. With every one-unit increase in Fall General Knowledge Score, there is a significant increase of 0.1993 in the change in Math score, which is statistically significant ($p < 0.001$). This suggests that general knowledge has a positive impact on the improvement of Math scores. The `C(Income_Group)` variable does not significantly contribute to the variance in Math score changes (`PR(>F)` = 0.536e-01), aligning with the OLS regression results. The `Fall_General_Knowledge_Score` significantly affects the variance in Math score changes (`PR(>F)` < 0.001), highlighting its importance as a predictor.

Null Hypothesis (H_0): There is no significant difference in the changes in reading scores from fall to spring across different income groups in early childhood, after accounting for the impact of general knowledge.

Alternative Hypothesis (H_a): There is a significant difference in the changes in reading scores from fall to spring among at least two of the income groups in early childhood, after accounting for the impact of general knowledge.

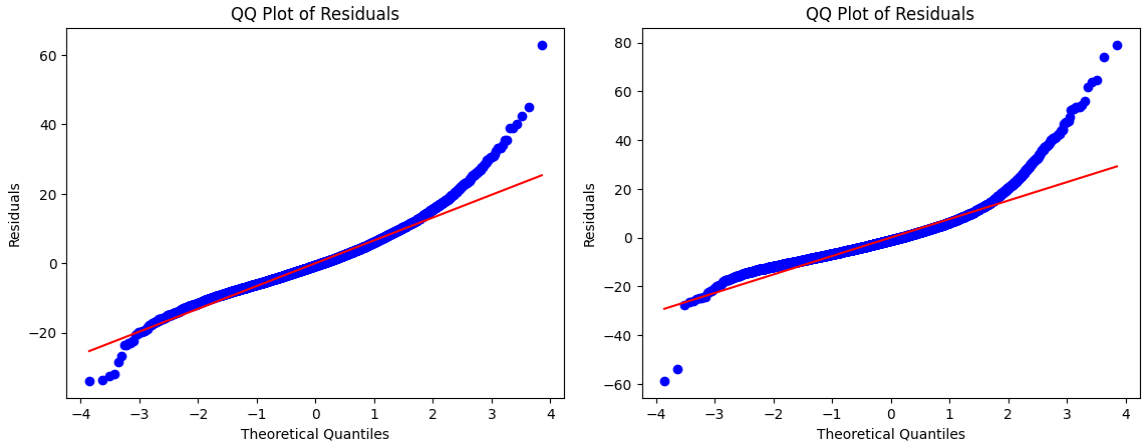
Metric	Value
Dependent Variable	Change_in_Reading_Score
F-statistic	95.49
Prob (F-statistic)	4.52e-61
Intercept (coef)	7.7308
ANCOVA - C(Income_Group)	PR(>F) = 1.053126e-01
ANCOVA - Fall_General_Knowledge_Score	PR(>F) = 2.354473e-49

The base change in Reading score, with other variables held at zero, is approximately 7.73 points. A one-unit increase in the Fall General Knowledge Score is associated with a significant increase of 0.1578 in the change in Reading score ($p < 0.001$), highlighting the importance of general knowledge in reading improvement. The `C(Income_Group)` variable's contribution to variance in Reading score changes is not significant at the 0.05 level (`PR(>F)` = 0.105e-01), though it is close, suggesting a potential, albeit weak, income effect. The `Fall_General_Knowledge_Score` significantly affects variance in Reading score changes (`PR(>F)` < 0.001), reinforcing its predictive value.

Fall General Knowledge Score: It plays a significant role in both Math and Reading score changes. The positive coefficients in both models indicate that as general knowledge increases, so do the changes in scores for Math and Reading. This underscores the broad impact of foundational knowledge across different subject areas, suggesting that interventions aimed at increasing general knowledge could be beneficial for improving academic performance across the board.

Income Group Impact: The analysis presents a nuanced view of socioeconomic status's influence on educational outcomes. For Math scores, income group differences were not statistically significant, suggesting that within the context of this analysis, income level alone does not strongly predict changes in Math scores. However, in Reading scores, the highest income group (group 3) showed a significant difference from the baseline, indicating some level of socioeconomic impact on Reading improvements. This difference in income group impact between Math and Reading might suggest that reading improvements are more sensitive to socioeconomic factors, or perhaps that Math achievement is influenced by other unmeasured variables.

Both the Shapiro-Wilk and Levene test results indicate that the assumptions of normality and homogeneity of variances are violated in our dataset. As the p values are extremely small, implies that the null hypothesis of normality and the null hypothesis that variances are equal across groups are rejected.



Both QQ plots suggest significant deviations from normality in the residuals, which violates one of the key assumptions of linear regression that the residuals should be normally distributed. This non-normality can affect the reliability of hypothesis tests related to regression coefficients and may suggest that linear regression is not the best model for these data, or that transformations of the data might be necessary.

group1	group2	meandiff	p-adj	lower	upper	reject
1.0-39800	40000-69700	0.9377	<0.001	0.5871	1.2883	True
1.0-39800	70000-150000	1.4406	<0.001	1.0831	1.7982	True
40000-69700	70000-150000	0.503	0.0051	0.1256	0.8803	True

The Tukey HSD test indicates significant differences in the changes in math scores between all pairs of income groups:

- Low vs. Middle (1.0-39800 vs. 40000-69700) with a mean difference of 0.9377 (p-adj < 0.001)
- Low vs. High (1.0-39800 vs. 70000-150000) with a mean difference of 1.4406 (p-adj < 0.001)
- Middle vs. High (40000-69700 vs. 70000-150000) with a mean difference of 0.503 (p-adj = 0.0051)

These results suggest that income level is associated with changes in math scores, with higher income groups showing greater improvements.

group1	group2	meandiff	p-adj	lower	upper	reject
1.0-39800	40000-69700	0.8387	<0.001	0.4246	1.2527	True
1.0-39800	70000-150000	1.4301	<0.001	1.0079	1.8523	True
40000-69700	70000-150000	0.5915	0.0053	0.1458	1.0371	True

The Tukey HSD test indicates significant differences in the changes in math scores between all pairs of income groups:

- Low vs. Middle (1.0-39800 vs. 40000-69700) with a mean difference of 0.8387 (p-adj < 0.001)
- Low vs. High (1.0-39800 vs. 70000-150000) with a mean difference of 1.4301 (p-adj < 0.001)
- Middle vs. High (40000-69700 vs. 70000-150000) with a mean difference of 0.5913 (p-adj = 0.0053)

These results suggest that income level is associated with changes in reading scores, with higher income groups showing greater improvements.

Conclusion: The results reveal that at during the kindergarten years a students income level has a limited impact in improving their reading and math scores. Both of the null hypothesis can not be rejected. Thus educational interventions should be specifically tailored to accommodate not only the socio-economic backgrounds of students but also their initial levels of general knowledge to maximize effectiveness. These nuanced findings advocate for a departure from the 'one-size-fits-all' approach towards more customized educational strategies that recognize the diverse needs across different socio-economic and knowledge backgrounds. This approach should inform educational policies and teaching methodologies, ensuring they are adaptive and responsive to the varied profiles of student groups to facilitate more equitable and effective educational outcomes.