INF2178
Assignment 3
Xuying Du

Influence of Income Group on Academic Score Changes: A One-Way ANCOVA Analysis

1. Introduction

This report is devoted to investigating the effect of income level groups on the variation in students' performance in math, reading, and general knowledge scores over the academic year. The raw data is derived from the 'INF2178_A3_data.csv' has a total of **8 columns** with **11934 rows**. Via Analysis of Covariance (ANCOVA), we can control the initial scores which allows us to determine the only real effect of income group on the changing factor in performance.

The linkage between socioeconomic status and academic achievement is a well-established domain of educational research. The current study addresses this intersection by investigating the following research questions:

- Does income group influence the change in math scores from fall to spring, after accounting for initial general knowledge?
- Is the change in reading scores from fall to spring across income groups moderated by initial reading scores?
- How does income group interact with changes in math and reading scores to influence the change in general knowledge?

2. Methodology:

A quantitative approach using ANCOVA is employed to adjust for initial score differences while examining the influence of income groups on score changes. The dataset is built using scores gathered from students from three sets of income groups, which allows a much more meaningful interpretation.

3. Exploratory Data Analysis (EDA):

We proceeded with a comprehensive EDA to leverage insight that could point in the direction of research questions. Firstly, we started by describing our quantitative data as shown in Figure 1 below. Distribution patterns are visually inspected, not indicating significant outliers. shown in Figure 2.

	Income group	n	mean	std
0	1	4729	25.069492	7.248229
1	2	3726	29.143605	6.965300
2	3	3478	31.567718	6.928347

Figure 1: Descriptive Statistics Summary Table

INF2178
Assignment 3
Xuying Du

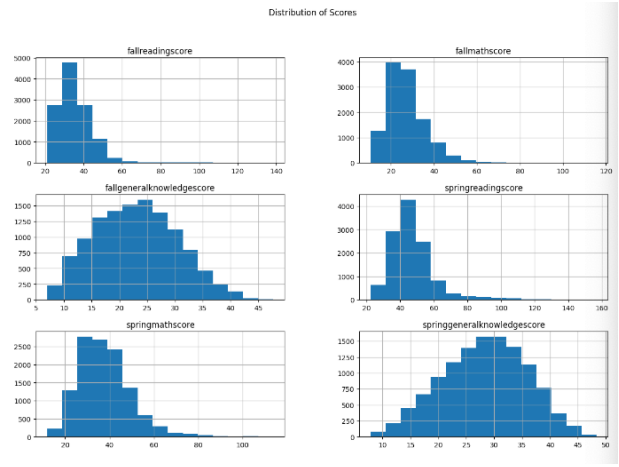

Figure 2: Distribution of Scores

Figure 3 portrays the beneficial features of a correlation heat map which visually elucidates the link between the academic scores belonging to different subjects, identifying potential factors that influence students' performance. Based on the heatmap, there is a strong positive correlation between fall and spring scores within the same subjects, with coefficients above 0.8 indicating a strong relationship. For instance, fall math and spring math scores also show a particularly strong correlation, as do the reading scores from the two terms. The general knowledge scores between the terms are also highly correlated. On the other hand, correlations between different subjects, like fall reading and spring general knowledge, are lower, indicating a weaker relationship. No significant outliers were recognized in the data, as evidenced by Figure 4's box plots, confirming a uniform distribution.

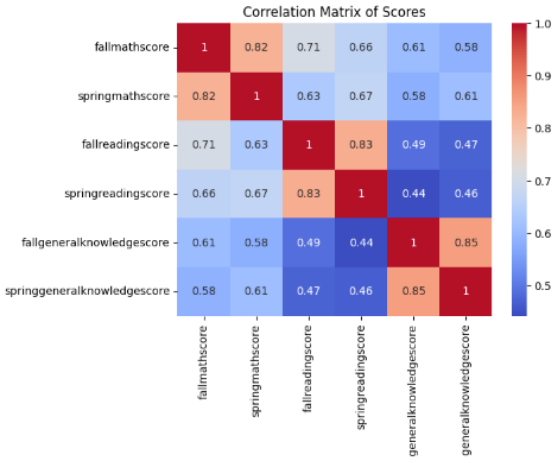

Figure 3: Correlation Matrix of Scores

INF2178
Assignment 3
Xuying Du

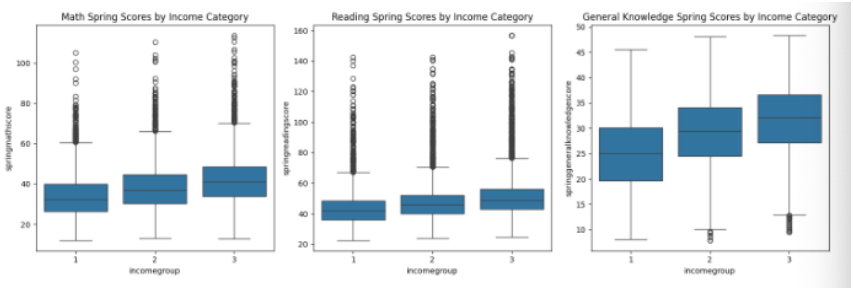

Figure 4: Boxplots

4. Assumption Checks:

Before interpreting the results of ANCOVA, we conducted tests to verify the key assumptions of normality and homogeneity of variances.

Normality of Residuals:

Shapiro-Wilk tests conducted by our models indicated significant departures from normality:

Income Group	Spring Math Score	Spring Reading Score	General Knowledge Score
Group **1**	0.949 p-value < 0.001	0.859 p-value < 0.001	0.994 p-value < 0.001
Group **2**	0.944 p-value < 0.001	0.827 p-value < 0.001	0.995 p-value < 0.001
Group 3	0.938 p-value < 0.001	0.797 p-value < 0.001	0.989 p-value < 0.001

Figure 5 Shapiro-Wilk Test Results for Spring Term Scores by Income Group

Shapiro-Wilk tests for math and reading score changes, both with and without interaction terms, also indicated significant departures from normality.

Variable	Shapiro-Wilk Statistic	P-value
Math Score Change (No Interaction)	0.966	< 0.001
Math Score Change (With Interaction)	0.966	< 0.001
Reading Score Change (No Interaction)	0.912	< 0.001
Reading Score Change (With Interaction)	0.912	< 0.001

Figure 6: Shapiro-Wilk Test Results for Score Changes

INF2178
Assignment 3
Xuying Du

We conducted tests on changes in scores, including general knowledge scores controlling for reading, general knowledge scores controlling for math, and math scores controlling for reading.

Controlled Variable	Test Statistic (w)	P-value
General Knowledge (Reading)	0.9968	< 0.001
General Knowledge (Math)	0.9972	< 0.001
Math Score (Reading)	0.9651	< 0.001

Figure 7: Shapiro-Wilk Normality Test Results for Score Changes with Control Variables

These results suggest that the residuals from our models deviate from a normal distribution, thus violating one of the ANCOVA assumptions. However, it's important to note that while these p-values indicate non-normality, the significance is likely influenced by our large sample size, which can detect even small deviations from normality. Luckily, the central limit theorem reassures us that means across samples are normally distributed for large sample sizes, potentially mitigating the practical impact of this violation on our ANCOVA results.

Homogeneity of Variances:

The Levene's tests were utilized to test the assumption of equality of variances within the groups based on the distribution of the independent variable, the income group. However, the results of the tests were inconclusive, with p-values returned as NaN. This lack of definitive p-values presents a challenge in validating this assumption. The absence of conclusive p-values could be attributed to the large sample size or potentially to other underlying issues with the dataset, such as missing data or outliers. The failure to establish homogeneity of variances suggests that caution should be exercised when interpreting the ANCOVA results. This assumption ensures that group variances are similar, and its violation may affect the fairness of the tests.

5. Results:

For Math Scores:

For Math Scores: The analysis reveals that without interaction terms, the income groups' coefficients (0.1523 and 0.1442) suggest a modest influence on math score changes, overshadowed by fall general knowledge scores (coefficient of 0.1993). The significance of these coefficients is underscored by p-values less than 0.001, indicating a robust model. With interaction terms, a notable trend emerges higher fall general knowledge scores lead to diminished income effects on math score changes, as seen by negative coefficients (-0.0683 and -0.0945) with strong statistical significance ($p < 0.001$).

Figure 8 illustrates this interaction, showing how different income groups' influence on math score changes varies with general knowledge scores. The slopes of the lines in Figure 8 indicate the nature and degree of these interactions.

INF2178
Assignment 3
Xuying Du

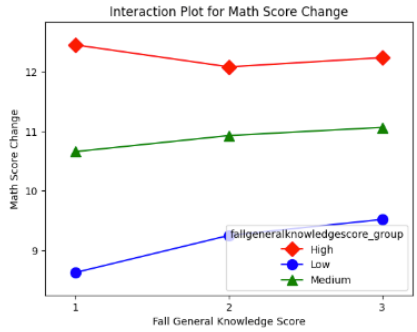

Figure 8: Interaction Plot for Math Score Change

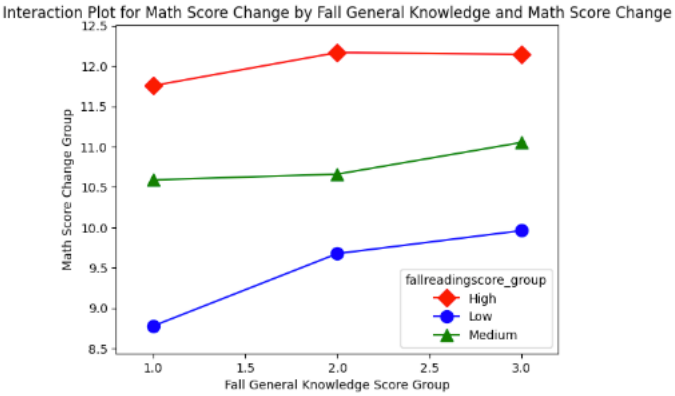

Figure 9: Interaction Plot for Math Score Change by Fall General Knowledge and Math Score Change

For Reading Scores:

In the model predicting changes in reading scores, the absence of significant interaction terms between income groups and fall reading scores resulted in the lack of an interaction plot. However, we still observed substantial effects of income groups and fall reading scores on changes in reading scores, as indicated by the coefficients for income groups (0.3751 and 0.4898) and fall reading scores (0.1322), again validated by p-values below 0.001.

For General Knowledge Scores:

Next, when exploring changes in general knowledge scores, we find that in the model controlling for changes in reading and math scores, the coefficients for income groups are 0.0835 and -0.0919, while the coefficients for changes in reading and math scores are 0.0601 and 0.0873 are positively associated with changes in general knowledge scores, a relationship substantiated by p-values less than 0.001, respectively. This suggests that controlling for changes in reading and math scores, the influence of income on changes in general knowledge scores is relatively small, while changes in reading and math scores have a positive effect on changes in general knowledge scores.

Finally, in the model with interaction terms, we observe that the interaction terms between changes in reading scores and changes in math scores affect changes in general knowledge scores. Specifically, when changes in reading scores and changes in math scores are both high, the influence of income groups 2 and 3 shows a decreasing trend, as evidenced by the coefficients of -0.0076 and -0.0076.

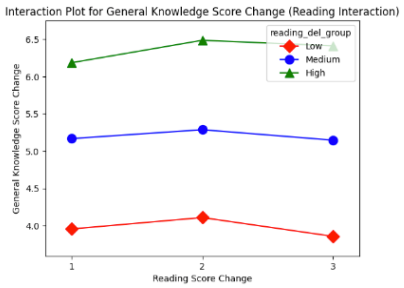

Figure 10: Interaction Plot for General Knowledge Score Change (Reading Interaction)

6. Discussion:

The analysis confirms that the significant interaction effects in the models underscore that it is hard to comprehend the influence of an income group' on academic performance without the sight of students' initial competencies and advanced in other subjects as well. Through the inclusion of this significant interaction with general knowledge scores, the difference in income brackets could compound the advantages and disadvantages gained from the initial knowledge of the students. This aligns with the claim of a positive relationship between the initial performance and intervention figure in the model that foundational reading skills could set off a trajectory for educational attainment and this was also influenced by the socio-economic variation. Full of Education policies and salvation from this, these insides are of paramount importance. It implies that strategies for closing the performance gap should have diverse approaches, focusing not just on economic inequality but also on the elements that develop critical thinking, reasoning, and knowledge in the foundational subjects.

7. Conclusion:

The investigation results raise concrete evidence on the tremendous position of income category in academic achievements increments other factors apart from the initial learner's knowledge that may mediate this relationship. There is a marked relationship that the interaction effects reflect an interrelated interplay between economic background and early academic abilities coupled with the educational advancement being made. Thus, these findings highlight the necessity of the methods of learning that are multi-aspectual, geared towards overall needs, socioeconomic specifics, and early academic development, ensuring the fairness of the educational process among the students.