INF2178 Technical Assignment 3

Name: KA YUEN LEE

Student Number: 1010073974

Professor: Shion Guha

1. Introduction

As society develops, the relationship between economic factors and academic success becomes increasingly pertinent. This report aims to explore the change in early children's scores by leveraging a dataset from a longitudinal study conducted between 1998 and 1999. The dataset, titled 'INF2178_A3_data.csv' will be used. This report aims to cover the impact of income status on the development of general knowledge and math skills over the learning period. By examining changes from fall to spring, this report seeks to offer insights into the educational differences that may exist and how they evolve within a single academic year.

There are 2 research questions that will be discussed in the following analysis:

1. How does income group influence spring semester general knowledge scores when controlling for fall semester general knowledge scores?
2. How does income group influence spring math scores when controlling for fall math scores?

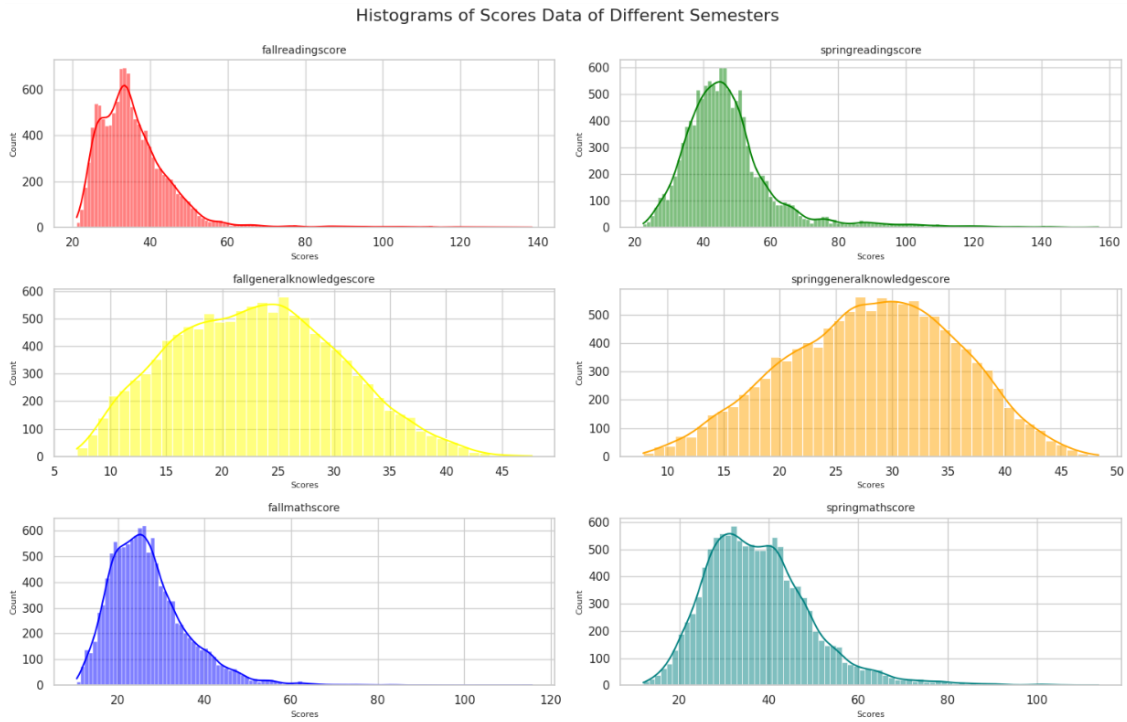
By addressing these questions, we can have a report that not only charts the impact of income status on student's academic results but also serves as a foundational analysis for policy development and long-term educational planning.

2. Data cleaning

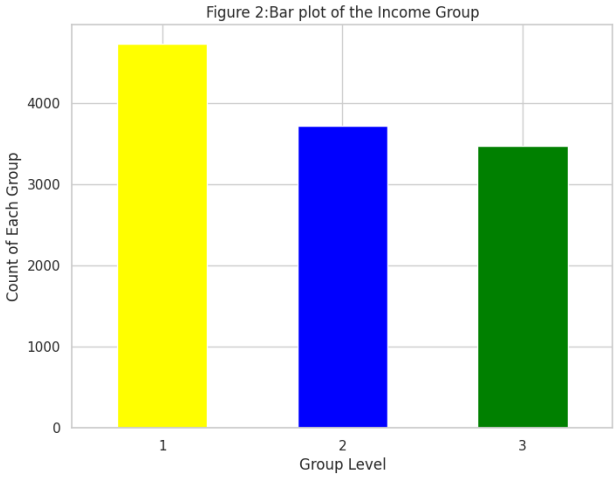
The raw dataset has a total of 9 columns with 11,933 entities (rows). The dataset comprises numerous variables and categorical variables. After initial review, the dataset is good, with a wealth of information across several dimensions, and does not have missing values. For this study, we concentrated on the following columns with short description :

- ✧ fallmathscore: Early children's math score on fall 1998
- ✧ fallgeneralknowledgescore: Early children's general knowledge score on fall 1998
- ✧ springmathscore: Early children's math score on spring 1999
- ✧ springgeneralknowledgescore: General knowledge' math score in spring 1999
- ✧ totalhouseholdincome: Total household income
- ✧ incomegroup:Group by total household income

3. Exploratory Data Analysis



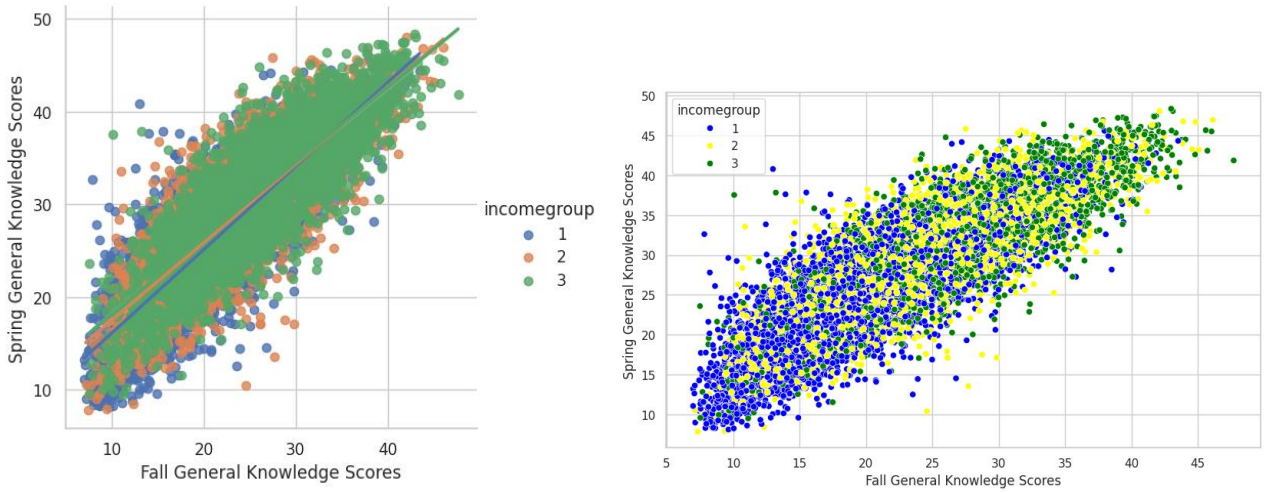
The six histograms illustrate the distribution of different scores across two semesters, indicating a roughly normal distribution for reading, general knowledge, and math scores. From the figure, it can be inferred that there's a consistency in academic performance from fall to spring, suggesting that the foundational knowledge acquired in the fall may have contributed positively to the outcomes in the spring. The consistency in the distribution shape across the semesters suggests that students who performed well in the fall term were likely to maintain their performance in the spring term.



From the Bar plot, we can find that income group 1 has the most frequency, and income group 3 has the lowest frequency.

4. ANCOVA Test

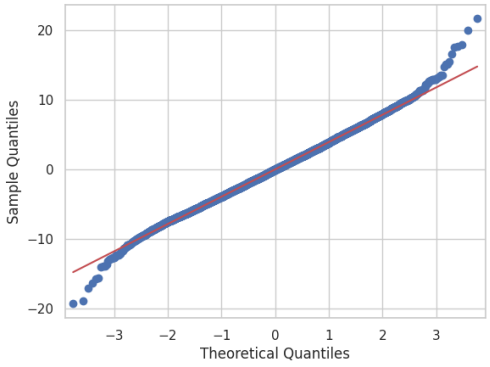
Research question 1: How does income group influence spring semester general knowledge scores when controlling for fall semester general knowledge scores?



From the above figure, shows that there appears to be a positive correlation between fall and spring scores, suggesting that higher scores in the fall are associated with higher scores in the spring across all income groups. And, the trend line indicates the general direction of this relationship, sloping upwards, which represents the positive correlation. Also, the distribution of points for income groups 2 and 3 (orange and green) is denser at the higher end of both axes compared to income group 1 (blue), which might suggest that these groups generally had higher general knowledge scores in both semesters.

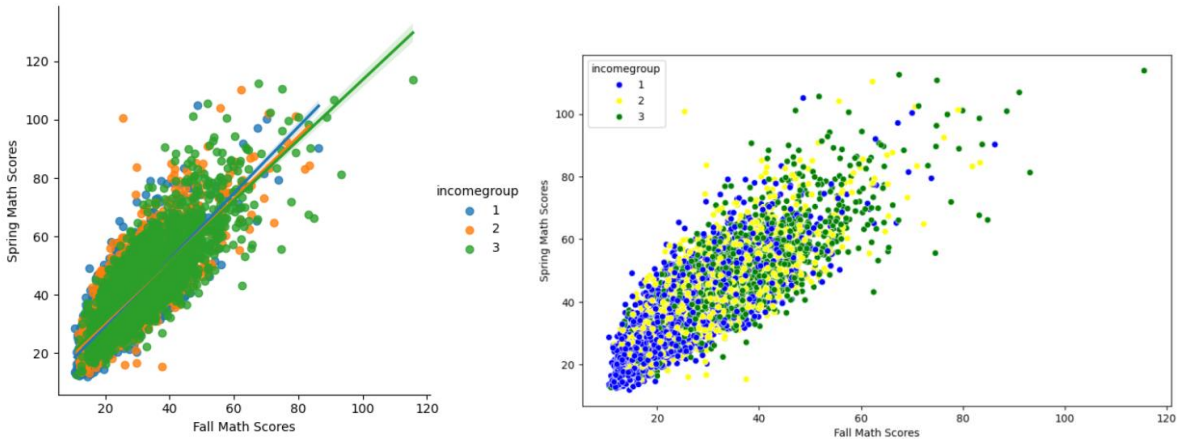
	Source	SS	DF	F	p-unc	np2
0	incomegroup	1756.904	2	56.908	2.53E-25	0.00945
1	fallgeneralknowledgescore	411876.768	1	26682.3	0.00E+00	0.69105
2	Residual	184140.18	11929	NaN	NaN	NaN

In the ANCOVA test conducted for Research Question 1. The sum of squares (SS) for the income group is 1756.904 and the F-value is 56.908. This indicates a statistically significant effect of income group on the spring general knowledge scores. Despite this, the effect size (np^2) associated with the income group is quite small (0.00945), suggesting that the practical impact of the income group, while statistically significant, may be minimal on the variance of spring general knowledge scores. The fall general knowledge score has an SS of 411876.768 with an F-value of 26682.27, which is extremely high. This strongly suggests that the fall general knowledge score is a significant predictor of the spring general knowledge scores. In summary, while the income group has a significant effect on the spring general knowledge scores, the fall general knowledge score is a much stronger predictor, explaining a substantial amount of variance in the outcome. This suggests that while the income group is statistically significant, but improving general knowledge in the fall may have a greater overall impact on spring general knowledge scores.



The QQ-plot and the Shapiro-Wilk test results indicate that the residuals from the model do not follow a normal distribution. Despite the w value of 0.998 from the Shapiro-Wilk test suggesting that the sample distribution is close to normal, the extremely small p-value (3.15e-11) signifies a statistically significant departure from normality. Additionally, Levene's test shows a p-value of 0.0001, confirming that the variances across different income groups are not equal and therefore violate the homogeneity of variance assumption required for ANCOVA. While the above analysis may indicate an effect of income group on spring general knowledge scores, these assumption violations could potentially affect the validity and reliability of these findings.

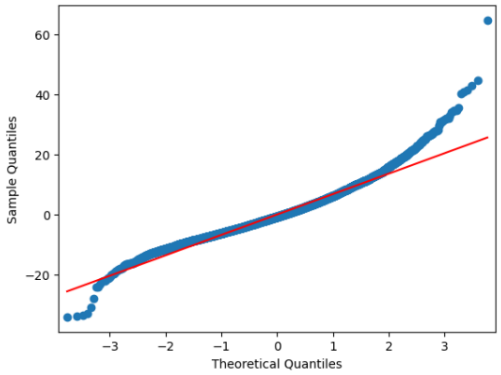
Research question 2: How does income group influence spring math scores when controlling for fall math scores?



In both figures, there is a positive correlation between fall and spring math scores, students who scored higher in the fall tended to also score higher in the spring. This relationship is represented by an upward-sloping trend line, suggesting that improvement in fall scores is predictive of better performance in the spring. Overall, these figures support the notion that there is a general trend of higher scores being associated with higher income levels.

	Source	SS	DF	F	p-unc	np2
0	incomegroup	1.71E+03	2	18.5236	9.28E-09	0.0031
1	fallmathscore	1.03E+06	1	22203.1	0.00E+00	0.65051
2	Residual	5.51E+05	11929	NaN	NaN	NaN

The ANCOVA analysis shows a sum of squares (SS) for the income group of 1.71E+03 and an F-value of 18.5236, indicating a statistically significant effect of the income group on spring math scores. Despite this, the effect size (np^2) associated with the income group is quite small (0.0031), suggesting that the practical impact of the income group, while statistically significant, may be minimal on the variance of spring math scores. In contrast, the fall math scores show an SS of 1.03E+06 and an F-value of 22203.1, which is very high. This denotes a very strong predictive relationship between fall math scores and spring math scores, with a large effect size (np^2) of 0.6505, indicating that fall math scores are a dominant predictor of spring math scores. In conclusion, both the income group and fall math scores significantly impact spring math scores. However, student's fall math scores have a greater impact on spring math scores.



The assumption checks for the ANCOVA applied to Research Question 2. The QQ-plot shows a deviation from the expected normal distribution of residuals in the ANCOVA model, particularly at the extreme ends of the distribution. The Shapiro-Wilk test corroborates this visual indication with a w value close to 1 (0.9649), which would typically suggest a distribution close to normal. However, the p-value is close to 0, indicating a statistically significant departure from normality. Furthermore, Levene's test for equality of variances across the different income groups shows a very low p-value (3.68004527698148e-09), which is less than the alpha level of 0.05. In conclusion, violations of these assumptions suggest that the results of the ANCOVA may not be reliable.

5. Conclusion

This analysis demonstrates that household income significantly influences the academic progress of kindergarten students, particularly in general knowledge and math. The study's findings indicate that after controlling for fall scores, students from higher-income groups statistically show greater improvements in both general knowledge and math scores over the academic year compared to their peers. While the ANCOVA tests confirm the impact of income on academic progress, assumption checks for Research Questions 1 and 2 reveal potential breaches in normality and homogeneity of variances. These statistical considerations warrant a cautious approach to interpreting the results and may necessitate further exploration using alternative methods. Nevertheless, the analysis underscores the crucial interaction between income and education, highlighting the importance of policies to ensure equitable learning opportunities for all children.