Analysis of the impact on kindergarten children's academic progress

1. Introduction

In recent years, the discourse surrounding educational equity has intensified. Among these factors, household income emerges as a critical determinant of educational outcomes, influencing access to resources, learning environments, and opportunities for enrichment outside the classroom. This study measures the relationship between household income group and kindergarten children's growth in academic progress in one school year (fall 1998 to spring 1999) to understand more about how socioeconomic status shapes students' trajectories of change.

This report uses the dataset named "INF2178_A3_data.csv", which contains the data of kindergarten student study information from fall 1998 - winter 1999. This report is focusing on students' reading, math, general knowledge score within this period, using analysis of covariance-ANCOVA to discover the potential factors that will affect the scores of children, our exploration will address three fundamental research questions:

Research question 1: How does income group affect childrens' reading score in spring, when controlling their reading score in fall term?

Research question 2: How does income group affect childrens' math score in spring, when controlling their math score in fall term?

Research question 3: How does income group affect childrens' general knowledge score in spring, when controlling their general knowledge score in fall term?

2. Data Cleaning

The raw data set has 11933 rows and 9 columns, and we have reduced our working data to 7 columns.

fallreadingscore: Kindergarten student's reading score in 1988 fall.

fallmathscore: Kindergarten student's math score in 1988 fall.

fallgeneralknowledgescore: Kindergarten student's general knowledge score in 1988 fall.

springreadingscore: Kindergarten student's reading score in 1999 Spring.

springmathscore: Kindergarten student's math score in 1999 Spring.

springgeneralknowledgescore: Kindergarten student's general knowledge score in 1999 spring.

incomegroup: household income level.

Now we have 11933 rows and 7 columns in our data set. After check, there is no missing value in this data set, I also change the datatype of Incomegroup variable from integer to category.

3. EDA

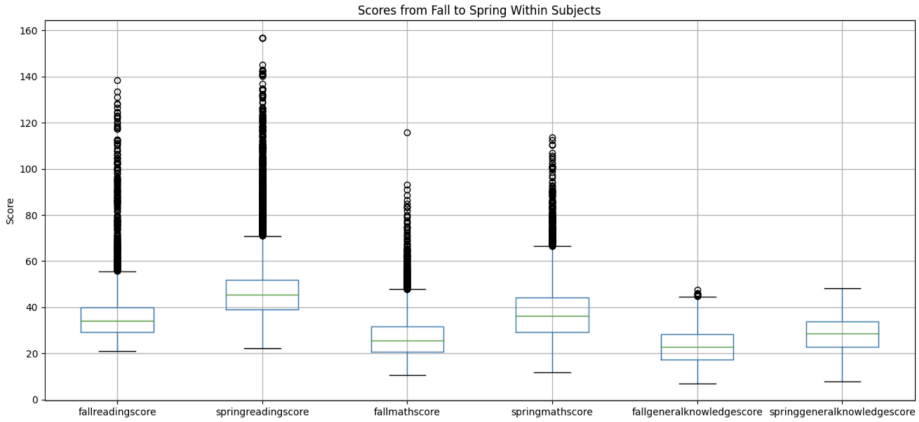
Over all summary statistics:

index	fallreadingscore	fallmathscore	fallgeneralknowledgescore	springreadingscore	springmathscore	springgeneralknowledgescore
count	11933.0	11933.0	11933.0	11933.0	11933.0	11933.0
mean	35.95	27.13	23.07	47.51	37.80	28.24
std	10.47	9.12	7.40	14.33	12.03	7.58
min	21.01	10.51	6.99	22.35	11.9	7.86

25%	29.34	20.68	17.39	38.95	29.27	22.80
50%	34.06	25.68	22.95	45.32	36.41	28.58
75%	39.89	31.59	28.32	51.77	44.22	33.78
max	138.51	115.65	47.69	156.85	113.8	48.35

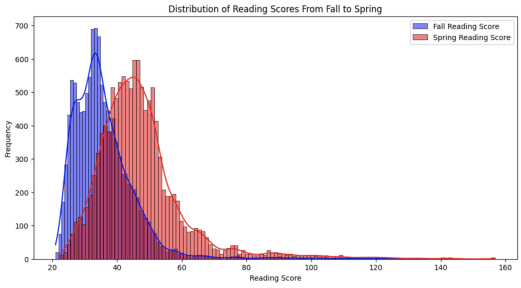
This table presents a summary of students' academic performance across three subjects—reading, math, and general knowledge—comparing fall and spring terms. The mean scores in all subjects show an increase from fall to spring, indicating overall progress. The reading scores have the highest averages and a wide range of scores, suggesting a great variability in students' reading abilities. General knowledge scores, with the lowest on average, but still follow a trend of improvement from fall to spring. The variability in student's math score is still obvious, the mean in math score is second highest in these subjects in both fall and spring.

Now, create a boxplot displays a comparison of students' scores in reading, math, and general knowledge from the fall term to the spring term.



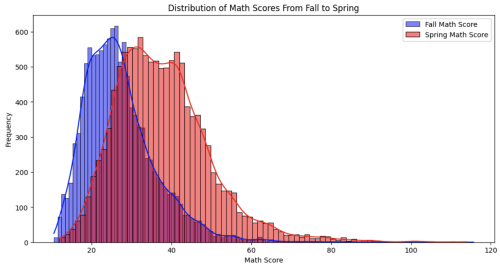
The boxplot shows a comparison of student scores for reading, math, and general knowledge from fall to spring. Obviously, there's an upward shift in median scores from fall to spring in all subjects, indicating general academic improvement over time. Reading median scores with a notable increase in the spring. Math and general knowledge scores also show an upward trend, although the median spring math and general knowledge scores exhibit a less rise.

Now let's see histogram distribution of these three subjects from fall to spring:

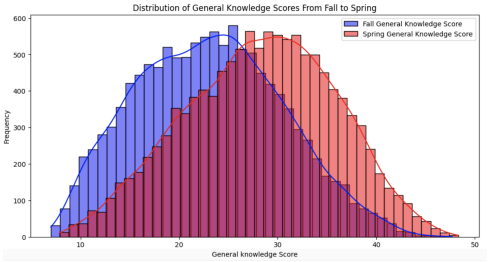


This histogram compares the frequency of fall and spring reading scores for students. The fall reading scores, generally cluster around lower values, while the spring scores, shift towards

higher values, indicating improvement. Both distributions appear normally shaped, suggesting a consistent progression in reading ability across the student.



This histogram compares the distribution of math scores from fall to spring. Both distributions have a similar bell-shaped curve, but the spring scores are shifted to the right, indicating overall higher performance in the spring term. This suggests that students are generally improving in math as the academic year progresses.

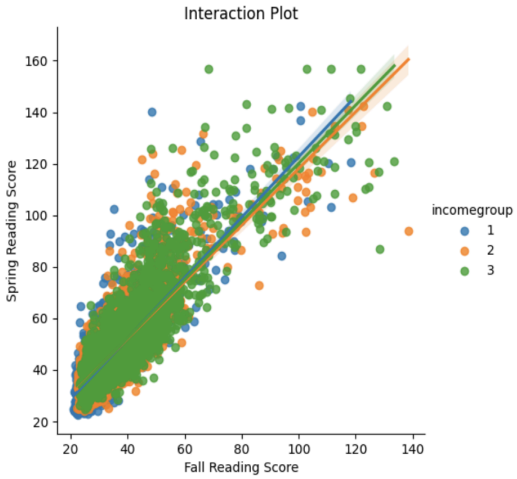


This histogram shows the distribution of general knowledge scores for students from fall to spring. Both distributions are similar in shape, suggesting consistent performance in general knowledge over time. However, the spring scores have a slight shift to the right, indicating an overall increase in scores when comparing the end of the school year to the beginning

4. ANCOVA Analysis

In order to discover our three research questions, we will contact the ANCOVA method:

Now we focus on the first research question: How does income group affect childrens' reading score in spring, when controlling their reading score in fall term?



This interaction plot visualizes spring reading scores against fall reading scores, by income group. The general upward trend suggests that higher fall scores are associated with higher spring scores across all income groups. Spread of points suggests that children from the highest income group general supports the greater improvement of score, the higher income family may experience a higher enhancement of their child’s reading score from fall to spring.

	F-statistic	P-value	np2
Income Group	4.056	0.017	0.001
Fall reading score	24455.398	<0.001	0.672

Result: For the income group: the F-statistics is 4.056 and the p value is 0.017, which shows there is a statistically significant difference in reading scores between different income groups, however, the np2 is really smaller, about 0.001, means the impact of income for reading scores might be minimal.

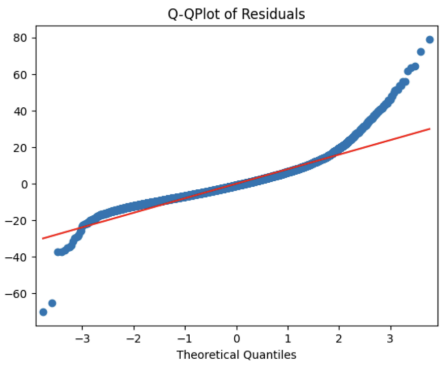
Fall reading score: The F-statistics is really high, a p value<0.001, indicate that fall reading score has a strong statistically effect on the dependent variable, the np2 is about 0.672, the large value of np2 suggests that fall reading score is the major factor for the outcome variable.

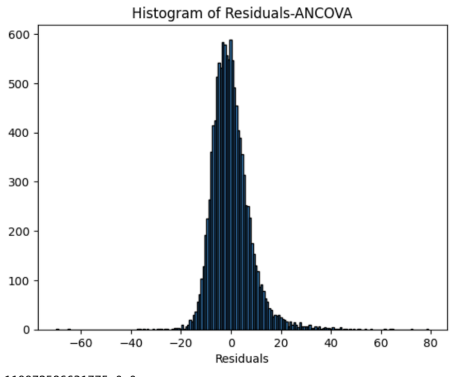
ANCOVA Model Result:

	Coef	P>|t|
Intercept	6.543	<0.001
C(incomegroup)[T.2]	0.375	0.033
C(incomegroup)[T.3]	0.490	0.008
fallreadingscore	1.132	<0.001

The result shows that both the second(with 0.375 coef) and third income group(with 0.490 coef) have a statistically significant effect on improvement in reading scores from fall to spring, this is also supported by the p value which is smaller than 0.05. The fall reading score has a strong positive association(coef 1.132) with the spring reading score, this relationship is highly significant by p value <0.001. Overall, both income group and fall reading score are important predictors of spring reading score improvement.

Assumption check:

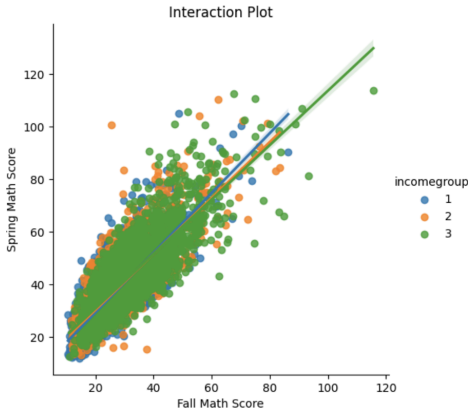




	statistic	P value
Shapiro test	0.912	<0.001
Levene test	39.553	<0.001

QQ plot(some points have deviation from fit line) and histogram of ANCOVA residuals suggest that the normality assumption might be violated, also supported by Shapiro test-p value < 0.001, Homogeneity of variance is also violated, shown by the levene's test p value < 0.001.

For second research question: How does income group affect childrens' math score in spring, when controlling their math score in fall term?



This interaction plot visualizes spring math scores against fall math scores by income group. Spread of points shows that student with the highest household income group general supports the greater improvement of math score, the higher income family may experience a higher enhancement of their child's math score from fall to spring.

	F-statistic	P-value	np2
Income Group	18.524	9.285e-09	0.003
Fall math score	22203.081	<0.001	0.651

For the income group: the p value is <0.001, which shows there is a statistically significant difference in math scores between different income groups, however, the np2 is only 0.003, showing the minimal impact of income for reading scores.

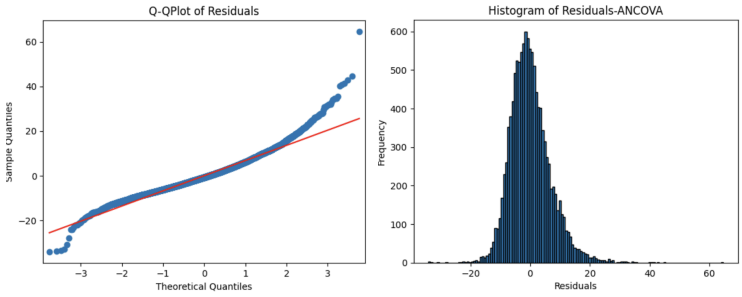
Fall math score: The F-statistics is really high, a p value<0.001, there is a strong statistical effect of fall math score on dependent variables, the np2 has a large value which indicates fall math score is a major factor.

ANCOVA Model Result:

	Coef	P>|t|
Intercept	8.201	<0.001
C(incomegroup)[T.2]	0.670	<0.001
C(incomegroup)[T.3]	0.920	<0.001
fallmathscore	1.074	<0.001

The results show that both income groups 2 and 3 have significant effects on increasing math scores(supported by positive coefficient and small p value). However, the fall math score is a significant major predictor of spring math score as the high coefficient value and p value <0.001.

Assumption check:



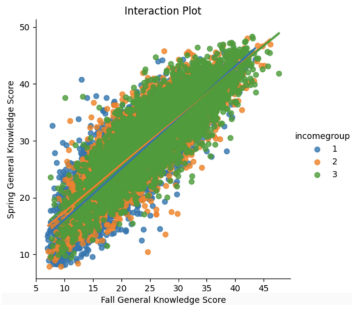
Test	statistic	P value
Shapiro	0.965	<0.001
Levene	18.900	<0.001

Deviation of points of QQ plot and histogram of ANCOVA residuals suggest that the normality assumption might be violated, also supported by Shapiro test-p value < 0.001.

By the levene's test p value < 0.001, therefore homogeneity of variance is also violated,.

Now last research: How does income group affect childrens' general knowledge score in

spring, when controlling their general knowledge score in fall term?



Similar with previous two plots, children with the higher household income group shows the larger increasing of general knowledge scores, and a stronger positive correlation between score with fall and spring.

	F-statistic	P-value	np2
Income Group	56.908	2.535e-25	0.009
Fall general knowledge score	26682.270	<0.001	0.691

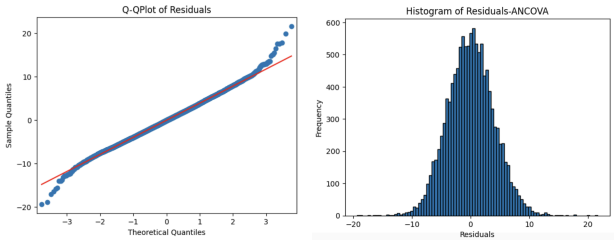
It indicates that there is a statistically significant difference in reading scores between different income groups, supported by F-statistic and small p value. Fall general knowledge has a stronger statistical effect on spring score, is also the major predictor for outcome variable. We can tell this by checking the high F, small p value(<0.001), and the large np2 value-0.691.

ANCOVA Model Result:

	Coef	P>|t|
Intercept	8.030	<0.001
C(incomegroup)[T.2]	0.708	<0.001
C(incomegroup)[T.3]	0.942	<0.001
fallgeneralknowledgescore	0.854	<0.001

It has indicated that children in higher household income group tends to have greater improvement on general knowledge score from fall 1998 to spring, both income 2,3 has significant effects on score improvement(p value <0.002 and positive coefficient), fall general Knowledge score has a significant major effect on spring math score as the high positive coefficient and small p value.

Assumption check:



Test	statistic	P value
Shapiro	0.998	3.155e-11
Levene	9.406	<0.001

Similarly with previous, based on the qqplot, histogram and the shaprio/Levene test, we have concluded that normality and homogeneity of variance assumption are violated.

5. Conclusion

We have concluded that socioeconomic status-income groups play a crucial role in kindergarten student's academic performance. This study demonstrates that children from higher household income groups will have a significant improvement on their reading, math, general knowledge scores, however children's individual performance has a larger effect on score improvement. For the assumption check, ANCOVA test does not meet the normality and Homogeneity of variance, therefore we need to pay more attention when interpreting the result and be careful.