Exploring the Impact of Household Income on Kindergarten Students' Academic Progress

Songjia Liu 1005137503

1. Introduction

In recent years, the conversation around educational equity has gained momentum. Among these, household income stands out as a key determinant, which may affect access to educational resources, the quality of learning environments, and the availability of enrichment opportunities outside the classroom. This study delves into the relationship between household income and the academic progress of kindergarten students over one school year, from fall 1998 to spring 1999. By conducting EDA and one-way ANCOVAs, the study aims to uncover deeper insights into how economic conditions influences disparities in academic outcomes among young learners.

2. Data cleaning and warning

a. Select the necessary columns

The raw dataset has a total of 9 columns and 11933 rows, the columns I need are:

- "fallreadingscore", "fallmathscore", "fallgeneralknowledgescore":

These columns represent the scores of students in reading, math, and general knowledge at the beginning of the study period (Fall 1998).

- "springreadingscore", "springmathscore", "springgeneralknowledgescore":

These columns contain the scores of students in the same subjects at the end of the study period (Spring 1999), allowing for the measurement of academic growth.

- "incomegroup":

This column categorizes households into different income groups, facilitating analysis based on socioeconomic strata.

b. Check whether there are missing values or outliers

There is no missing value because the count of nulls in all columns is zero.

c. Change data type as necessary

Change incomegroup's data type from "int" to "categorical" for One-Way ANCOVA ensures the statistical analysis appropriately handles the distinct, non-ordered income groups, thereby accurately assessing their impact on children's academic progress without implying any numerical or ordinal relationship between the groups.

3. Research questions

Research Question 1 (using One-Way ANCOVA):

How does income group impact childrens' **reading score** in the spring term, after controlling their **reading score** in the fall term?

Research Question 2 (using One-Way ANCOVA):

How does income group impact childrens' **math score** in the spring term, after controlling their **math score** in the fall term?

Research Question 3 (using One-Way ANCOVA):

How does income group impact childrens' **genernal knowledge score** in the spring term, after controlling their **general knowledge score** in the fall term?

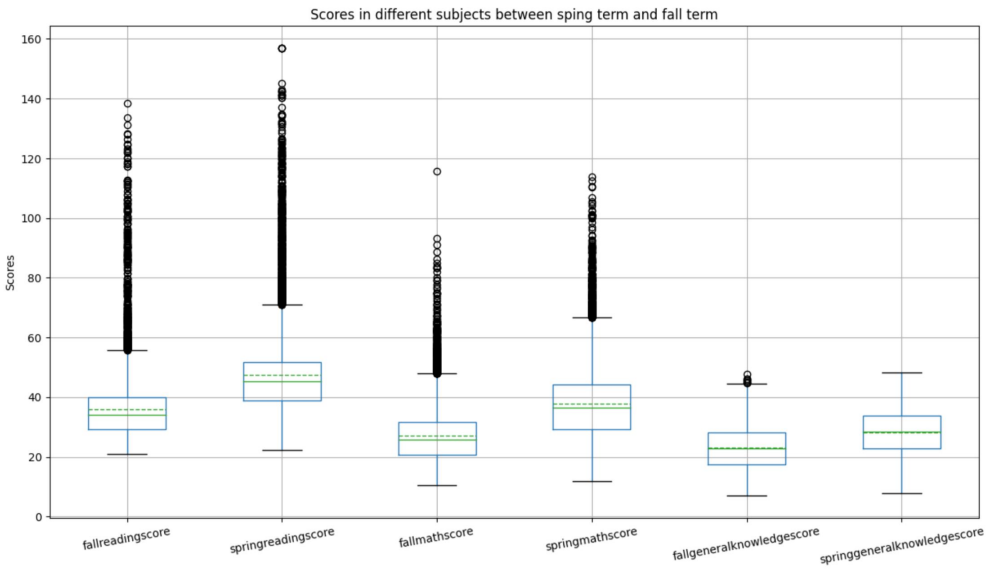
4.Exploratory Data Analysis (EDA):

a. Statistics summary of columns

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	14.32	12.03	7.58
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

Interpretation: The summary statistics for the academic scores of kindergarten children indicate improvements from fall to spring across reading, math, and general knowledge. Scores increased, reflecting academic growth over the school year, with reading scores showing the most substantial rise. The spread of scores, as indicated by the standard deviation, suggests variability in academic progress among students.

b. Boxplot for the scores in different subjects between sping term and fall term

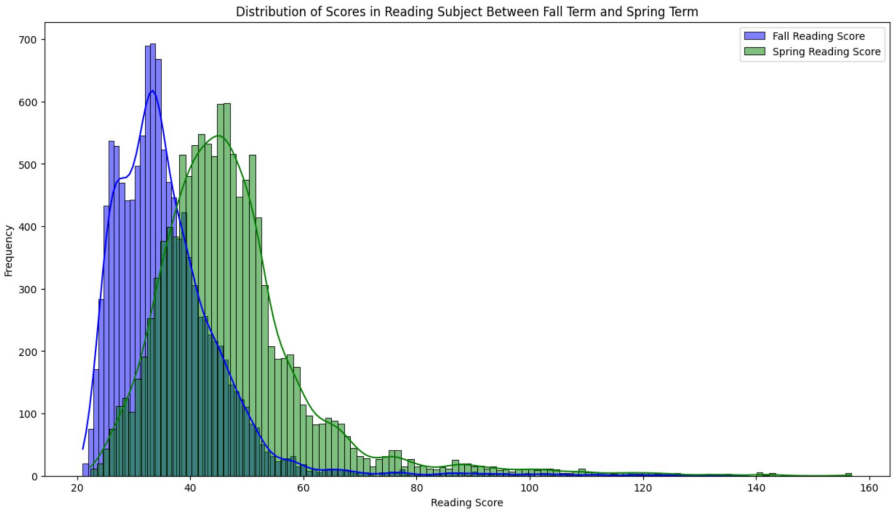


Interpretation:

From fall to spring, median scores rose across all subjects, with reading showing the most significant increase.

The spread of scores, as shown by the interquartile range, also expanded, particularly in reading, indicating a greater variability in student achievement as the year progressed.

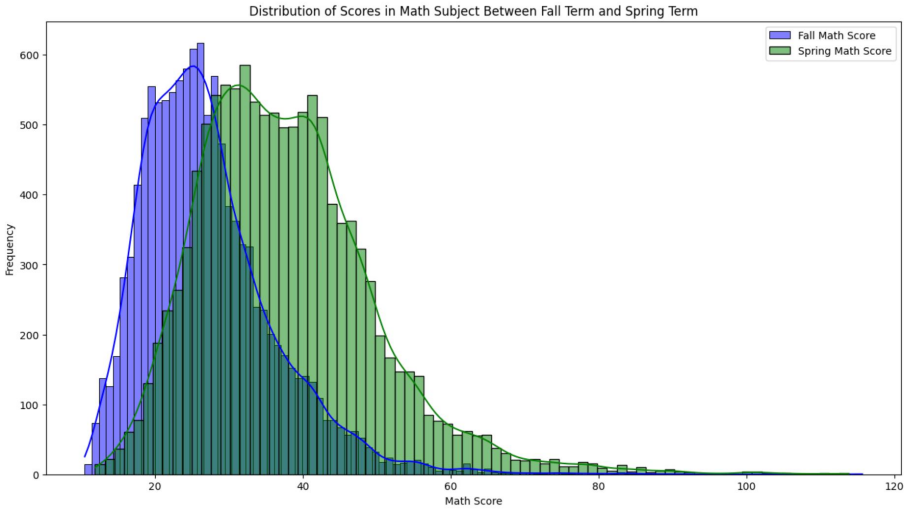
Outliers are present in all subjects, with some students scoring significantly higher than their peers.



Interpretation:

The distribution in spring is right-shifted compared to fall, showing an improvement in reading scores.

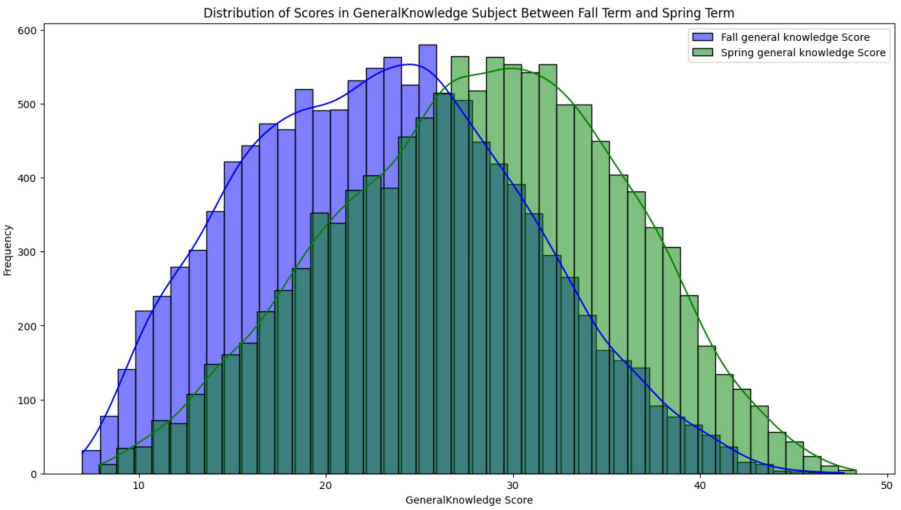
The shape of the distribution suggests that while the majority of students improved, a significant number achieved remarkably higher scores, possibly indicating effective reading interventions or developmental gains in literacy.



Interpretation:

Similarly, spring math scores have improved, as evidenced by the rightward shift. The spring distribution appears slightly more spread out, indicating increased variability in student performance.

The long tail to the right suggests some students have achieved scores much higher than the mean, which could be indicative of a greater range of mastery in math skills among students.



Interpretation:

The general knowledge scores exhibit a rightward shift as well, with a notable increase in the mode from fall to spring.

The distribution appears to maintain a similar shape, suggesting consistent improvement across students.

c. Boxplot for the scores in different subjects between sping term and fall term

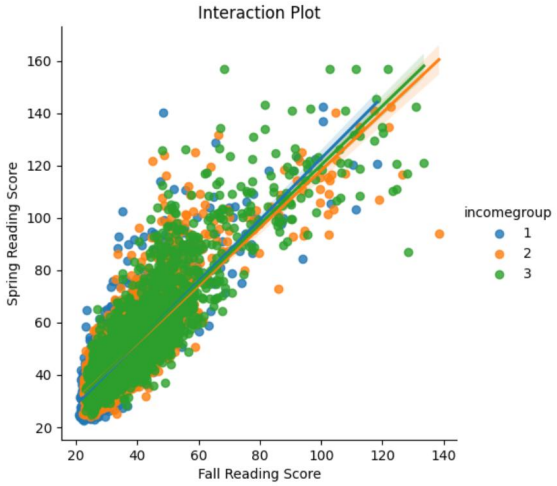
Comparison: Comparing across subjects, the improvement in reading scores appears to be the most pronounced, with math and general knowledge showing notable but less dramatic increases. The consistency of the shape in general knowledge scores indicates a more uniform improvement among students in that area. The increased spread in math and reading could imply a greater divergence in students' abilities or possibly reflect the impact of curriculum and instruction varying across these subjects.

4.One-Way ANCOVA Analysis

Research Question 1:

How does income group impact childrens' reading score in the spring term, after controlling their reading score in the fall term?

a. Interaction Plot



Interpretation:

The interaction plot reveals a positive correlation between fall and spring reading scores across all income groups, with the degree of improvement being associated with income levels. The gentle upward trajectory for Group 1 implies more moderate progress in reading scores from fall to spring. In contrast, Group 2 displays a noticeably steeper ascent, indicative of more pronounced gains. Group 3, representing the highest income category, shows a similarly sharp increase, signifying significant advancements in reading proficiency. This variation in slopes suggests that a student's socioeconomic status could play a role in their academic development, particularly in reading, over the course of the school year.

b. Ancova Result Table

index	Source	SS	DF	F	p-unc	np2
0	incomegroup	513.12	2	4.06	0.02	0.0007
1	fallreadingscore	1547042.33	1	24455.40	0.0	0.6721
2	Residual	754625.55	11929	NaN	NaN	NaN

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.5430	0.264	24.779	0.000	6.025	7.061
C(incomegroup) [T.2]	0.3751	0.176	2.130	0.033	0.030	0.720
C(incomegroup) [T.3]	0.4898	0.185	2.648	0.008	0.127	0.852
fallreadingscore	1.1322	0.007	156.382	0.000	1.118	1.146

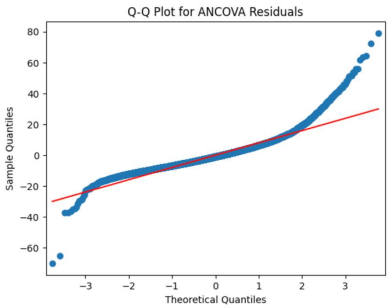
Interpretation:

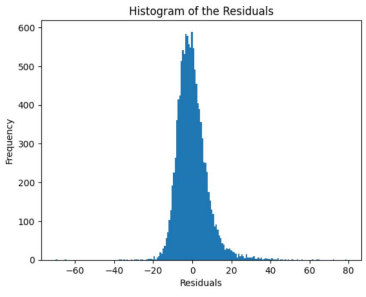
The ANCOVA results indicate that after controlling for fall reading scores, there is a statistically significant impact of income group on children's reading scores in the spring term. The coefficients for income groups 2 and 3 are 0.3751 and 0.4898, respectively, both with p-values below 0.05, suggesting that children from higher income groups have higher spring reading scores compared to the reference group (presumably group 1). The large F value (24455.40) for the fall reading score and its low p-value confirm that the fall reading score is a strong predictor of the spring reading score.

The eta squared value for the income group, which is 0.0007, indicates a very small effect size. This means that while the income group differences are statistically significant, their actual impact on the reading scores is small in practical terms.

The model explains a substantial amount of variance in spring reading scores, as indicated by the eta squared value for the fall reading score (0.6721), which is a large effect size. This suggests that while income group does have a significant effect, the fall reading score is a much stronger predictor of the spring reading score.

c. Ancova Assumption Check





index	Parameter	Value
0	Test statistics (W)	39.5528
1	Degrees of freedom (Df)	2.0
2	p value	0.0

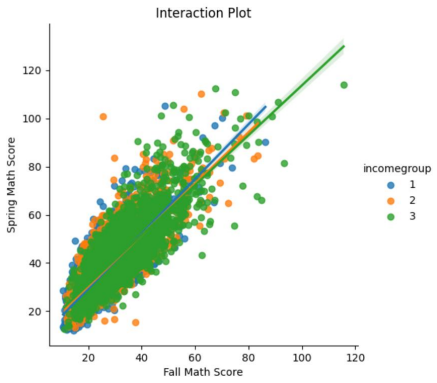
Normality of residuals: The Q-Q plot and histogram for the ANCOVA residuals, shows that the assumption of normality may not hold, because it is indicated by the deviation from the line in the Q-Q plot and the p-value <0.01 in the test.

Homogeneity of variances: Levene's test result with a p-value less than 0.01 shows the variances are not equal across groups, which violate the homogeneity of variance assumption for ANCOVA.

Research Question 2:

How does income group impact childrens' math score in the spring term, after controlling their math score in the fall term?

a. Interaction Plot



Interpretation:
The interaction plot for math scores suggests a positive relationship between fall and spring scores across all income groups. The slope of the lines for each income group appears to be similar, indicating that the increase in math scores from fall to spring does not differ dramatically between income groups, especially when fall scores are controlled for. This could imply that while income may influence math scores, it does not significantly impact the rate of improvement over the school year once initial math ability is accounted for.

b. Ancova Result Table

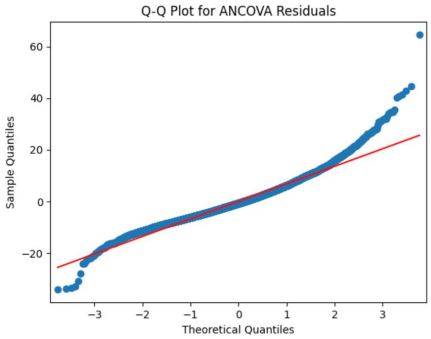
index	Source	SS	DF	F	p-unc	np2
0	incomegroup	1712.76	2	18.52	9.28e-09	0.003
1	fallmathscore	1026488.97	1	22203.08	0.0	0.65
2	Residual	551499.44	11929	NaN	NaN	NaN

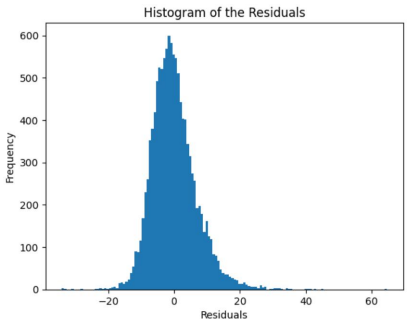
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.2011	0.199	41.273	0.000	7.812	8.591
C(incomegroup) [T.2]	0.6700	0.151	4.430	0.000	0.374	0.966
C(incomegroup) [T.3]	0.9199	0.160	5.741	0.000	0.606	1.234
fallreadingscore	1.0735	0.007	149.007	0.000	1.059	1.088

Interpretation:
The ANCOVA results indicate that income group has a statistically significant impact on children's spring math scores after controlling for fall math scores. With F values indicating strong effects and p-values less than 0.05, children from higher income groups (Groups 2 and 3) demonstrate higher spring math scores compared to the reference group (Group 1). The effect sizes, represented by the coefficients (0.6700 for Group 2 and 0.9199 for Group 3), are substantial, especially for Group 3. This suggests that the math score improvements are significantly associated with higher income levels.

c. Ancova Assumption Check

The Q-Q plot, Histogram of residuals, and the Shapiro-Wilk test with P < 0.001 shows that the assumption of normality may not hold. Levene's test result is that p-value < 0.01 shows it violates the homogeneity of variance assumption for ANCOVA.



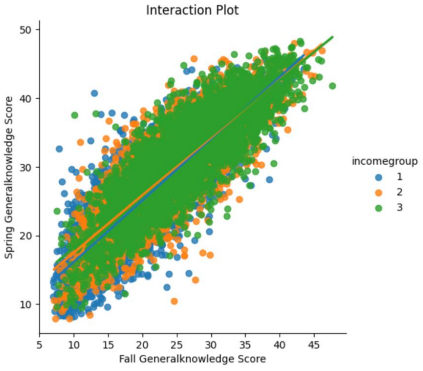


index	Parameter	Value
0	Test statistics (W)	18.90
1	Degrees of freedom (Df)	2.0
2	p value	0.0

Research Question 3:

How does income group impact childrens' generalknowledge score in the spring term, after controlling their generalknowledge score in thefall term?

a.Interaction Plot



Interpretation:
The interaction plot suggests a positive correlation between fall and spring general knowledge scores across all income groups. There's a visible trend indicating that children from higher income groups 2 and 3 tend to have steeper slopes, implying that they may have larger gains in general knowledge scores from fall to spring compared to the lowest income group 1. This pattern reflects the potential influence of socioeconomic status on the development of children's general knowledge over the school year.

b. Ancova Result Table

index	Source	SS	DF	F	p-unc	np2
0	incomegroup	1756.90	2	56.91	2.53e-25	0.009
1	fallgeneralknowledgescore	411876.77	1	26682.27	0.0	0.69
2	Residual	184140.18	11929	NaN	NaN	NaN

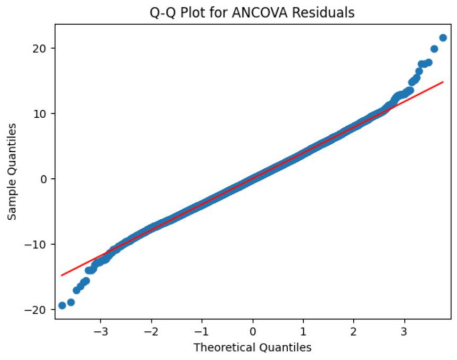
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.0303	0.119	67.519	0.000	7.797	8.263
C(incomegroup) [T.2]	0.7084	0.088	8.005	0.000	0.535	0.882
C(incomegroup) [T.3]	0.9424	0.094	10.013	0.000	0.758	1.127
fallreadingscore	0.8542	0.005	163.347	0.000	0.844	0.864

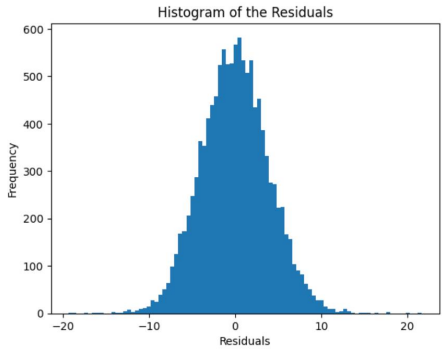
Interpretation:
The ANCOVA results indicate a significant effect of income group on children's general knowledge scores in the spring term, after controlling for their scores in the fall term. Both income groups 2 and 3 show significant positive coefficients (0.7084 and 0.9424 respectively), with p-values less than 0.001, indicating higher general knowledge scores in the spring term for these groups compared to the reference income group (1). The large F statistic for fall general knowledge scores highlights it as a strong predictor of spring scores. The eta squared values suggest a small to moderate effect size for income group on spring scores, while the fall scores have a large effect size on spring scores.

c.Ancova Assumption Check

Similarly with previous, based on Q-Q plot, Histogram of residuals and Levene's test, both assumption of normality and homogeneity of variance are violated.

index	Parameter	Value
0	Test statistics (W)	9.41
1	Degrees of freedom (Df)	2.0
2	p value	0.0





5. Conclusion:

This investigation into the academic progress of kindergarten students over a school year underscores the significant role of socioeconomic status, as delineated by income groups, in influencing academic outcomes. The analysis demonstrates that higher household income brackets are associated with marked improvements in reading, math, and general knowledge scores. Yet, it is individual performance during the fall term that most powerfully predicts spring term achievement, overshadowing the impact of socioeconomic status.

Despite the clarity of these trends, this study's findings are tempered by limitations in the underlying assumptions required for the ANCOVA analysis. Specifically, the data did not meet the criteria for normality and homogeneity of variance, as established by the Shapiro-Wilk and Levene's tests. This calls for a cautious interpretation of the results and signals the need for further methodological considerations in future research. Alternative statistical methods or data transformation techniques may provide more reliable insights and should be considered in ongoing efforts to understand and address the complexities of educational equity.