Yan Yan

1005387582

An Analysis of Score Changes Across Income Groups

Introduction

This narrative explores a dataset focused on educational metrics, particularly reading score change, math score change, and general knowledge score change. The dataset includes variables like income group and reading scores collected over different time points, allowing for an in-depth analysis of how these factors interact and influence student performance. The primary statistical method employed in this study was Analysis of Covariance (ANCOVA), a powerful tool for assessing group differences while controlling for covariates. The ANCOVA model was specified with the dependent variables being changes in reading, math, and general knowledge scores from fall to spring, the independent variable being income group membership, and the covariate being household income.

Research Questions

- How do reading, math, and general knowledge scores change over time?
- Is there a significant difference in score changes across different income groups?
- Does household income have a significant effect on the change in scores when controlling for income groups?

Visual Assessment for ANCOVA Assumptions

- For reading, math, and general knowledge score changes in the dataset, the Q-Q plots displayed a generally linear pattern. Most data points fell on or near the 45-degree reference line, which indicates that the distribution of the residuals closely follows a normal distribution. However, some deviations were noticed, particularly at the ends of the distribution, and these deviations might represent outliers or extreme values that do not conform perfectly to the expected distribution. Nonetheless, the overall trend suggested that the assumption of normality was not seriously violated. Upon visual inspection, the histograms for the residuals of the score changes were bell-shaped and symmetrical about the means. This symmetry is indicative of a normal distribution. The peaks of the histograms were well-centered, and there was an absence of significant skewness or kurtosis.

OLS Regression Result (reading score change)

The R-squared value is 0.023, suggesting that approximately 2.3% of the variance in reading score changes is explained by the model, which includes income group and fall general knowledge scores as predictors. This is relatively low, indicating that there may be other factors not included in the model that explain the changes in reading scores. The intercept (7.5314) represents the estimated mean reading score change for the baseline category of the income group when the fall general knowledge score is zero. The fall general knowledge score coefficient (0.1578) suggests that for each one-unit increase in the fall general knowledge score, there is an associated increase in the reading score change by approximately 0.158 units, holding income group constant. The p-values for the model coefficients (including the intercept) are less than 0.05, which implies that these variables have a statistically significant relationship with reading score changes. The F-statistic and the p-value relate to the hypothesis test concerning whether there is a difference in reading score changes across income groups, controlling for general knowledge scores. A p-value greater than 0.05 suggests that the null hypothesis (no difference in group means) cannot be rejected.

Reading score change	income group	n	mean	standard deviation
	1	4729	10.88	7.46
	2	3726	11.72	7.91
	3	3478	12.31	8.98
	R squared	0.023		
	F-statistic	143.2		
	intercept	7.5314		
	coefficient	0.1578		

OLS Regression Result (math score change)

The provided outputs from ANCOVA and OLS regression analyses offer insights into the relationship between income groups and math score changes, after adjusting for the covariate of fall general knowledge scores. With an R-squared value of 0.048, the model suggests that income group and initial knowledge only explain a small fraction of the variability in math score changes. Interestingly, the coefficients for the income groups (T.2 and T.3) were not statistically significant, indicating that, within the bounds of this analysis, the income group alone does not have a discernible impact on the math score progression. The positive coefficient for the fall general knowledge score (0.1993) suggests a positive relationship where students with higher initial knowledge scores may see the same level of improvement in math scores as their peers. Students with high knowledge scores have less room to improve, or it may reflect differences in the opportunities for advancement available to students at different knowledge levels.

Although the ANCOVA analysis yielded an F-statistic that approached significance (p = 0.076246), it did not firmly establish a statistically significant difference in math score changes across income groups when controlling for prior knowledge. This marginally significant result may indicate a subtle trend that could be clarified with a more robust sample or additional variables. The findings thus call for careful consideration of other potential influences on math score changes, such as educational resources, school quality, or learning environments, which may interact with income level and prior knowledge in complex ways.

Reading score change	income group	n	mean	standard deviation
	1	4729	10.88	7.46
	2	3726	11.72	7.91
	3	3478	12.31	8.97
	R squared	0.048		
	F-statistic	200		
	intercept	5.9826		
	coefficient	0.1993		

OLS Regression Result (general knowledge score changes)

The OLS regression model indicates a significant negative relationship between fall general knowledge scores and the change in general knowledge scores over time, with a coefficient of -0.1451 ($p < 0.001$). This could suggest that students with higher initial knowledge scores may have less room for improvement, or it could point to a potential ceiling effect where increments in knowledge become harder to achieve as students' initial knowledge base expands. The ANCOVA results further shed light on the impact of income on general knowledge score changes. While the model explains a relatively high proportion of the variance (R-squared = 0.061), indicating a better fit compared to the math score change model, the effect of income group on general knowledge score changes is not statistically significant at the conventional 0.05 level. The F-statistic of 2.25904 and the corresponding p-value suggest that there is a trend towards significance, hinting at a potential effect of income on general knowledge score changes that might be more apparent with a larger sample size or additional covariates. This lack of statistical significance indicates that while there may be observable differences in how general knowledge scores change across income groups, these differences are not strong enough to rule out the possibility that they occurred by chance.

Reading score change	income group	n	mean	standard deviation
	1	4729	5.12	4.08
	2	3726	5.26	4.06
	3	3478	5.12	4.03
	R squared	0.061		
	F-statistic	385.4		
	intercept	7.6		
	coefficient	-0.1451		

Interaction plot interpretation for reading score changes

The interaction plot for reading score change shows three distinct trajectories corresponding to three income groups. For income group 1, represented by the red line, reading scores seem to decline over time from fall to spring. Income group 2, shown with the green line, displays stability in reading scores across the two time points. Finally, income group 3, depicted by the blue line, indicates an slightly decrease in reading scores from fall to spring. The non-parallel lines suggest that the relationship between time and reading score changes vary by income group, implying an interaction between income level and time in their effect on reading score changes.

Interaction plot interpretation for math score changes

In this plot, income group 1, shown with the red line, experiences a decline in math scores over the school year, suggesting that students from this group may face challenges that hinder their academic progress in mathematics. Conversely, the blue line representing income group 3 rises, illustrating an improvement in math scores that could be attributed to factors like access to better resources or more support systems. Income group 2, represented by the green line, maintains a relatively stable score, indicating neither significant improvement nor declines. The non-parallel lines imply that the effect of time on math score changes is not consistent across income groups, indicating a potential interaction effect.

Interaction plot interpretation for general knowledge score changes

The interaction plot for general knowledge score change illustrates the disparities in educational progress across three different income groups over a school year. Income group 2, depicted by the green line, shows a notable upward trend, suggesting a substantial improvement in general knowledge from fall to spring. This could imply that students in this middle-income group are benefiting from effective educational practices or resources conducive to learning. In contrast, income group 3, represented by the blue line, experiences a decline in general knowledge scores. This downward trend might point to systemic issues or barriers that could be hindering the educational growth of

students from this higher-income group. Income group 1, denoted by the red line, shows a declining trend in general knowledge scores from the fall to the spring, which could point to educational disparities or gaps in the support structure that disproportionately affect this group. It raises questions about whether these students have equal access to quality educational resources or if there are other external factors.

EDA analysis

Students from income group 1 exhibit a mean reading score change of approximately 10.87, which suggests a positive trajectory in reading proficiency over the academic year. However, the considerable standard deviation of about 7.46 within this group points to a wide range of individual student experiences, indicating that while some students may have made significant strides, others have improved less or not at all. In contrast, the highest income group (income group 3) shows an even higher mean reading score change of around 12.30, but with a comparable standard deviation of approximately 8.97. This larger variation might reflect a diversity of educational experiences and outcomes even among those with higher socioeconomic status. For income group 1, the mean general knowledge score changes to approximately 5.12, with a relatively tight standard deviation of about 4.08, indicating less variability in score changes within this group. Income group 2 shows a slightly higher mean change of around 5.25 and a modestly higher standard deviation of approximately 4.06, suggesting a similar level of variability in their score changes. Income group 3 has a mean change of approximately 5.12 and standard deviation of roughly 4.03. The statistics suggest that while all groups, on average, experience an increase in general knowledge over time, there is a noteworthy spread within each group's score changes. The largest spread within income group 2 could imply a disparity in educational experiences or access to resources that facilitate knowledge gain.

Implications

The analysis, through ANCOVA, provides initial answers to the research questions. It shows that reading, math, and general knowledge scores do change over time, but these changes are not uniformly experienced across income groups. The discovery that household income isn't a significant predictor of score changes, once income group classifications are accounted for. It indicates that factors beyond mere income might play a more crucial role in the fluctuation of educational achievements over a period. Such factors could include the education, availability of academic resources, the extent of parental engagement, or participation in activities beyond the classroom. This suggests that the nuances of educational success are not solely tied to financial capacity but are also dependent on a broader spectrum of educational and social variables.