TECHNICAL

ASSIGNMENT 3

‘Affording Academic Success’:

Exploring the Impact of Socioeconomic Status on

Educational Outcomes: An Analysis of Kindergarten Scores

INF2178

SUBMITTED TO: Dr. Shion Guha

SUBMITTED BY: Adil Mahmood Ashraf

Student Number: 1010727017

1- Introduction:

The role of socioeconomic status in shaping educational outcomes has been a topic of extensive research and debate among educators, policymakers, social scientists, and even young couples. My study aims to delve into this issue by examining the influence of income levels on the academic progress of kindergarten students over the Spring and Fall. Specifically, I would analyze two questions:

RQ1: Does the level of income, considering the effect of general knowledge, influence the evolution of reading scores from fall to spring?

Our Null Hypothesis to answer RQ1 is as follows: The level of income does not significantly influence the evolution of reading scores from fall to spring, considering the effect of general knowledge on these scores.

RQ2: Does the level of income, while accounting for the influence of general knowledge, affect the progression of math scores from fall to spring?

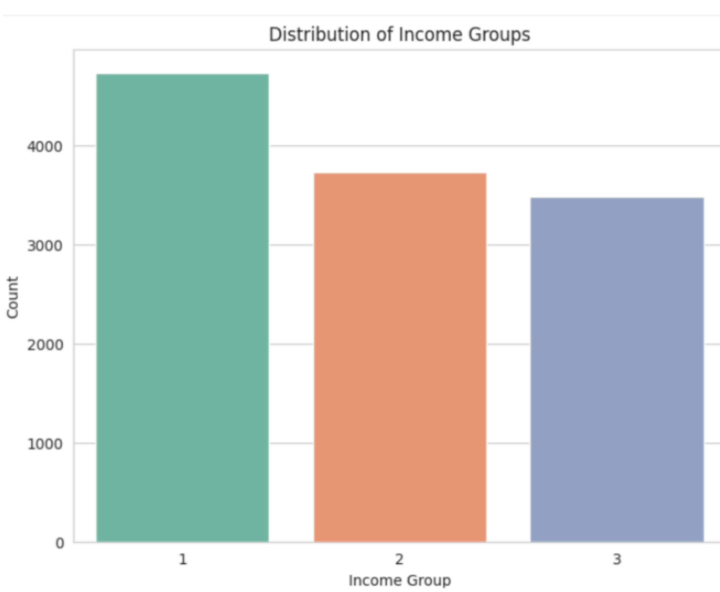
Our Null Hypothesis to answer RQ2 is as follows: The level of income does not significantly affect the progression of math scores from fall to spring, when accounting for the influence of general knowledge on these scores.

2- Data Cleaning and Wrangling:

I began by having a look at the data. It had 11933 rows. On initial visual inspection, there seemed to be no apparent null values. However, as a best standard practice, I ran a python script to drop null values and remove any duplicate entries.

3- Exploratory Data Analysis:

Figure 1: Distribution of Income Groups



In our dataset, we categorize households into three distinct groups based on income levels, labeled as 1, 2, and 3. This classification is derived from a continuous variable that measures total household income. Figure 1 visualizes the various income groups. In order to understand these classes better, I created Table 2 to decipher what income threshold constitutes these classes.

Table 2: A summary of Income Groups

Income Group	From (min)	mean	median	To (max)
1	$1	$22,219.71	$23,000	$39,800
2	$40,000	$51,742.76	$50,000	$69,700
3	$70,000	$100,989.75	$90,000	$150,000

4- Checking Assumptions for ANCOVAS:

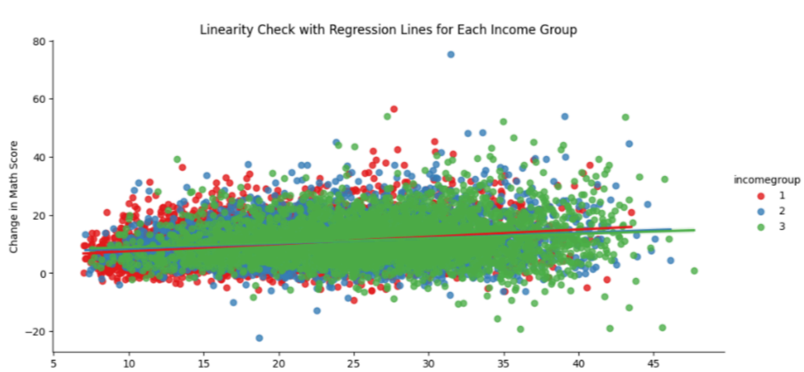
Before we address RQ1 and RQ2, we must check for certain assumptions for Math and Reading Scores. The details are as follows:

Table 3: Math Scores

Test	Statistic	p-Value
Shapiro- Wilk Test	0.966404	0.00E+00
Levene Test	354.449601	1.69E-78

As indicated in Table 3, the Shapiro-Wilk test indicated a significant deviation from normality with a p-value of 0.00E+00. This is a critical finding as it suggests that the math score data do not conform to the normal distribution typically assumed in parametric testing. Non-normality can affect the test's robustness and may lead to an increased chance of making type I errors—incorrectly rejecting the null hypothesis when it is true.

Figure 2: Scatter Plot and Linearity Check for Math Scores

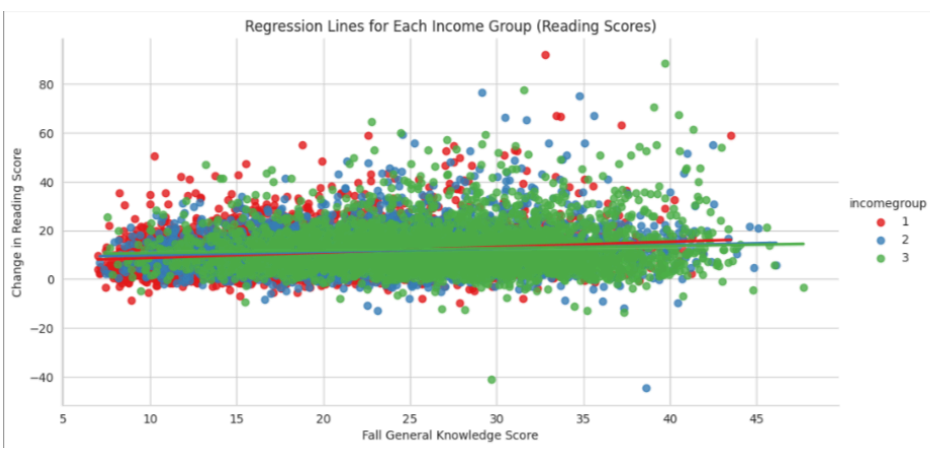


Levene's test for equality of variances also presented a notable statistic of 354.449601 with a p-value of 1.69E-78, indicating that the variances of math scores across the groups defined by income level are not equal. This inhomogeneity of variance violates another core assumption of ANCOVA and could potentially lead to biased estimates of the effect sizes and significance levels.

Table 4: Reading Scores		
Test	Statistic	p-Value
Shapiro- Wilk Test	0.899632	0.00E+00
Levene Test	49.198451	2.38E-12

For the reading scores, the Shapiro-Wilk test yielded a statistic of 0.899632 with a p-value of 0.00E+00, firmly rejecting the null hypothesis of normality. Similarly, for the math scores, the Shapiro-Wilk test statistic was 0.966404 with the same p-value, again indicating a departure from normality. The implications of these results are significant; they suggest that the reading and math score data do not follow a normal distribution, which is an assumption of ANCOVA.

Figure 3: Scatter Plot and Linearity Check for Reading Scores



Levene's test for the reading scores resulted in a statistic of 49.198451 and a p-value of 2.38E-12, while for math scores, the statistic was markedly higher at 354.449601 with a p-value of 1.69E-78. Both outcomes suggest that the assumption of homogeneity of variances is violated for our dependent variables.

These violations of ANCOVA assumptions necessitate caution in the interpretation of the analysis. The non-normality of the data could be addressed by applying a transformation to the scores or by employing a non-parametric alternative to ANCOVA. The inhomogeneity of variances points to differences across groups that are not consistent, which could affect the type II error rate. One may consider using robust ANCOVA techniques or adjusting the model to account for the variance discrepancy. Figures 2 and 3 offer a comprehensive view of the data, showcasing not only the positive linear relationship between fall and spring scores but also the spread and distribution of data points

Despite these statistical challenges, the significant results obtained from the ANCOVA analysis are informative. However, the assumptions check reminds us that while the models provide evidence of the impact of socioeconomic status on academic outcomes, the precise estimation of these effects is complex and must account for the distributional properties of the data. Future analyses could explore methods that are less sensitive to these assumptions or collect data that better meet these conditions.

5- RQ1: Does the level of income, considering the effect of general knowledge, influence the evolution of reading scores from fall to spring?

ANCOVA table for Spring Reading Score				
	sum_sq	**df**	**F**	**PR(>F)**
Income Group	4.70E+02	1	7.43	0.0064
Fall Reading Score	1.55E+06	1	24455.58	< 0.0001
Residual	7.55E+05	11930	NaN	NaN

Utilizing 'springreadingscore' as my dependent variable, 'incomegroup' as the factor, and 'fallreadingscore' as the covariate, the ANCOVA results presented an intriguing revelation. The

F-value of 7.43 for the 'Income Group' indicated a statistically significant effect at a p-value of 0.0064. This finding implies that income levels had a discernible influence on the reading scores of the students when the baseline reading scores in the fall were controlled for.

The overwhelming F-value of 24455.58 for the 'Fall Reading Score' with a p-value of less than 0.0001 cannot be overstated. It signals a very strong predictive relationship between the fall reading scores and those in the spring, suggesting that the students' initial performance was a robust indicator of their subsequent scores.

The residual, while not yielding an F or p-value, represented the variation in spring reading scores not explained by the income group or the fall reading scores. This unexplained variance could potentially be attributed to factors not accounted for in the model.

Based on the results of the ANCOVA analysis, we reject the null hypothesis for RQ1 as the level of income does significantly affect the progression of math scores from fall to spring, even when accounting for the influence of general knowledge.

6- RQ2: Does the level of income, while accounting for the influence of general knowledge, affect the progression of math scores from fall to spring?

ANCOVA Table for Spring Math Score				
	sum_sq	**df**	**F**	**PR(>F)**
Income Group	1.61E+03	1	34.62	4.13E-09
Fall Math Score	1.03E+06	1	22203.54	< 0.0001
Residual	5.52E+05	11930	NaN	NaN

Moving to math scores, with 'springmathscore' as the dependent variable and following the same model structure, the ANCOVA painted a similar yet more pronounced picture. The 'Income

Group' produced an F-value of 34.62, with a strikingly significant p-value of 4.13E-09, suggesting that socioeconomic status played an even more critical role in the progression of math scores compared to reading scores.

Again, 'Fall Math Score' stood out as a highly significant predictor with an F-value of 22203.54 and a p-value of less than 0.0001. This underlines the importance of early academic performance as a strong predictor for future achievement in mathematics.

Similarly, for RQ2, we reject the null hypothesis since the analysis indicates that the level of income significantly influences the evolution of reading scores from fall to spring, beyond the effect of general knowledge.

7- Conclusion:

In conclusion, my investigation into the research questions regarding the role of socioeconomic status in the academic evolution of children—specifically in math and reading performance—reveals that disparities in income are intricately linked to academic outcomes from a young age. For math scores (RQ2), socioeconomic status distinctly contributed to the growth in scores over time. Similarly, for reading scores (RQ1), income levels had a significant effect on their progression. These results support the initial hypotheses and underscore the necessity of integrating socioeconomic considerations into educational interventions. Finally, the challenge to afford academic success for children of young parents remains a pressing pain point. This study taps into the very real concerns for many low income couples who could delay becoming parents knowing that their children may not be set up for academic success from the get-go.