INF2178 Assignment 3
Quantitative Analysis of Spring Academic Performance by Income Group

Liyanze Liu
1009172324

March 2024

1 Introduction

The academic achievements of students during their formative years are critical indicators of future educational outcomes. This report delves into the dynamics of academic performance among Kindergarten students during the spring term, highlighting the potential impact of income groups on learning progress. Utilizing data from a comprehensive early child longitudinal study from the academic year 1998-99, this investigation seeks to untangle the complex relationship between socioeconomic factors and educational attainment.

At the core of this study lies the premise that income groups may differentially influence student performance in reading and math, two foundational areas of early education. By incorporating general knowledge scores from the previous fall as a covariate, the analysis aims to provide a nuanced understanding of how economic backgrounds could correlate with or predict academic growth over the school year.

The importance of this study is underscored by ongoing discussions on educational equity and the role of socioeconomic status in shaping children's academic journeys. Insights gleaned from this research are intended to inform educators, policymakers, and stakeholders committed to fostering learning environments where every child can thrive, regardless of their economic background.

In summary, the ensuing analysis provides a quantitative exploration of how disparities in income influence the academic trajectories of young learners, setting the stage for targeted interventions and informed educational strategies.

2 Data and Methods

2.1 Data

The dataset, named 'INF2178_A3_data.csv,' contains a subset of data from a longitudinal study conducted during the 1998-99 academic year, which followed a group of Kindergarten students. The focus is on the students' reading and math performance recorded in the spring term. The analysis evaluates academic achievement in relation to income groups, using the students' general knowledge scores from the fall as a baseline. The aim is to understand the influence of socioeconomic factors on early educational progress.

2.2 Statistical Approach

The analysis of the relationship between income group and spring academic performance, while controlling for general knowledge scores from the fall, was carried out using a combination of Analysis of Covariance (ANCOVA) and regression modeling.

2.2.1 ANCOVA Analysis

To assess the impact of income group on spring reading and math scores, an ANCOVA was conducted with the income group as the independent variable and fall general knowledge scores as the covariate. This statistical technique allows for the comparison of means across different income groups while adjusting for the baseline scores, providing a clearer picture of the income group's effect on academic performance.

2.2.2 Regression Modeling

Following the ANCOVA, Ordinary Least Squares (OLS) regression models were fitted to further explore the relationships between the variables. The initial models included income group and fall general knowledge scores as predictors to quantify their direct effects on spring academic outcomes.

Subsequent models incorporated interaction terms to investigate whether the relationship between baseline knowledge and spring performance varied across income groups. This step aimed to uncover any nuanced interactions that might exist between socioeconomic status and prior knowledge in influencing educational achievement.

2.3 Hypotheses for ANCOVA Analysis

The null and alternative hypotheses for the ANCOVA analysis comparing spring academic performance across income groups, while controlling for fall general knowledge scores, can be formulated as follows:

2.3.1 Spring Reading Scores

- Null Hypothesis (H_0):
  The adjusted mean spring reading scores are equal across all income groups, after controlling for fall general knowledge scores.
$$H_0 : \mu_{\text{low}}^{\text{reading}} = \mu_{\text{middle}}^{\text{reading}} = \mu_{\text{high}}^{\text{reading}} \tag{1}$$
- Alternative Hypothesis (H_a):
  There is at least one income group whose adjusted mean spring reading score is significantly different from the others.
$$H_a : \text{Not all } \mu_{\text{income group}}^{\text{reading}} \text{ are equal} \tag{2}$$

2.3.2 Spring Math Scores

- Null Hypothesis (H_0):
  The adjusted mean spring math scores are equal across all income groups, after controlling for fall general knowledge scores.
$$H_0 : \mu_{\text{low}}^{\text{math}} = \mu_{\text{middle}}^{\text{math}} = \mu_{\text{high}}^{\text{math}} \tag{3}$$
- Alternative Hypothesis (H_a):
  There is at least one income group whose adjusted mean spring math score is significantly different from the others.
$$H_a : \text{Not all } \mu_{\text{income group}}^{\text{math}} \text{ are equal} \tag{4}$$

In these hypotheses, $\mu_{\text{low}}^{\text{reading}}$, $\mu_{\text{middle}}^{\text{reading}}$, and $\mu_{\text{high}}^{\text{reading}}$ represent the adjusted mean spring reading scores for the low, middle, and high-income groups, respectively. Similarly, $\mu_{\text{low}}^{\text{math}}$, $\mu_{\text{middle}}^{\text{math}}$, and $\mu_{\text{high}}^{\text{math}}$ represent the adjusted mean spring math scores for the respective income groups, adjusted for the fall general knowledge scores.

3 Results

3.1 Spring Reading Scores

3.1.1 Descriptive Statistics

Descriptive statistics were generated to provide a summary of the spring reading scores across different income groups. The number of observations, mean scores, and standard deviations were calculated to offer an initial understanding of the data distribution (Table 1).

incomegroup	count_reading	mean_reading	std_reading
1	4729	43.67	12.00
2	3726	48.01	13.51
3	3478	52.21	16.45

Table 1: Spring Reading Scores by Income Group

3.1.2 Visual Analysis

The graphical representations reveal a discernible correlation between students' foundational knowledge in the fall and their reading achievements in the spring, segmented by income groups (Figure 1). The scatter plot elucidates a positive correlation where students with higher general knowledge scores tend to achieve higher reading scores, a trend that is consistent across all income levels. Higher income groups consistently show higher reading and general knowledge scores, indicating socioeconomic advantages. The box plots confirm this trend, with median scores increasing with income level. These findings suggest that socioeconomic status influences educational outcomes and point towards the potential benefits of focused educational support for lower-income groups.

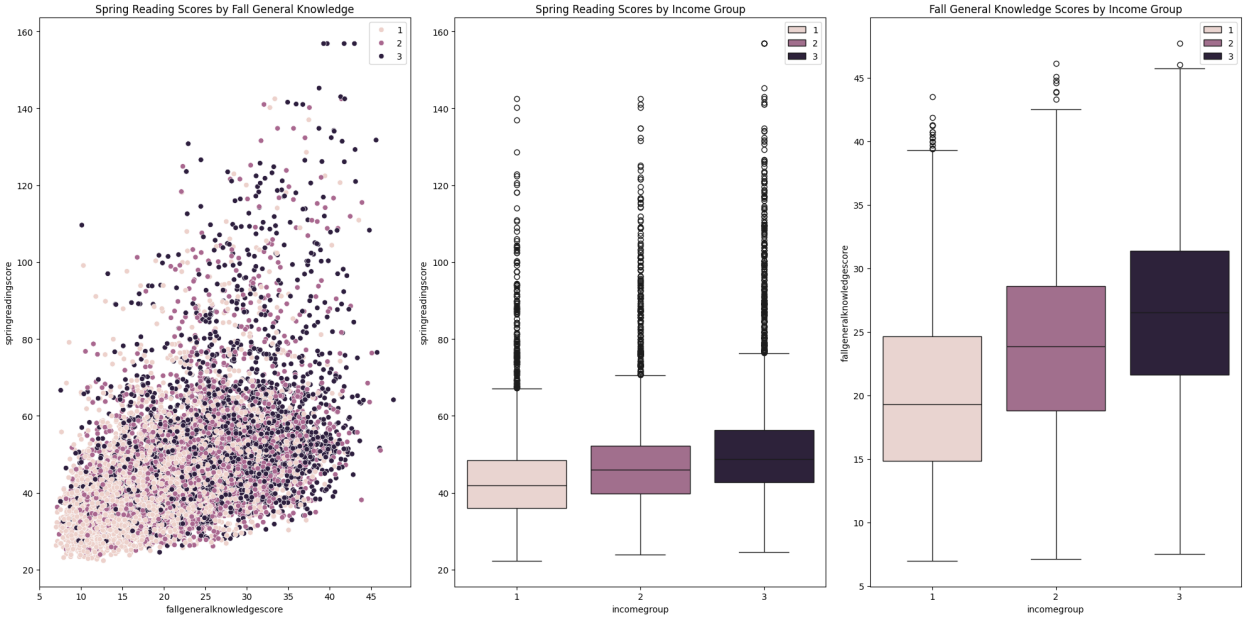

Figure 1: Spring Reading Scores

3.1.3 ANCOVA

The ANCOVA results presented focus on understanding the impact of income group on spring reading scores while controlling for the variability due to students' general knowledge scores from the previous fall (Table 2).

Source	SS	DF	F	p-unc	np2
incomegroup	2.04×10^4	2	62.53	9.69×10^{-28}	0.01
fallgeneralknowledgescore	3.52×10^5	1	2152.46	0.00	0.15
Residual	1.95×10^6	11929	NaN	NaN	NaN

Table 2: ANCOVA Table: Spring Reading Scores

The F-statistic for the income group is 62.52695, with an associated $p < .001$, which is highly significant. This suggests that there are substantial differences in spring reading scores across different income groups. Moreover, the partial eta squared (η^2) for the income group is 0.010374, indicating that although the income group has a significant effect, it accounts for just over 1% of the variance in spring reading scores.

The general knowledge scores from the fall have a much more pronounced effect, with an F-statistic of 2152.459429, which is substantial and accompanied by a p-value ($p < .001$). This indicates a very strong relationship between students' general knowledge in the fall and their reading scores in the spring. The partial eta squared for the fall general knowledge score is 0.152858, suggesting that this covariate alone accounts for approximately 15% of the variance in the spring reading scores, underscoring its importance as a predictor.

The residual sum of squares is 1.949e+06 with 11929 degrees of freedom, representing the variation within spring reading scores that is not explained by either income groups or fall general knowledge scores. The magnitude of this residual variation highlights that there are other factors, not included in this model, that influence spring reading scores. The null hypothesis that income group and fall general knowledge do not affect spring reading scores is soundly rejected, affirming that both are significant factors in the variance of spring reading outcomes among kindergarten students.

3.1.4 OLS Regression Models

The OLS regression analysis indicates that socioeconomic status, as categorized by income groups, alongside initial general knowledge, significantly predicts spring reading scores in Kindergarten. Higher income groups are associated with improved scores: group 2 by 1.2337 points and group 3 by 3.4069 points, in comparison to the baseline group. A positive correlation exists between fall general knowledge and spring reading achievement, with each additional point in fall equating to a 0.7895 point increase in spring scores, confirmed by p-values below 0.001. These findings suggest that early educational strategies should focus on these factors to boost reading skills.

Including interaction terms in the regression model, which has an R-squared value of 0.204, reveals the nuanced interplay between socioeconomic status, baseline knowledge, and reading proficiency. Interaction terms for income groups 2 and 3 do not significantly alter the impact of general knowledge on reading outcomes ($p = 0.465$ and $p = 0.852$, respectively). These interaction model results reveal that the beneficial effect of early general knowledge on reading proficiency is comparable across income groups two and three relative to the baseline. This pattern implies that while socioeconomic status certainly affects reading outcomes, the level of initial general knowledge consistently predicts reading success, irrespective of income variations. It also indicates that higher income levels do not modify the positive influence of early knowledge on reading achievements in the spring term.

3.2 Spring Math Scores

3.2.1 Descriptive Statistics

The analysis began with assessing spring math scores against fall general knowledge and income groups. Higher income was associated with increased math scores, suggesting a correlation between economic status and academic achievement. (Table 3).

incomegroup	count_math	mean_math	std_math
1	4729	33.88	10.73
2	3726	38.46	11.36
3	3478	42.41	12.61

Table 3: Spring Math Scores by Income Group

3.2.2 Visual Analysis

Visual inspection of scatter plots and box plots revealed distinct patterns (Figure 2). The scatter plots displayed a positive correlation between fall general knowledge scores and spring math scores, with higher income groups tending towards higher scores, suggesting a compounding advantage. Box plots corroborated these trends, with median scores for math increasing with income, and the spread of scores indicating variability within income groups. Outliers observed mainly in higher income categories may point to exceptional cases or data entry errors, warranting further investigation.

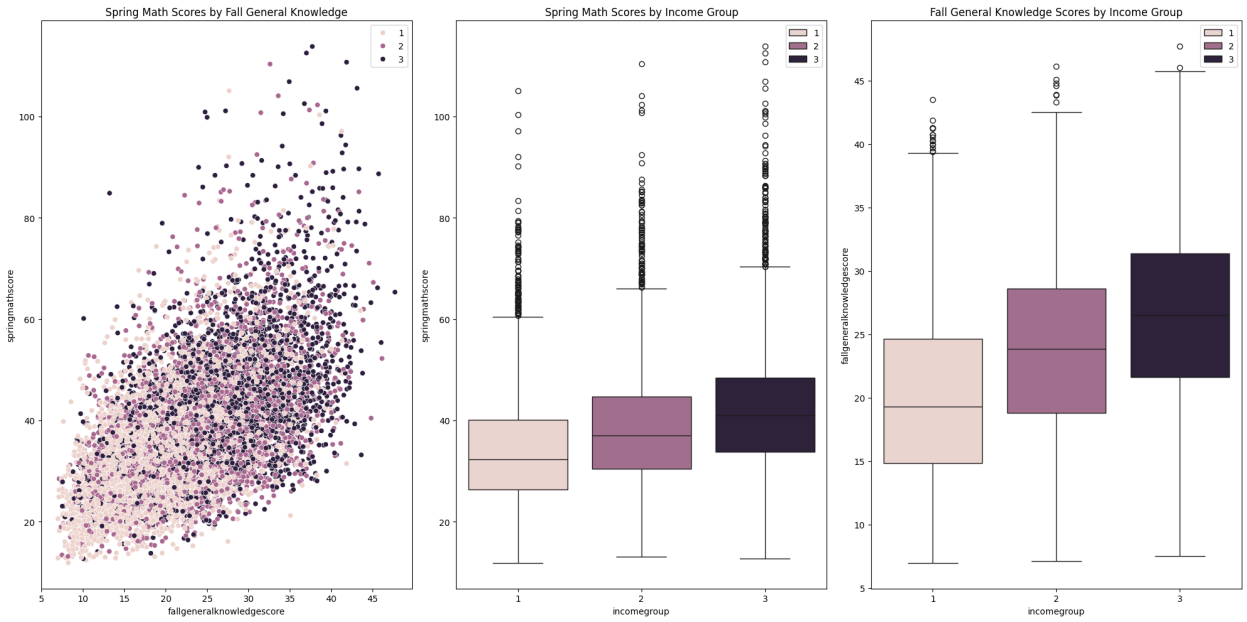

Figure 2: Spring Math Scores

3.2.3 ANCOVA

The ANCOVA results for spring math scores, while accounting for students' general knowledge in the fall, suggest significant effects of both income group and prior general knowledge on the math scores (Table 4).

The F-value for the income group factor is 67.93 with a highly significant p-value ($p < .001$), indicating strong evidence against the null hypothesis of equal mean math scores across income groups. The partial eta squared (η^2) for the income group is 0.01162, which, although relatively small, indicates a non-negligible effect.

The effect of general knowledge from the fall on spring math scores is even more pronounced, as suggested by the F-value of 4795.24 and an accompanying p-value effectively at zero ($p < .001$). The large eta squared value of 0.28672 for this factor indicates that approximately 28.67% of the variance in spring math scores can be attributed to differences in general knowledge scores from the fall, after controlling for income group. This large effect size denotes the substantial role of general knowledge in influencing math outcomes.

In summary, the ANCOVA analysis affirms that both socioeconomic factors, as represented by income groups, and the foundational general knowledge acquired prior to the spring term are key determinants of math performance among kindergarten students. The statistical significance of these effects, backed by exceptionally low p-values, underscores their robustness. The reported eta squared values reinforce the conclusion that while income group has a notable impact, the effect of fall general knowledge is particularly powerful in explaining the variability in spring math scores. These results underscore the importance of both socioeconomic context and early educational foundations in shaping academic trajectories in mathematics.

Source	SS	DF	F	p-unc	np2
incomegroup	1.28×10^4	2	67.93	4.60×10^{-30}	0.01
fallgeneralknowledgescore	4.52×10^5	1	4795.24	0.00	0.29
Residual	1.13×10^6	11 929	NaN	NaN	NaN

Table 4: ANCOVA Table: Spring Math Scores

3.2.4 OLS Regression Models

The regression analysis reveals a correlation between higher income levels and increased math scores for kindergarten students in the spring term, after adjusting for baseline general knowledge from the fall term. Specifically, the average math scores for students in the second income group were 1.0541 points higher, while those in the third income group were 2.7058 points higher than those in the baseline first income group, when other variables are held constant. Additionally, a one-point increase in the fall general knowledge score was linked to a 0.8953 point increase in spring math scores, emphasizing the influence of both socioeconomic status and early knowledge on math achievement in early childhood.

Further examination of the OLS model with interaction terms indicates that belonging to the second or third income group continues to be a significant predictor of math scores. However, the interaction between economic status and fall general knowledge shows intricate patterns. While the influence of fall general knowledge on math scores is decidedly significant ($p < .001$), the interaction term for income group 2 does not show significance ($p > .05$), implying that the benefits of prior knowledge do not vary between the baseline group and income group 2. In contrast, the interaction term for income group 3, though statistically significant ($p = .002$), suggests a potentially different effect of prior knowledge for students in this higher income group. This complex relationship points to the varying impact of socioeconomic status on educational outcomes and the unequal advantages of early academic readiness across different income levels.

4 Conclusion

This study underscores the intertwined nature of socioeconomic factors and early educational foundations in shaping the academic trajectories of young learners. The insights gained from this research emphasize the need for targeted educational strategies and support systems that address both the socioeconomic disparities and the critical role of early learning experiences in fostering academic success.