INF2178: Experimental Design for Data Science
Fatima Zohra (1004167876)

# Assignment 3

## Introduction

This report explores the educational progress of kindergarten students over the academic year 1998-1999, utilizing a subset of data from a longitudinal study. The data includes scores in reading, mathematics, and general knowledge from both fall 1998 and spring 1999. Additionally, income categories are included, offering a lens to examine the influence of socioeconomic status on educational outcomes.

The following research questions (RQs) were investigated to guide our analysis:

(1) How do reading and math scores improve from fall to spring across different income groups?
(2) Is there a significant difference in general knowledge scores between the fall and spring semesters, and does income level affect this?
(3) Does the starting level (fall scores) in one subject (e.g., reading) predict improvement in another (e.g., math) by spring across income groups?
(4) Are there any observable patterns in the scores that can be associated with income bracket changes (e.g., from just below to just above a threshold)?

## RQ1: Income Group Impact on Reading and Math Score Improvement

In the first research question, it was hypothesized that the improvement in reading and math scores from fall to spring would vary across different income groups. To test the hypothesis, a one-way ACOVA was used to examine variance in spring reading scores while controlling for fall reading scores, across the income groups. As seen in Table 1, the ANCOVA results showed a significant effect of income groups on the spring reading scores ($F(2, 11929) = 4.056$, $p = 0.017$), affirming that changes in income brackets were indeed associated with variances in reading scores. Furthermore, fall reading scores exhibited a highly significant impact on spring reading scores ($F(1, 11929) = 24455.398$, $p < 0.0001$), indicating that students' initial performance could predict their subsequent improvement.

**Table 1.** ANCOVA results for income group impact on reading score improvement

| Source | df | Sum of Squares | F | p-value |
|---|---|---|---|---|
| Income Group | 2 | 513.12 | 4.056 | 0.017 |
| Fall Reading Score | 1 | 1,547,042 | 24455.398 | < 0.0001 |
| Residual | 11929 | 754,625.6 | --- | --- |

The violin plots in Figure 1 further illustrated a shift toward higher reading scores from fall to spring across all income groups, suggesting a general trend of improvement. The variability in performance also appeared to increase in the spring, indicating a more diverse range of student outcomes as the academic year progressed.

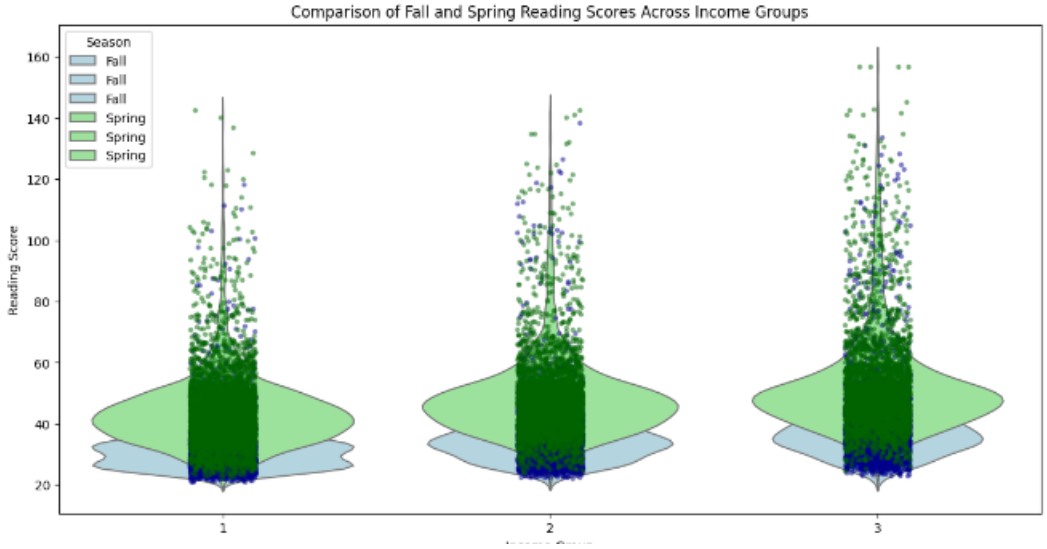


**Figure 1**. Violin plots comparing fall and spring reading scores across income groups

The findings highlight the need for early literacy support, such as after-school reading programs and resource provision like books and educational technology, particularly for students from lower-income groups. These tailored interventions can decrease the influence of socioeconomic status on educational outcomes and foster a fairer educational environment.

**RQ2: General Knowledge Score Improvement and Income Level Influence**

In the second research question, it was hypothesized that the changes in general knowledge scores from fall to spring semesters would be significantly different and that these changes would be influenced by income level. A one-way ANCOVA was conducted to evaluate the variance in the improvement of general knowledge scores while considering income groups and initial scores. As seen in Table 2, the

ANCOVA findings indicated that the income groups did not have a significant effect on the improvement of general knowledge scores ($F(1, 11930) = 0.541$, $p = 0.462$), suggesting that the variation in these scores was not significantly related to income levels.

**Table 2**. ANCOVA Results for Income Group Impact on General Knowledge Score Improvement

| **Source** | **df** | **Sum of Squares** | **F** | **p-value** |
|---|---|---|---|---|
| Income Group | 1 | 8.898 | 0.541 | 0.462 |
| Total Household Income | 1 | 12.858 | 0.782 | 0.377 |
| Residual | 11930 | 196176.093 | --- | --- |

The histogram in Figure 2 shows the distribution of changes in math scores from fall to spring for the lowest and highest income groups. The peaks of the histograms for both groups were on the positive side of the score change axis, indicating overall score improvements. However, the highest income group displayed a broader range of score changes, including some notably high increases, compared to the lowest income group. These insights suggest that while general improvements were seen regardless of income level, the degree and variability of this progress might be affected by other factors not captured by income alone. It underlines the necessity for educators and policymakers to identify and support these additional factors that contribute to students' knowledge acquisition and academic growth.

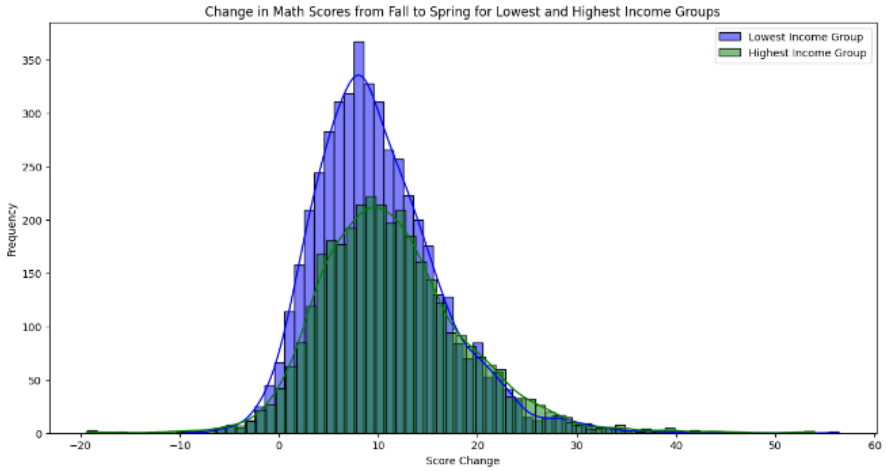


**Figure 2.** Histogram showing the distribution of changes in math scores from fall to spring for the lowest and highest income groups

**RQ3: Cross-Disciplinary Academic Growth Prediction**

In the third research question, it was hypothesized that the starting level (fall scores) in one subject (e.g., reading) would predict improvement in another (e.g., math) by spring, across income groups. A one-way ANCOVA was conducted to examine this, using fall general knowledge scores to predict spring math score improvements while controlling for income groups. According to Table 3, the ANCOVA results revealed that fall general knowledge scores were a significant predictor of math score improvements ($F(1, 11930) = 503.431$, $p < 0.0001$), suggesting that a student's initial general knowledge can forecast their progression in math.

**Table 3**. ANCOVA Results for the Impact of Fall General Knowledge on Math Score Improvement

| **Source** | **df** | **Sum of Squares** | **F** | **p-value** |
|---|---|---|---|---|
| Income Group | 1 | 39.55 | 0.884 | 0.347 |
| Fall General Knowledge Score | 1 | 22529.78 | 503.431 | < 0.0001 |
| Residual | 11930 | 533896.83 | --- | --- |

The correlation matrix in Figure 2 showed moderate positive correlations between fall general knowledge and both spring reading and math scores, indicating that early competence in general knowledge may be associated with later academic success in other domains. The significance of fall general knowledge in predicting math score improvements, while not influenced by income group disparities, suggests the importance of a strong foundational knowledge base. This finding emphasizes the role of cross-disciplinary skills in academic growth and supports the implementation of educational strategies that foster these foundational competencies from an early age.

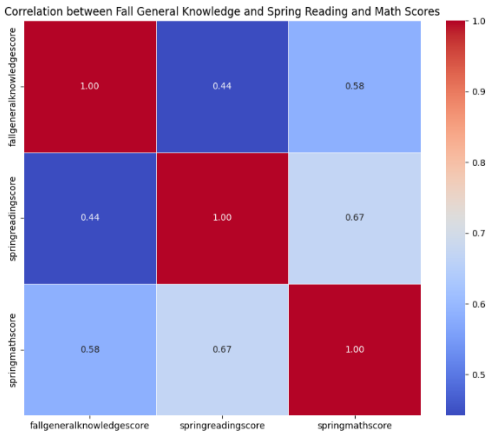


**Figure 2.** Correlation matrix between Fall General Knowledge and Spring Reading and Math Scores

**RQ4: Observable Patterns in Scores Associated with Income Bracket Changes**

In the fourth research question, it was hypothesized that small differences in income, especially around thresholds defining income groups, may have a noticeable impact on academic scores. To investigate this, a one-way ANCOVA model assessed spring reading scores as a dependent variable with income group as an independent variable and fall reading scores as a covariate. The ANCOVA results, as seen in Table 4, indicate a significant effect of income groups on spring reading scores ($F(1, 11930) = 7.43$, $p = 0.0064$), suggesting that slight changes in income brackets can indeed be associated with observable differences in reading scores.

**Table 4.** ANCOVA Results for Observable Patterns in Scores and Income Bracket Changes

| **Source** | **df** | **Sum of Squares** | **F** | **p-value** |
|---|---|---|---|---|
| Income Group | 1 | 469.9812 | 7.43 | 0.0064 |
| Fall Reading Score | 1 | 1,547,013 | 24455.58 | < 0.0001 |
| Residual | 11930 | 754,668.7 | --- | --- |

As seen in Figure 4, the box plots further demonstrated variability in spring reading scores across income groups, with a slight increase in median scores corresponding to higher income brackets, although the range of scores within each group remained similar, indicating that the spread of scores was not significantly impacted by income. These findings suggest that while income may not drastically change the variability within each group, it can influence the median reading performance, potentially due to factors such as access to resources or educational support that correspond with higher income levels.

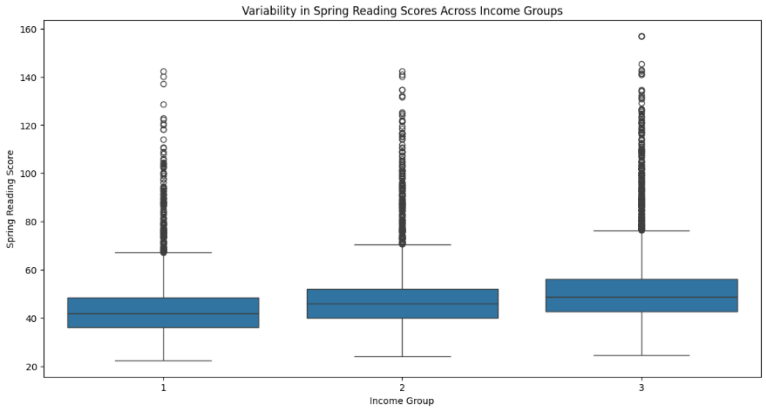


**Figure 4.** Boxplots of Spring Reading Scores Across Income Groups

## Conclusion

This report provided an analysis of the educational progress of kindergarten students during the 1998-1999 academic year, focusing on the relationship between income levels and academic performance in reading, mathematics, and general knowledge. Key findings from the analysis include the significant role of initial academic performance in predicting subsequent improvement, the varying impact of income on different subject areas, and the importance of foundational knowledge in cross-disciplinary academic growth. Income level was found to influence reading score improvements significantly, indicating that socioeconomic status plays a critical role in early literacy development. Conversely, general knowledge score improvements did not significantly correlate with income, suggesting that other factors may be more influential in this domain. The predictive power of initial general knowledge on subsequent math performance highlighted the value of a strong foundational knowledge base across different subject areas.

Further investigation is needed to explore the effects of income on educational outcomes. Questions such as how non-economic factors (e.g., parental involvement, access to quality educators, and extracurricular opportunities) intersect with income to influence academic performance could provide deeper insights. Additionally, longitudinal studies extending beyond a single academic year would help in understanding the long-term impacts of income on educational trajectories.