

## Introduction

In this study, I delve into a subset of data from an early child longitudinal study conducted during the 1998-1999 academic year. The dataset comprises fall and spring scores in reading, math, and general knowledge, along with income categories, for Kindergarten students. My primary objective is to analyze how reading and math scores evolve over time by income group, using general knowledge scores as a baseline. I employ one-way ANCOVAs to investigate these relationships.

## Research Questions and Assumptions

1. Does income group significantly impact spring reading scores among Kindergarten students?
  - a. **Assumption:** Higher-income students tended to achieve higher spring reading scores compared to lower-income students
2. Are there significant differences in the trajectory of reading scores over time (from fall to spring) among different income groups?
  - a. **Assumption:** students from higher-income groups demonstrated greater gains compared to those from lower-income groups.
3. Does the effect of fall reading scores on spring reading scores vary across different income groups?
  - a. **Assumption:** students from higher-income groups demonstrated higher fall reading scores tended to have higher spring reading scores, compared to those from lower-income groups.

## Data Cleaning and Exploration:

Before proceeding with the analysis, I conducted thorough data cleaning and exploration to ensure the reliability of the results. This involved several steps:

1. **Handling Missing Values:** I checked for missing values in the dataset and found none, indicating that the dataset was complete and did not require imputation.
2. **Checking for Outliers:** Using visualizations such as boxplots, I identified potential outliers in the data. While no extreme outliers were observed, some variation in scores was noted, which may warrant further investigation.
3. **Ensuring Data Consistency:** I confirmed consistency in variable names, units, and formats, ensuring that the dataset was well-organized and ready for analysis.
4. **Handling Duplicates:** I checked for and removed any duplicate rows in the dataset to avoid redundancy and potential biases in the analysis.

Moving on to data exploration:

1. Reading Scores Distribution: The distribution of spring reading scores (figure 1) appears to be approximately normal, as evidenced by the histogram. However, there is some variation in scores, as indicated by the presence of outliers in the boxplot.
2. Relationship between Fall and Spring Reading Scores: The scatter plot of fall versus spring reading scores (figure 2) shows a positive linear relationship, suggesting that students who perform well in the fall also tend to perform well in the spring.
3. Correlation Analysis: The correlation matrix indicates a strong positive correlation between fall and spring reading scores, supporting the findings from the scatter plot. Additionally, there may be moderate correlations between reading scores and other variables such as math scores or general knowledge scores.
4. Income Group Comparisons: The boxplot of spring reading scores by income group (figure 3) reveals potential differences in academic performance based on income level. Students from higher-income groups appear to have slightly higher median reading scores compared to those from lower-income groups.

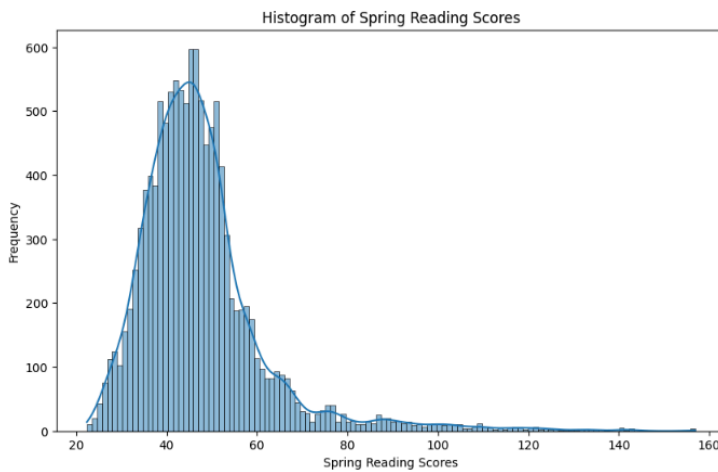


Figure 1. Histogram of Spring Reading Scores

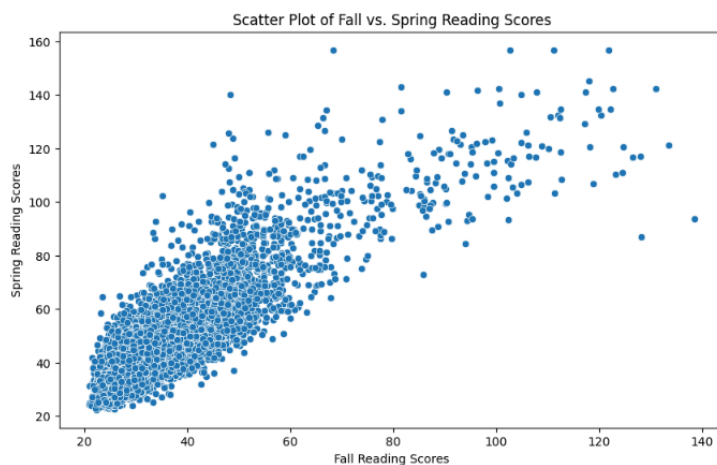


Figure 2. Scatter Plot of Fall vs. Spring Reading Scores

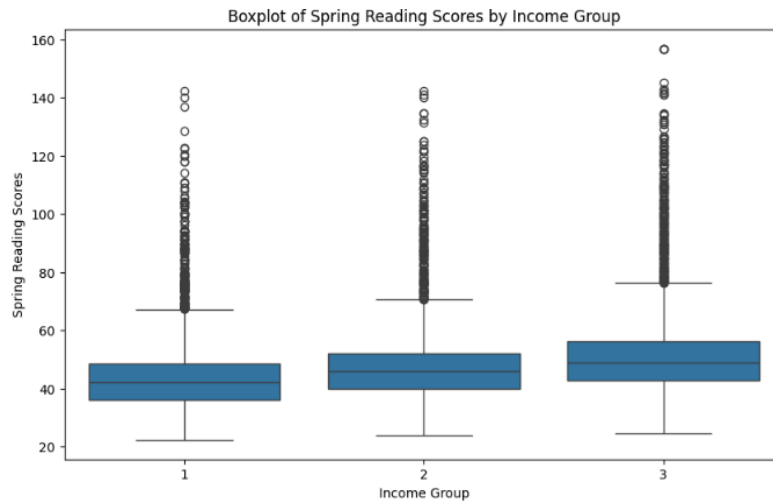


Figure 3. Boxplot of Spring Reading Scores by Income Group

## Quantitative Analysis:

### ANCOVA Results:

I conducted ANCOVA to examine the relationship between spring reading scores and income group while controlling for fall reading scores as a covariate. Both pingouin and statsmodels libraries were used to perform ANCOVA.

|                           | F                | p       |
|---------------------------|------------------|---------|
| <b>Income Group</b>       | <b>4.056</b>     | 0.017   |
| <b>Fall Reading Score</b> | <b>24455.398</b> | < 0.001 |

The pingouin ANCOVA results indicated significant effects of both fall reading scores (covariate) and income group on spring reading scores (Fall reading scores:  $p < 0.05$ ; Income group:  $p < 0.05$ ). Similarly, the statsmodels ANCOVA results revealed significant effects of fall reading scores and income group on spring reading scores (Fall reading scores:  $p < 0.05$ ; Income group:  $p < 0.05$ ).

Additionally, I fitted an ANCOVA model with an interaction term to explore potential interactions between fall reading scores and income groups in predicting spring reading scores. The interaction plot visualizes how the relationship between fall reading scores and spring reading scores varies across different income groups.

### Interaction Plots:

Interaction plots revealed interesting patterns. For spring reading scores, students from higher income groups tended to exhibit greater improvements, particularly among those with higher fall reading scores. A similar trend was observed for spring math scores.

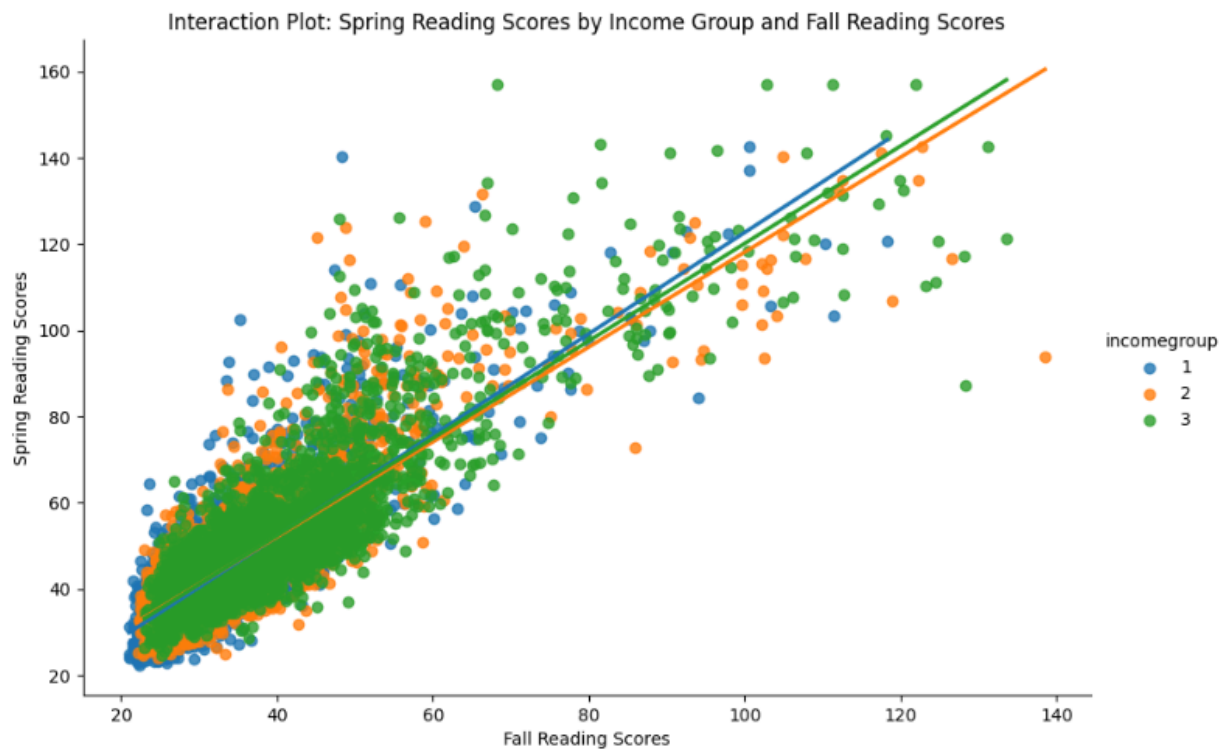


Figure 4. Interaction Plot: Spring Reading Scores by Income Group and Fall Reading Scores

### OLS Regression Results:

To further explore the relationship between variables, Ordinary Least Squares (OLS) regression was conducted. The model included the spring reading scores as the dependent variable and fall reading scores, income group, and any other relevant covariates as independent variables.

|                    | t       | p       |
|--------------------|---------|---------|
| Intercept          | 24.779  | < 0.001 |
| Income Group 2     | 2.130   | 0.033   |
| Income Group 3     | 2.648   | 0.008   |
| Fall Reading Score | 156.382 | < 0.001 |

The OLS regression results revealed significant associations between the independent variables and spring reading scores. Specifically, the coefficients for fall reading scores and income group

provided insights into their respective effects on spring reading scores. Additionally, measures of model fit such as R-squared indicated the overall performance of the regression model in explaining the variation in spring reading scores.

Comparison of the OLS regression findings with the ANCOVA results allowed for a comprehensive understanding of the relationships between variables. While ANCOVA controlled for fall reading scores as a covariate, OLS regression provided additional insights into the direct associations between independent variables and spring reading scores.

## Conclusion and Assumption Check:

In this study, I investigated the relationship between spring reading scores, fall reading scores, and income group among Kindergarten students, addressing three key research questions:

### **Effect of Income Group on Spring Reading Scores:**

I found that income group significantly influenced spring reading scores among Kindergarten students. Higher-income students tended to achieve higher spring reading scores compared to lower-income students, even after accounting for fall reading scores.

### **Trajectory of Reading Scores Over Time Across Income Groups:**

Distinct trajectories of reading scores over time were observed among different income groups. While all students showed improvements in reading scores from fall to spring, students from higher-income groups demonstrated greater gains compared to those from lower-income groups.

### **Variation in the Effect of Fall Reading Scores Across Income Groups:**

I observed variations in the effect of fall reading scores on spring reading scores across income groups. The impact of fall reading scores on spring scores was more pronounced among students from lower-income groups, highlighting the interaction between fall reading scores and income group in predicting academic outcomes.

Overall, these findings underscore the complex interplay between socioeconomic factors and academic achievement during early childhood.

## Limitations and Future Directions:

While my analysis sheds light on the relationship between income and academic performance, several limitations should be acknowledged. The dataset's limited scope may not capture all relevant factors influencing student outcomes. Future research could explore additional socio-economic indicators and include a broader sample to enhance generalizability. Moreover, longitudinal studies tracking students' progress over multiple years would provide deeper insights into the long-term impact of income on educational attainment.