

Exploring Aspects Affecting Early Child Longitudinal Study

1. Introduction

There is more and more focus put on early child longitudinal study since last 2000. However, the overall performance or grades of these young students can be affected by multiple factors such as their pre-educated level, some socio-economic factors like family income level and the ability to learn new knowledge.

This report conducted the quantitative analysis using ANCOVA with a goal to uncover the underlying factors that affect the Kindergarten scores by using the dataset title INF2178_A3_data. This report will first analyze some data in this dataset, and then conduct some data wrangling process before delving into the formal ANCOVA analysis. Finally, some useful insights will be derived based on the result of ANCOVA analysis.

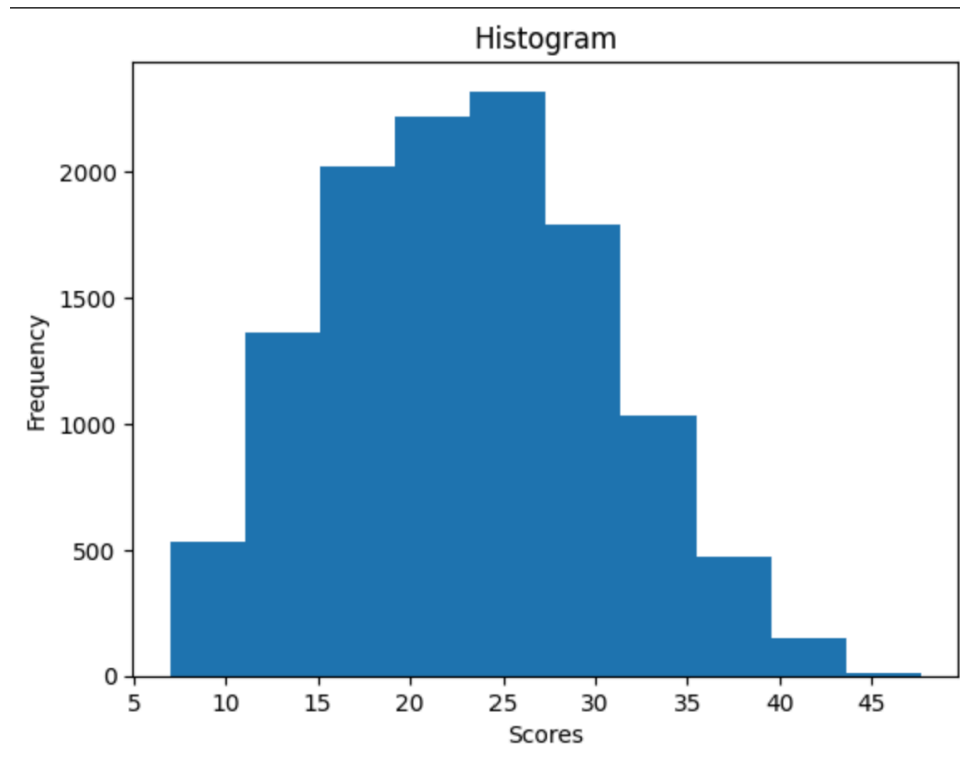
The research will address three fundamental research questions.

1. **Research Question 1:** How the income group will affect the spring math score while controlling the fall math score
2. **Research Question 2:** How the level of fall general knowledge score will affect the spring general knowledge score while controlling 3 different scores in fall
3. **Research Question 3:** How the level of fall general knowledge score will affect the improvements on the score of general knowledge while controlling the improvements on math score

As it mentioned in some research questions, this research has involved some aspects like improvements on score. Moreover, we have created a general knowledge level based on the fall general knowledge score served as the baseline. Hence, some data wrangling is conducted as follows.

Data Wrangling

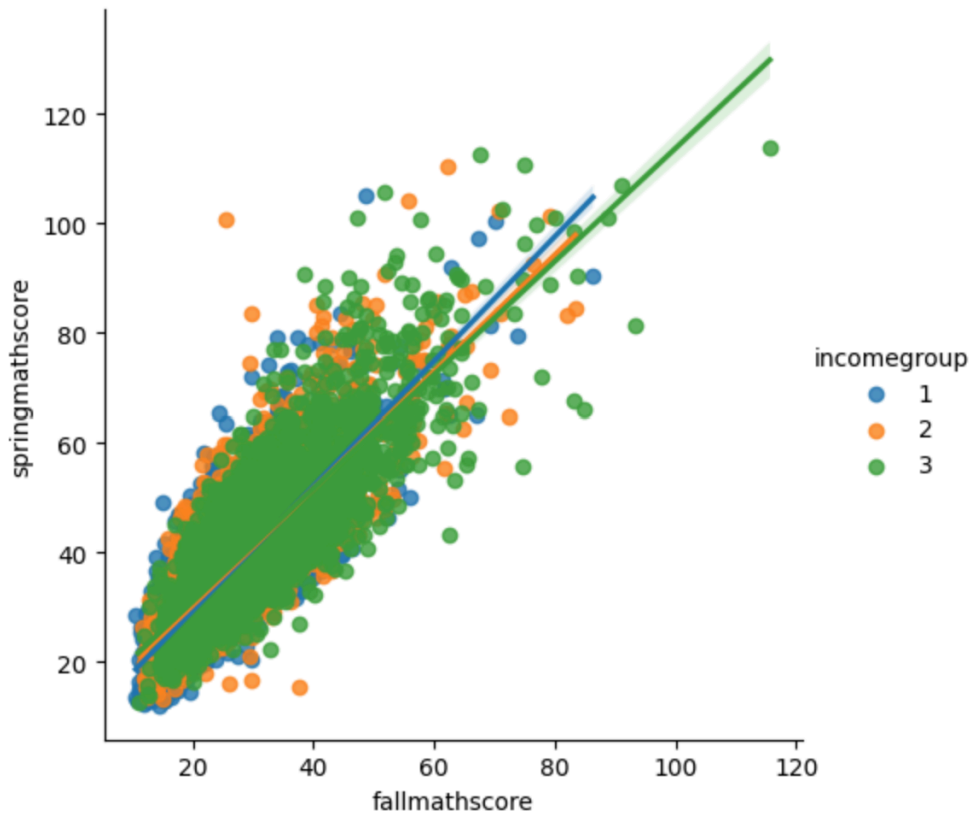
Before adding or deleting some data in the dataset, the first thing done in this research is plotting the distribution of general knowledge scores in fall semester in order to create a new variable called *fallscorelevel* to serve as the baseline for further analysis.



Based on the histogram, an A of *fallscorelevel* is corresponding to the fall general knowledge score of 35+, a B is corresponding to the score between 25-35, a C is corresponding to the score between 15 and 25 and D is corresponding to the score less than 15. Besides creating the variable of *fallscorelevel*, this report also calculates the improvements on different scores by calculating the difference of two semesters, and named variables as *mathscorediff*, *generaldiff* and *readingdiff*.

ANCOVA Analysis

In this report, we will conduct 3 different analyses corresponding to 3 different research questions. In order to make sure the ANCOVA analysis can output an effective result and provide useful insights, we have created some scatterplots to make sure the assumptions of ANCOVA have been met before delving into the formal ANCOVA analysis.



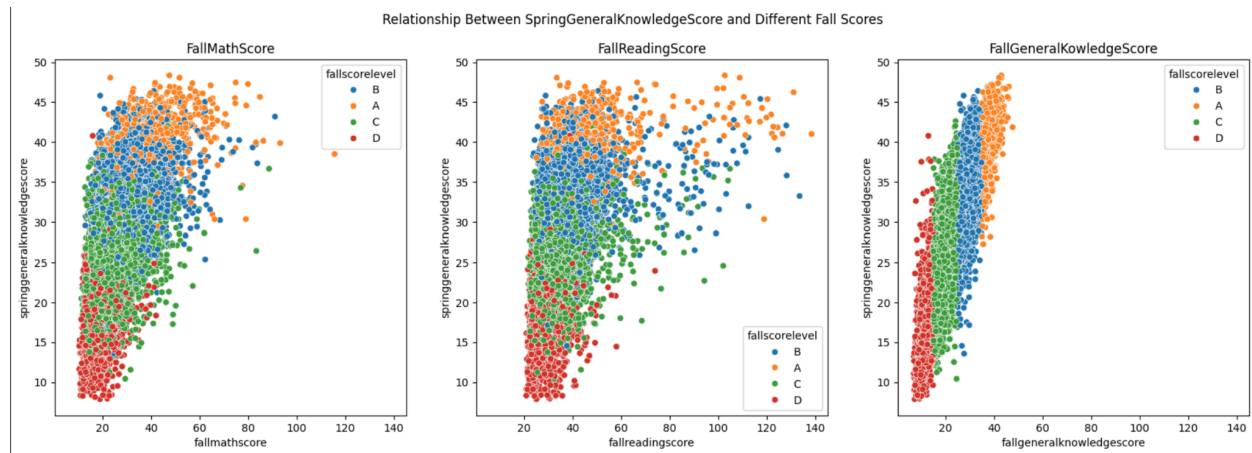
The first test is to see how the income group will affect the spring math score while controlling the fall math score. Here we have created a scatterplot to see if the dependent variable and covariates have the linear relationship and if the homogeneity of slope assumptions have been met. Based on the graph, we can see that both assumptions have been met due to a strong linear relationship and relatively similar slope. Hence, we have gone to the formal ANCOVA analysis, and got the following result.

#1 ANCOVA Result

- **Income Group** F-statistic: 18.523585 | P-value: 9.284861e-09 | np2: 0.003096
- **Fall Math Score** F-statistic: 22203.081238 | P-value: <0.001 | np2: 0.650505

Based on the results, the F-statistic and corresponding p-value of the income group suggest that there are statistically significant differences between the means of different income groups, after adjusting for the covariate. However, the effect size (np2) is very small, indicating that the practical significance of this difference might be minimal. On the other hand, the F-statistic and p-value for fall math score indicate a very strong and statistically significant effect on the dependent variable. The effect size is large, suggesting that fall math score is a major predictor of

the outcome variable and explains a significant portion of the variance. Based on this analysis, we can derive some insights that although difference caused by income group exists, they may not be substantial when it comes to their practical impact on the future math score, especially when previous math scores are considered, indicating that while socioeconomic status has an influence, other factors like previous math level are more pivotal.



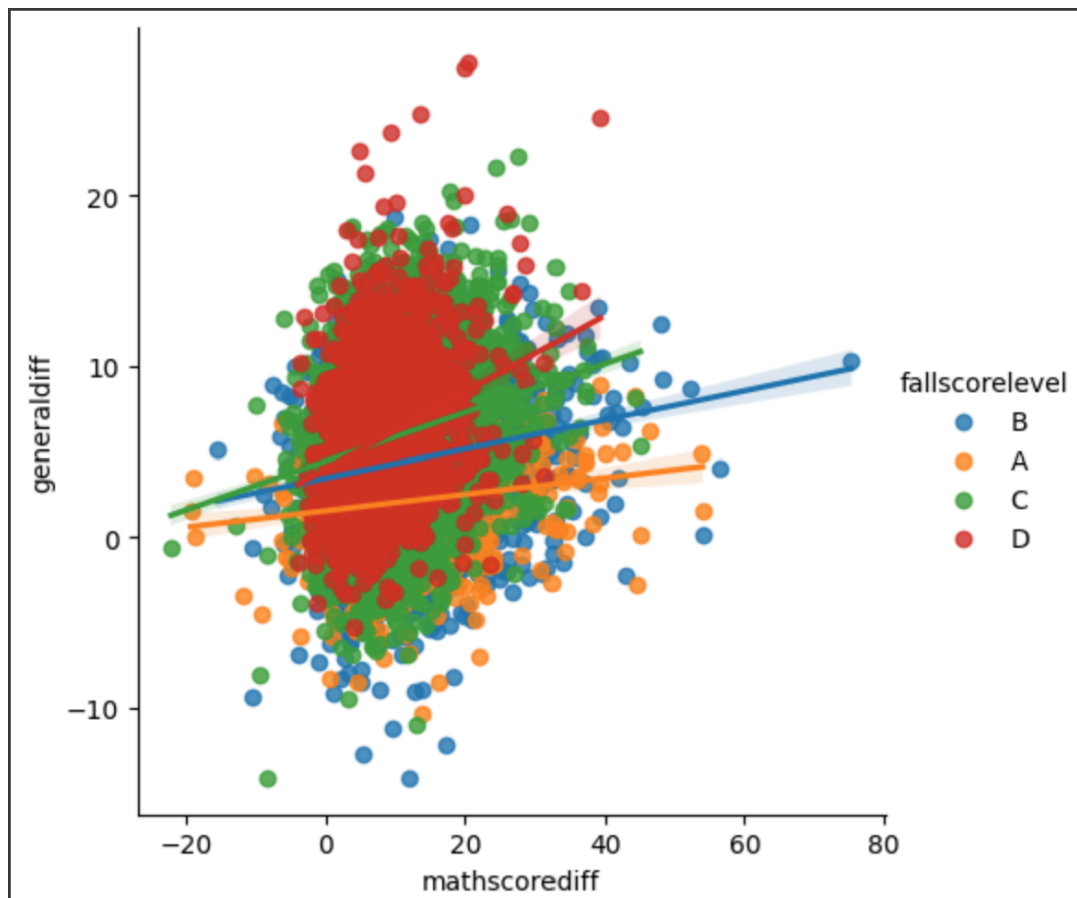
The second ANCOVA focused on investigating how the level of fall general knowledge score will affect the spring general knowledge score while controlling 3 different scores in fall. Similar to the previous test, we have still created a scatterplot to see if assumptions are met. As the results, we could see that both assumptions of linear relationship and homogeneity of slopes are met.

#2 ANCOVA Result

- **Fallscorelevel [T.B]** coef: -4.8766 | p-value < 0.001
- **Fallscorelevel[T.C]** coef: -11.2154 | p-value < 0.001
- **Fallscorelevel[T.D]** coef: -18.4356 | p-value < 0.001
- **Fall Math Score** coef: 0.1464 | p-value < 0.001
- **Fall Reading Score** coef: 0.0358 | p-value < 0.001

From the above results, we can see that Level B, C and D all are associated with a decrease in the spring general knowledge score compared to level A. Also the very low p-values for the coefficients indicate that these variables are statistically significant predictors of the spring general knowledge score. In conclusion, the model indicates that both math and reading scores from the fall positively influence the spring general knowledge score, with math having a stronger impact than reading. The significant negative coefficients for fallscorelevel categories

B, C and D suggests a clear trend that lower fall score levels are associated with lower spring general knowledge scores. This trend is both statically significant and likely of practical significance, given the magnitude of the coefficients. Lastly, the positive coefficients for fall math score and fall reading score reinforce the importance of academic performance in math and reading as predictors of broader knowledge outcomes. This particular test result has highlighted some specific areas where interventions could be most effective, such as enhancing math and reading skills, to improve general knowledge outcomes.



The third ANCOVA test aims to find how the level of the general knowledge score will affect the improvements on the score of general knowledge while controlling the improvements on math score. Also, we still make sure that assumptions of linear relationship and homogeneity of slopes are satisfied.

#3 ANCOVA Result

- **Fallscorelevel [T.B]** coef: 0.4109 | p-value < 0.001
- **Fallscorelevel[T.C]** coef: 4.0447 | p-value < 0.001

- **Fallscorelevel[T.D]** coef: 4.1735 | p-value < 0.001
- **Mathscorediff** coef: 0.1149 | p-value < 0.001

This particular ANCOVA analysis provides insights into how different levels of performance in a previous assessment, the fall score level and changes in math scores are associated with differences in the general knowledge score. The model indicates that higher levels of initial performance and improvements in math scores are associated with higher differences in general performance, although these factors only explain a small portion of the variability in general performance differences.

Conclusion

This report offered a multifaceted view into the relationships between children's academic performance across different metrics and conditions. Three ANCOVA analyses collectively highlight the critical role of early academic achievement, particularly in math, in determining future academic outcomes. They suggest that interventions aimed at improving math skills could have broad implications for overall academic success. Moreover, the findings point to the nuanced ways in which initial performance levels set the stage for future academic gains or losses, emphasizing the importance of early support and target educational strategies to address disparities and foster academic growth.