Jingxian Chen

1005354572

INF2178: Assignment3

Background and Introduction

This report examines the influence of socioeconomic status on the academic progress of kindergarteners. Using data from a 1998-1999 longitudinal study, we analyze how income levels correlate with improvements in reading, math, and general knowledge over a school year. Through one-way ANCOVAs, we explore the changes in scores by income group, controlling for initial knowledge. This analysis seeks to uncover patterns that could inform educational strategies and policies aimed at fostering equitable learning environments.

Research Questions

1. Does income groups have an effect on the improvement of reading scores from fall to spring, after controlling for general knowledge?
2. Do income groups and the improvement of reading score have an effect on the improvement of math scores from fall to spring, after controlling for general knowledge?
3. How does the relationship between reading score improvement and math score improvement vary across different income groups when controlling for baseline general knowledge scores?

Exploratory Data Analysis

The Exploratory Data Analysis presentes consists of distribution plots for reading, math, and general knowledge scores for fall and spring semesters, income distribution, a correlation heatmap, scatter plots with boxplots for scores by income group, and distribution plots for reading score improvement.

Score Distribution Plots: The histograms for fall and spring reading, math, and general knowledge scores display the frequency distribution of scores among kindergarten students. Scores seem to be normally distributed, indicating a standard spread of student abilities.

Income Distribution Plot: This histogram shows the distribution of total household income. The data exhibits right skewness, suggesting that most students come from lower-income households, with fewer students from higher-income families.
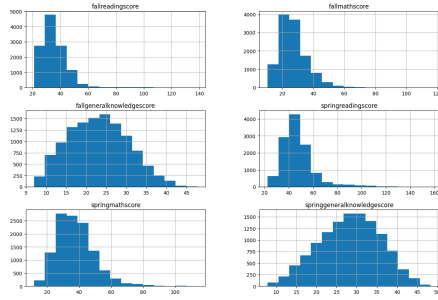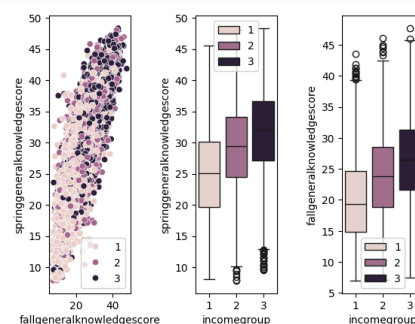
(Figure 1: Distribution of Scores)
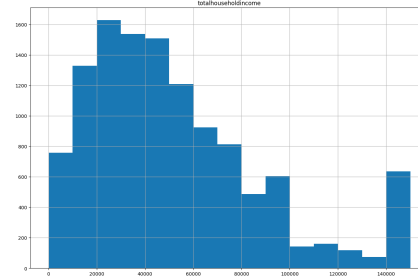
(Figure 2: Distribution of Income)

Correlation Heatmap: The heatmap indicates the strength of the relationship between different variables. There are strong correlations between fall and spring scores for reading, math, and general knowledge, which is expected as student performance is likely to be consistent across semesters. The correlation between income variables and scores appears to be weaker.
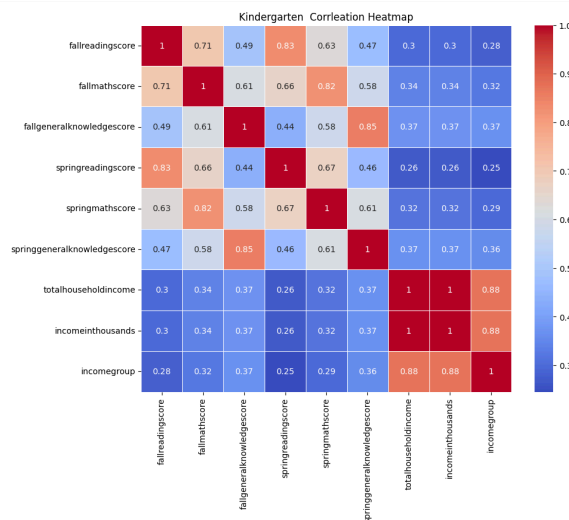
(Figure 3: Heatmap)

Scatter and Boxplots: These show the relationship between general knowledge scores in fall and spring and how they are distributed across income groups. The scatter plot suggests a positive correlation between fall and spring general knowledge scores. Boxplots indicate the spread and median of scores within each income group, with median scores increasing from income group 1 to 3.

(Figure 4: Scatter and Boxplots)

Reading Score Improvement Distribution: A histogram shows the distribution of reading score improvements from fall to spring. It resembles a normal distribution, with a slight skew towards positive improvement, suggesting most students improved their reading scores.

Boxplot of Reading Score Improvement by Income Group: The boxplot shows that reading score improvements are relatively consistent across income groups. All income groups have a similar interquartile range, median improvement, and presence of outliers.

Summary Statistics: These provide a quick numerical overview of score improvements and income groups. The statistics for reading score improvements indicate variability with a wide range of scores, while income group categories are evenly distributed.

Results of Testing the Assumptions for Running One-way ANCOVA

1. Does income groups have an effect on the improvement of reading scores from fall to spring, after controlling for general knowledge?

One-Way ANCOVA Results Interpretation

The results of the ANCOVA analysis indicate that, after controlling for general knowledge scores, there is a statistically significant difference in reading score improvement between at least one of the income groups compared to the reference group. Specifically, incomegroup[T.3] shows a statistically significant difference in reading score improvement, while incomegroup[T.2] does not. However, the overall model explains only a small fraction of the variability in reading score improvement, suggesting that other unaccounted factors may also be influencing the outcome. This finding prompts consideration of additional variables that could affect reading score improvement and suggests that while income groups can have an impact, it is not the sole predictor of academic progress.

Assumption Analysis in One-way ANCOVA
**Assumption 1: Normality Assumption**
1- Normality of Residuals
The Q-Q plot shows a clear deviation of the residuals from the line in the tails, especially in the right tail. This indicates that the residuals have more extreme values than would be expected under a normal distribution, suggesting that the residuals are not normally distributed.

2 - Statistical Test for Normality
The histogram shows that the residuals are fairly symmetric but have a peak (leptokurtic distribution), which deviates from the bell-shaped curve of a normal distribution. Moreover, the p-value from the Shapiro-Wilk test is extremely low (practically zero). These suggest that the residuals may not be normally distributed.
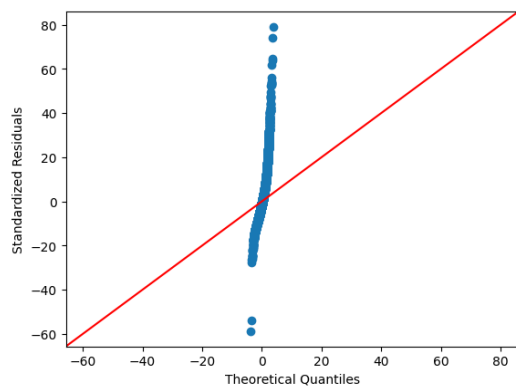
(Figure 5: Q-Q Plot)
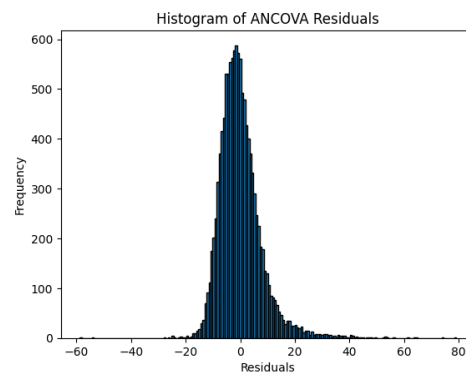
(Figure 6: Histogram of Residuals)

Assumption 2: Homogeneity Test Interpretation
A p-value of 2.79e-09 is well below the conventional alpha level of 0.05, indicating that there is a statistically significant difference in variances across the groups. This result suggests a violation of the homogeneity of variances assumption for ANCOVA. When this assumption is violated, the results of the ANCOVA may not be reliable.

2. **Do income groups and the improvement of reading score have an effect on the improvement of math scores from fall to spring, after controlling for general knowledge?**

One-Way ANCOVA Results Interpretation

When examining the specific predictors, income group classifications and reading score improvements did not have a statistically significant impact on math score improvements, with p-values well above the conventional 0.05 threshold for significance. In contrast, general knowledge scores were found to be a significant predictor; a one-unit increase in fall general knowledge score was associated with an average increase of 0.1619 units in math score improvement. This suggests that interventions aimed at improving general knowledge could be beneficial for enhancing math scores, irrespective of income group or reading score improvement.

Assumption Analysis in One-way ANCOVA

Assumption 1: Normality Assumption
1 - Normality of Residuals

The residuals seem to deviate from the red line in the tails, which might indicate potential outliers or a departure from normality in the tails of the distribution.

2 - Statistical Test for Normality

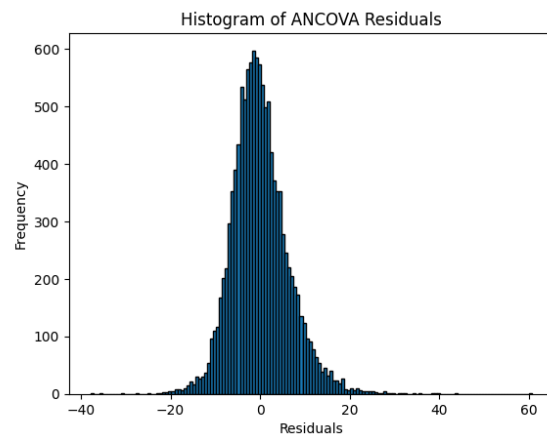The residuals seem to deviate from the red line in the tails, which might indicate potential outliers or a departure from normality in the tails of the distribution. Additionally, the p-value from the Shapiro-Wilk test is extremely low (practically zero), suggesting that the residuals do not follow a normal distribution.
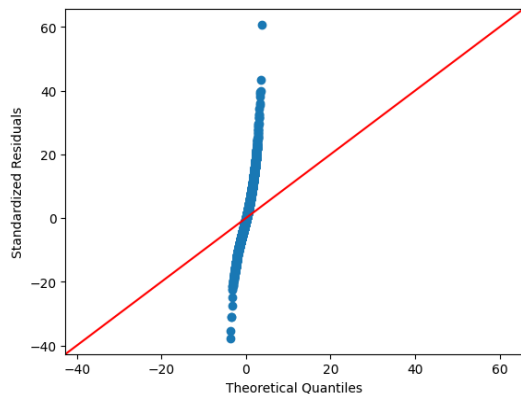
(Figure 7: Q-Q Plot)

(Figure 8: Histogram of Residuals)

Assumption 2: Homogeneity Test Interpretation

In the output provided, Levene's statistic is 19.728, and the p-value is approximately 2.79e-09, which is much less than 0.05. This would lead to the rejection of the null hypothesis of equal variances across the different income groups for the reading score improvement. Thus, we would conclude that the variances are not homogeneous and that this assumption of ANCOVA is violated.

3. How does the relationship between reading score improvement and math score improvement vary across different income groups when controlling for baseline general knowledge scores?

**One-Way ANCOVA Results Interpretation**

The results indicate that not only do income groups and reading score improvement have significant effects on math score improvement individually, but the impact of reading improvement on math improvement also changes depending on the income group. The interaction term's significance suggests that the relationship between reading and math improvements is not uniform across all income groups.

Interaction Plot

Income Group 1: Students in this group show a larger difference in mean spring reading scores based on the direction of their math score change, with a greater mean score for those whose math scores increased compared to those whose math scores decreased.

Income Groups 2 and 3: These groups display a smaller discrepancy in mean spring reading scores between students with increased math scores and those with decreased scores. Notably, for these income groups, the mean reading scores are higher overall compared to Income Group 1, irrespective of the math score trend.

For all income groups, students whose math scores increased tend to have higher mean spring reading scores than those whose math scores decreased. The convergence of the lines for Income Groups 2 and 3 suggests that the influence of changing math scores on spring reading scores is relatively consistent in these groups, whereas the steeper slope for Income Group 1 indicates a stronger relationship between the changes in math and reading scores in this group.
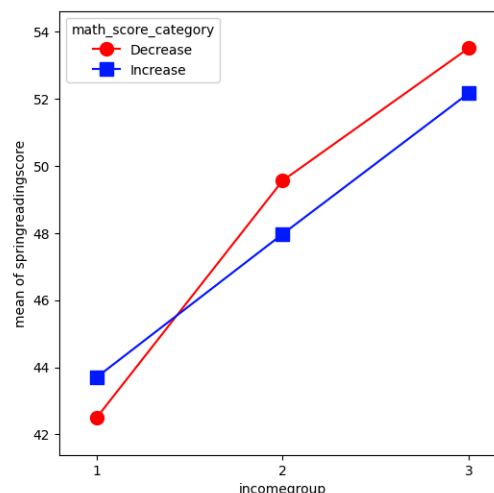
(Figure 9: Interaction Plot)

Conclusion

This analysis of kindergarteners' academic progress over a school year reveals the significant role of income groups in reading score improvements when baseline general knowledge is considered. Although income status proved to be a meaningful factor, its overall explanatory power was limited, pointing to other influences on learning outcomes. The ANCOVA results were further complicated by the violation of the homogeneity of variance assumption, suggesting the need for cautious interpretation. The interaction between math and reading score improvements varied across income groups, with a stronger dependency noted among the lowest income group. These insights reinforce the importance of addressing socioeconomic disparities to foster equitable educational growth.