INF2178 Assignment 3

Zonglin Li
1004910117

Introduction

In this study, we are focused on investigating the potential impact of household income levels on the academic performance of kindergarten children. To ensure the validity of our findings, we utilized data collected from an early child longitudinal study from the fall of 1998 to the spring of 1999, including children's reading, mathematics, and general knowledge performance. We used the fall scores as covariates to control for the children's initial academic abilities as they entered the school year. The one-way ANCOVA, a robust statistical tool, was used to isolate the effect of household income levels on students' academic progression from fall to spring, adjusting for their initial scores. This approach provides a straightforward statistical method to understand how family income may affect children's school grades for a year, offering insight into how family income levels may specifically affect children's academic performance observed from fall to spring.
Research question 1: How does household income affect general academic growth in children over an academic year?
Research question 2: Does household income influence children's academic performance in reading skills differently than in math or general knowledge?
My initial assumption is that children who come from the high-income level family will have higher scores at the beginning and score growth because families with high total incomes will spend more money on their children's education, such as hiring a personal tutor and purchasing more textbooks for different subjects, children's academic achievement may be linked to extra aid from outside school work. On the other hand, a high total income may also relate to parents' educational background, which is another factor that influences children's academic performance. In the next step, I will apply one-way ANOCOVA on our data to test our assumptions.

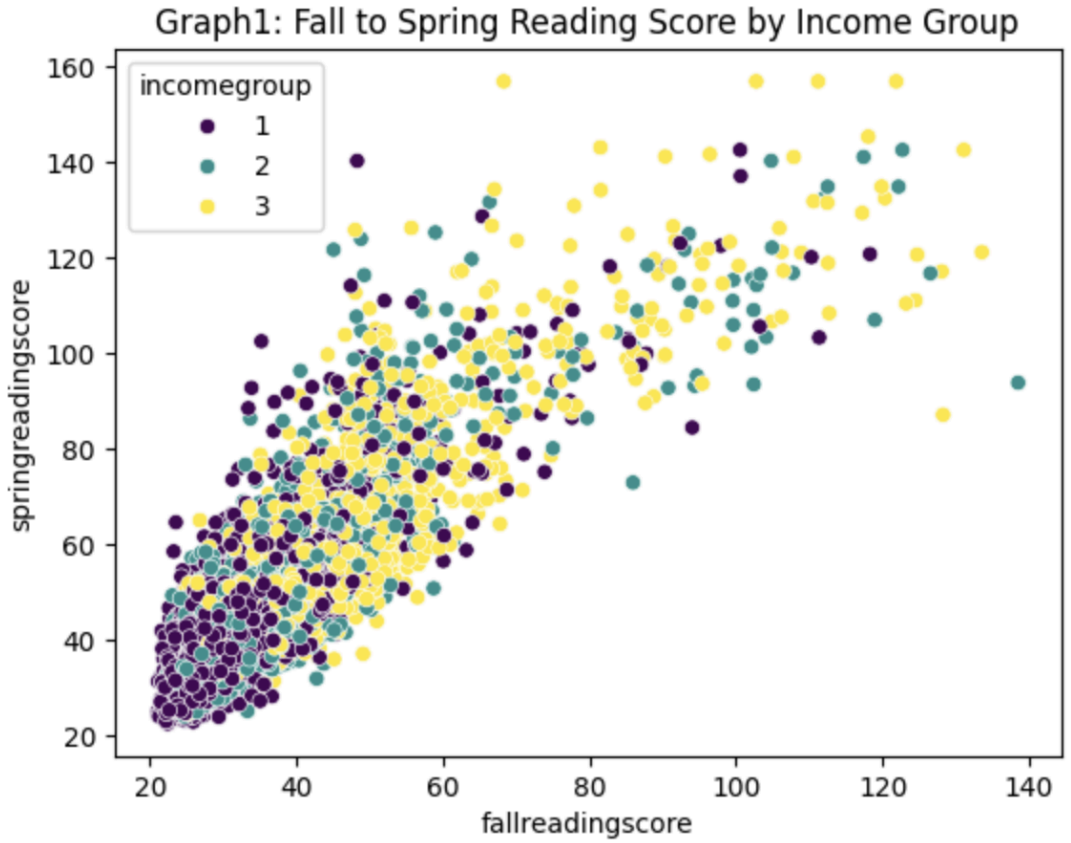
Data cleaning and preparation

There are 11,933 entries with 9 features in our datasets. Since this dataset already be cleanned with no null value, we will skip the data cleaning process. In this research we only following variables:

- **fallreadingscore: Reading score in the fall**
- **fallmathscore: Math score in the fall**
- **fallgeneralknowledgescore: General knowledge score in the fall**
- **springreadingscore: Reading score in the spring**
- **springmathscore: Math score in the spring**
- **springgeneralknowledgescore: General knowledge score in the spring**
- **incomegroup: Categorical income group classification (1, 2, or 3)**

Exploratory Data Analysis

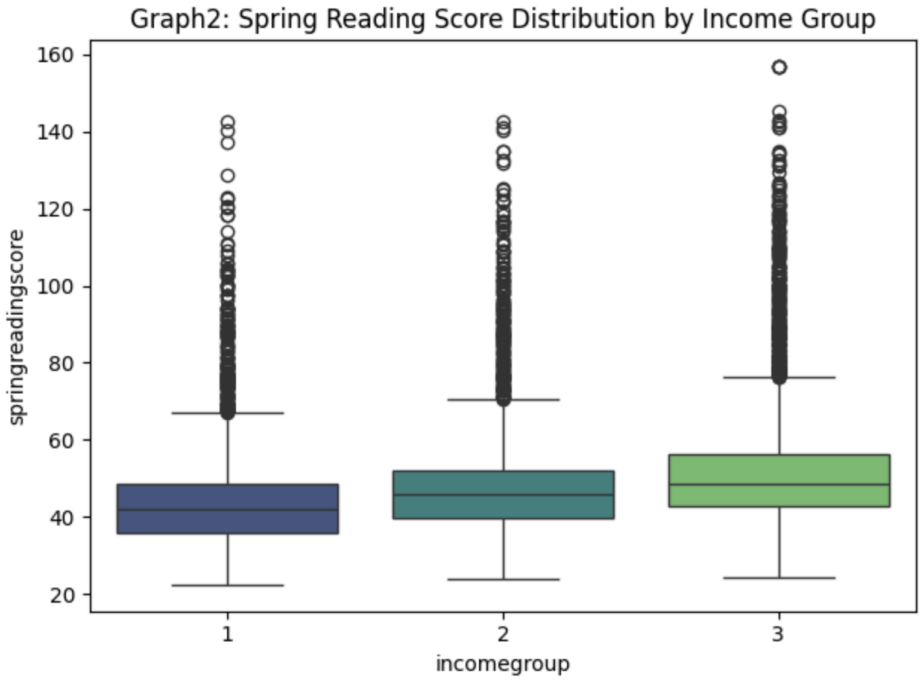
The scatterplot in Graph 1 illustrates the positive relationship between kindergarten children's reading scores in the fall and spring. According to the scatterplot, children's scores improved between fall and spring, although the amounts varied slightly. When we compare the data points from different levels to those for the highest income group—labeled with yellow dots—we discover that they also have the most significant reading scores in the spring. The group with the lowest income, shown in purple, has a lower spring reading score than the next highest income level. These results imply that higher reading scores may be related to income levels. More in-depth research will be conducted to explore whether different income groups positively impact academic achievement.

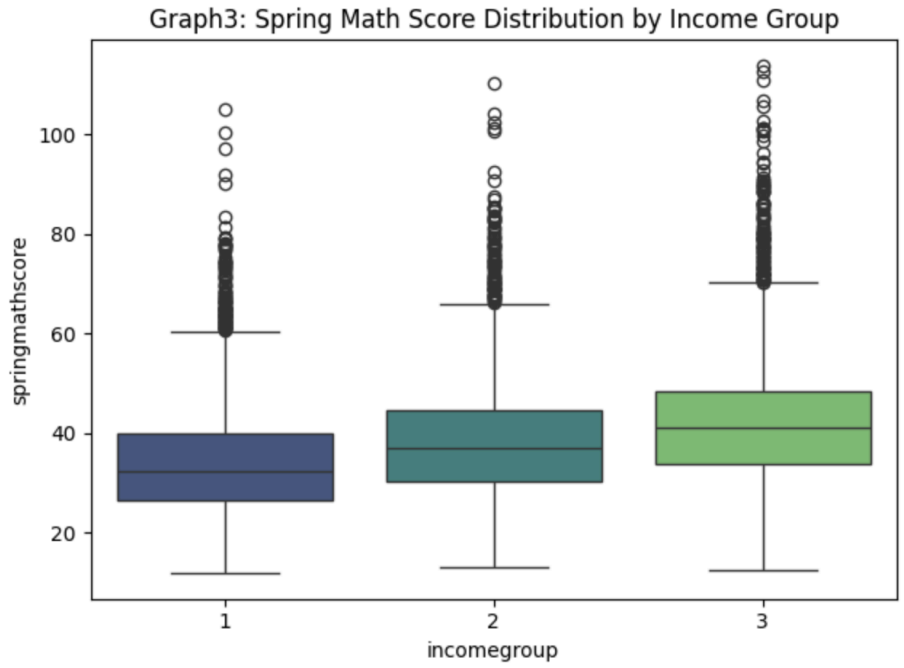


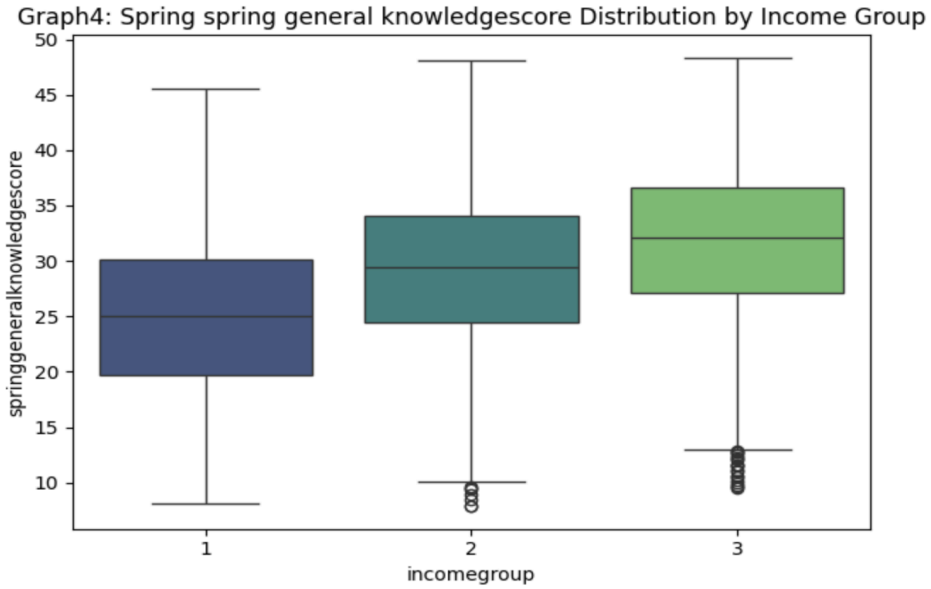
Graph 2 depicts the distribution of spring reading scores by income group. The boxplots reveal that the median score increases with each successive income group, indicating a potential correlation between higher income levels and better reading scores. Notably, the spread and range of scores, as evidenced by the interquartile range (IQR) and whiskers, expand with higher income groups, suggesting increased variability in scores among wealthier families. Outliers are present across all income groups, indicating individual cases of students with scores significantly different from the group's general trend.

Graph 3 illustrates the distribution of spring math scores by income group. Similar to the reading scores, a trend of increasing median scores with higher income groups suggests income may also correlate with math score outcomes. However, the IQRs appear more consistent across groups than reading scores, implying a more uniform distribution of scores within each income level. Like reading scores, outliers are present, which could indicate exceptions to the general pattern or potential areas for further investigation.

Graph 4 shows the distribution of spring general knowledge scores by income group. The median scores are similar across the income groups, with a less noticeable incremental increase than reading and math scores.







Research question 1:

Using one-way ANCOVA, we first controlled for the fall general knowledge limit scores from the prior year before examining the impact of income on spring general knowledge scores. The regression results showed that, due to p-values for both groups being less than 0.001, income group 3 (highest household income) had statistically substantially higher scores on general knowledge than income groups 2 and 1 (lowest household income). After adjusting for baseline general knowledge scores, we have two coefficients, one for Group 2 and one for Group 3, representing the corrected mean differences in spring general knowledge scores compared to Group 1. Furthermore, the model fits the data overall, as evidenced by the high R-squared value of 0.731, suggesting that the model can explain a

significant amount of the variance in spring general knowledge scores. Therefore, the results provide credence to the hypothesis that, within the given period, greater household income levels are linked to more notable general academic advancement.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     springgeneralknowledgescore   R-squared:                       0.731
Model:                                     OLS   Adj. R-squared:                  0.731
Method:                          Least Squares   F-statistic:                 1.082e+04
Date:                         Sat, 23 Mar 2024   Prob (F-statistic):               0.00
Time:                                 13:46:21   Log-Likelihood:                -33259.
No. Observations:                        11933   AIC:                         6.653e+04
Df Residuals:                            11929   BIC:                         6.656e+04
Df Model:                                    3
Covariance Type:                     nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
Intercept                        8.0303      0.119     67.519      0.000       7.797       8.263
C(incomegroup)[T.2]              0.7084      0.088      8.005      0.000       0.535       0.882
C(incomegroup)[T.3]              0.9424      0.094     10.013      0.000       0.758       1.127
fallgeneralknowledgescore        0.8542      0.005    163.347      0.000       0.844       0.864
==============================================================================
Omnibus:                       75.905   Durbin-Watson:                   1.867
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              101.391
Skew:                           0.090   Prob(JB):                     9.62e-23
Kurtosis:                       3.414   Cond. No.                         80.9
==============================================================================

Notes:
```

Research Question 2:

In this study, we once again used the one-way ANCOVA. The regression findings indicate that, over an academic year, household income appears to have a distinct effect on children's academic achievement in arithmetic and reading than general knowledge. In particular, there are notable disparities in reading between students in Income Group 2 and Group 3, with the former having better spring scores by an average of 0.3751 points (p=0.033) and the latter by 0.4898 points (p=0.008). In math, the pattern has comparable p-values, coefficient values, and directions of significance. The gains of 0.7084 points for Income Group 2 and 0.9424 points for Income Group 3, while positive, do not approach statistical relevance to general knowledge (p=0.535 and p=0.758, respectively). Our findings indicate that while household income appears to have a statistically significant impact on reading and math scores, its effect on general knowledge is not statistically significant. Therefore, we can conclude that household income seems to influence academic performance in reading and math but does not affect general knowledge scores.

OLS Regression Results

Dep. Variable:	springreadingscore	R-squared:	0.692
Model:	OLS	Adj. R-squared:	0.692
Method:	Least Squares	F-statistic:	8929.
Date:	Sat, 23 Mar 2024	Prob (F-statistic):	0.00
Time:	13:43:04	Log-Likelihood:	-41675.
No. Observations:	11933	AIC:	8.336e+04
Df Residuals:	11929	BIC:	8.339e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.5430	0.264	24.779	0.000	6.025	7.061
C(incomegroup)[T.2]	0.3751	0.176	2.130	0.033	0.030	0.720
C(incomegroup)[T.3]	0.4898	0.185	2.648	0.008	0.127	0.852
fallreadingscore	1.1322	0.007	156.382	0.000	1.118	1.146

Omnibus:	3968.363	Durbin-Watson:	1.729
Prob(Omnibus):	0.000	Jarque-Bera (JB):	32029.224
Skew:	1.371	Prob(JB):	0.00
Kurtosis:	10.543	Cond. No.	138.

OLS Regression Results

Dep. Variable:	springmathscore	R-squared:	0.681
Model:	OLS	Adj. R-squared:	0.680
Method:	Least Squares	F-statistic:	8469.
Date:	Sat, 23 Mar 2024	Prob (F-statistic):	0.00
Time:	13:46:03	Log-Likelihood:	-39804.
No. Observations:	11933	AIC:	7.962e+04
Df Residuals:	11929	BIC:	7.965e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.2011	0.199	41.273	0.000	7.812	8.591
C(incomegroup)[T.2]	0.6700	0.151	4.430	0.000	0.374	0.966
C(incomegroup)[T.3]	0.9199	0.160	5.741	0.000	0.606	1.234
fallmathscore	1.0735	0.007	149.007	0.000	1.059	1.088

Omnibus:	1782.310	Durbin-Watson:	1.808
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5069.769
Skew:	0.802	Prob(JB):	0.00
Kurtosis:	5.761	Cond. No.	94.8

Conclusion

By applying quantitive analytics and one-way ANCOVA, our result shows that household income significantly improves reading and math scores from fall to spring. Children from higher income groups tend to outperform their peers from the baseline group, suggesting that family income level plays a role in their academic growth in these areas. Our research results support my initial assumption that children from high-income families will have higher scores at the beginning and score growth. In future studies, we should consider adding more variables, such as specific income groups and racial or gender groups, to conduct more in-depth analyses to understand how these factors interact with academic achievement.