Rosa Lee
Assignment 3

Exploring Kindergarten Student Performance Trends

Introduction:

Early childhood education plays a vital role in laying the foundation for academic success. This report analyses Kindergarten students' performance using a subset of data from an early child longitudinal study conducted during the academic year 1998-1999. The dataset, titled "INF2178 A3 data.csv," includes fall and spring measurements of reading, math, and general knowledge scores, along with information on household income. The objective is to explore how students' reading and math scores evolve over time, considering income group differences. The research question this paper will explore are the following:

- How do fall reading and math scores, along with household income, predict spring reading and math scores among students?

Observations and Data Cleaning:

Upon initial examination, the dataset comprises six continuous variables: fall/spring reading, math, and general knowledge scores, total household income, income in thousands, and income group.

The latter is collected from total household income, categorizing it into income groups.The data cleaning process involved loading the dataset and checking for missing values, while the exploratory data analysis included understanding the dataset's structure, computing summary statistics, and visualizing the distribution of income groups.

Exploratory Data Analysis (EDA):

Upon examining the dataset, we focused on further analysing columns such as fall reading and math scores, spring reading and math scores, and income groups. The dataset consists of 11,933 observations with no missing values.

	fallreadingscore	fallmathscore	fallgeneralknowledgescore	springreadingscore	springmathscore	springgeneralknowledgescore	totalhouseholdincome	incomeinthousands	incomegroup
count	11933	11933.000000	11933.000000	11933.000000	11933.000000	11933.000000	11933.000000	11933.000000	11933.000000
mean	35.954215	27.128244	23.073694	47.511178	37.799461	28.235584	54317.199930	54.3172	1.895165
std	10.473130	9.120505	7.396978	14.327111	12.027753	7.577457	36639.061147	36.639061	0.822692
min	21.010000	10.510000	6.985000	22.350000	11.900000	7.858000	1	0.001	1
25%	29.340000	20.680000	17.385000	38.950000	29.270000	22.802000	27000	27	1
50%	34.060000	25.680000	22.954000	45.320000	36.410000	28.583000	47000	47	2
75%	39.890000	31.590000	28.305000	51.770000	44.220000	33.782000	72000	72	3

max	138.510000	115.650000	47.691000	156.850000	113.800000	48.345000	150000	150	3

Figure 1. Dataset Qualitative Data Statistics

The relevant columns for the analysis:fallreadingscore: Fall reading scores

- fallmathscore: Fall math scores
- springreadingscore: Spring reading scores
- springmathscore: Spring math scores
- incomegroup: Income group of the household

Descriptive statistics insights:

- Fall reading scores range from 21.01 to 138.51, with a mean of approximately 35.95.
- Fall math scores range from 10.51 to 115.65, with a mean of around 27.13.
- Spring reading scores range from 22.35 to 156.85, with a mean of roughly 47.51.
- Spring math scores range from 11.90 to 113.80, with a mean of about 37.80.
- Income groups range from 1 to 3, with a mean of approximately 1.90.

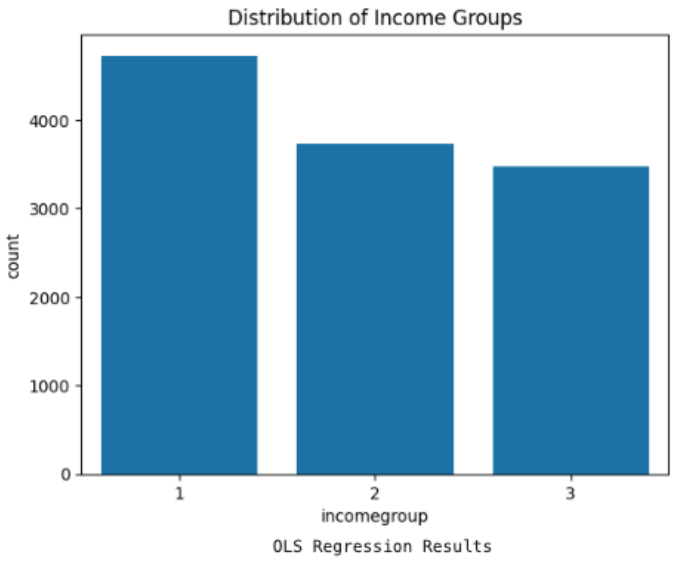

Fig 2. Bar chart for Distribution of income groups

The above Discribution of Income Groups chart groups indicates that there is representation across all three income groups, with a higher concentration in the lower income groups.

One-way ANCOVA test:

The correlation analysis showed strong positive relationships between fall and spring scores for reading, math, and general knowledge, suggesting consistency in students' performance over time.ANCOVA Analysis. There were two separate one-way ANCOVAs conducted to analyze the effects of fall reading/math scores and income on spring reading/math scores.

Model 1: Spring Reading Score ~ Fall Reading Score + Income

OLS Regression Results

Dep.Variable:	springreadingscore	R-squared:	0.692
Model:	OLS	Adj. R-squared:	0.692
Method:	Least Squares	F-statistic:	1.339e+04
Date:	Thu, 21 Mar 2024	Prob (F-statistic):	0.00

Time:	12:43	Log-Likelihood:	-41675
No.Observations:	11933	AIC:	8.336e+04
DF Residuals	11930	BIC:	8.338e+04
DF Model:	2	Covariance Type:	nonrobust

	coef	Std error	t	P>|t|	[0.025	0.975]
Intercept	6.3272	0.280	22.628	0.000	5.779	6.875
fallreadingscore	1.1322	0.007	156.383	0.000	1.118	1.146
incomegroup	0.2512	0.092	2.726	0.006	0.071	0.432

Omnibus:	3965.793	Durbin-Watson:	1.729
Prob(omnibus):	0.00	Jarque-Bera(JB):	31979.5
Skew:	1.371	Prob(JB):	0.00
Kurtosis:	10.537	Cond. No.	145

The ANCOVA model for predicting spring reading scores based on fall reading scores and income group yielded the following results:

- R-squared: 0.692, indicating a moderately strong fit for the model.
- Both fall reading scores and income group were significant predictors of spring reading scores ($p < 0.001$).
- The coefficient for fall reading scores was 1.1322, suggesting that for every one-unit increase in fall reading score, spring reading score is expected to increase by approximately 1.1322 units.
- The coefficient for income group was 0.2512, indicating that higher income groups were associated with slightly higher spring reading scores.

Model 2: Spring Math Score ~ Fall Math Score + Income

OLS Regression Results

Dep.Variable:	springmathscore	R-squared:	0.680
Model:	OLS	Adj. R-squared:	0.680
Method:	Least Squares	F-statistic:	1.270e+04
Date:	Thu, 21 Mar 2024	Prob (F-statistic):	0.00
Time:	12:43	Log-Likelihood:	-39805
No.Observations:	11933	AIC:	7.962e+04
DF Residuals	11930	BIC:	7.964e+04

DF Model:	2	Covariance Type:	nonrobust

	coef	Std error	t	P>|t|	[0.025	0.975]
Intercept	7.7862	0.215	36.286	0.000	7.366	8.207
fallreadingscore	1.0735	0.07	129.009	0.00	1.059	1.088
incomegroup	0.4699	0.080	5.884	0.000	0.313	0.626

Omnibus:	1782.878	Durbin-Watson:	1.808
Prob(omnibus):	0.00	Jarque-Bera(JB):	5083.773
Skew:	0.802	Prob(JB):	0.00
Kurtosis:	5.766	Cond. No.	100

Similarly, the ANCOVA model for predicting spring math scores based on fall math scores and income group yielded the following results:R-squared: 0.680, indicating a moderately strong fit for the model.

- Both fall math scores and income group were significant predictors of spring math scores ($p < 0.001$).
- The coefficient for fall math scores was 1.0735, suggesting that for every one-unit increase in fall math score, spring math score is expected to increase by approximately 1.0735 units.
- The coefficient for income group was 0.4699, indicating that higher income groups were associated with slightly higher spring math scores.

Analysis Result:

Using one-way ANCOVAs, the impact of fall scores and income groups on spring reading and math scores separately were assessed.

For reading scores, the ANCOVA yielded a significant effect of both fall reading scores ($\beta = 1.1322$, $p < 0.001$) and income group ($\beta = 0.2512$, $p = 0.006$) on spring reading scores. This indicates that higher fall reading scores and belonging to a higher income group are associated with higher spring reading scores.

Similarly, for math scores, the ANCOVA revealed significant effects of fall math scores ($\beta = 1.0735$, $p < 0.001$) and income group ($\beta = 0.4699$, $p < 0.001$) on spring math scores. Again, higher fall math scores and belonging to a higher income group correspond to higher spring math scores.

Interaction Plots

Interaction plots were created to visualize the relationship between fall scores, income groups, and spring scores. These plots illustrate how the effect of fall scores on spring scores varies across income groups.

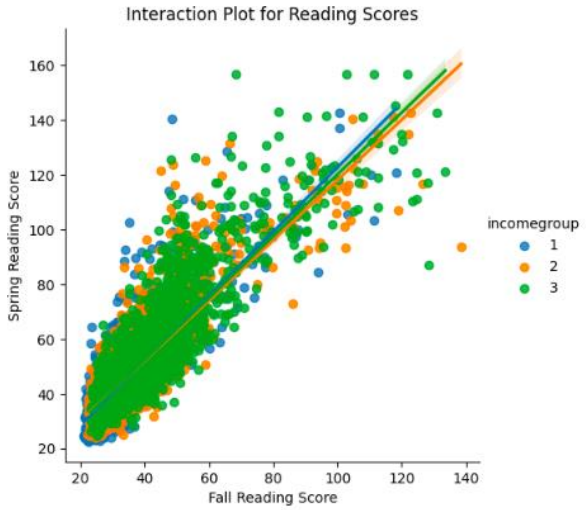

Figure 3. Interaction Plot for Reading Scores

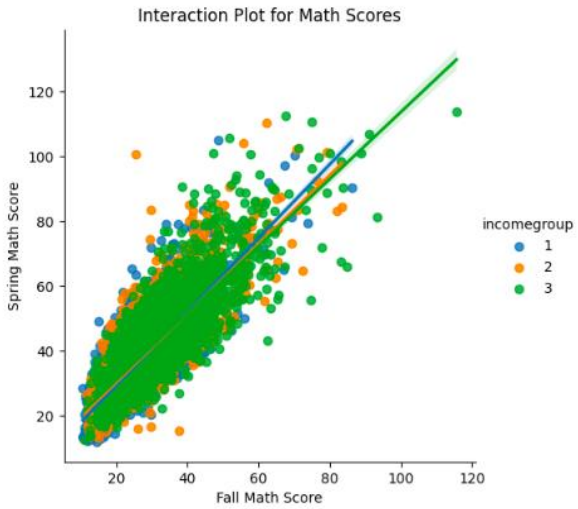

Figure 4. Interaction Plot for Math Scores

The interaction plot for reading scores illustrates how the relationship between fall reading scores and spring reading scores varies across different income groups. Similarly, the interaction plot for math scores demonstrates the relationship between fall math scores and spring math scores across different income groups.

1. The plots corroborate the statistical evidence, illustrating a marked positive correlation between fall and spring scores in both reading and math.
2. The gradient of improvement from fall to spring differs slightly by income group, with higher income levels generally translating into steeper ascents, implying that socioeconomic status plays a role in academic growth rates. These visual depictions echo the ANCOVA findings where income group coefficients were positive.
3. The spread of data points around the regression lines also sheds light on the within-group variability, hinting at a multitude of factors at play beyond income, which could include access to resources, parental involvement, or innate abilities.
4. The visual plots emphasize that while income is a clear factor, it is not the sole determinant of academic progress, as seen by the overlap between the income groups.

Assumptions Testing

An assumptions test was conducted for the ANCOVA model to ensure the validity of results.

1. **Normal Distribution Departure:** The deviation of the residuals from the straight line suggests that the assumption of normality may not hold for the residuals
2. **Left and Right Tail Deviations:** The residuals exhibit deviations from the expected normal distribution at both ends. The lower tail shows larger negative residuals, while the upper tail shows larger positive residuals, indicating possible heavy tails and non-normality
3. **Possible Heavy Tails and Outliers:** The pattern of deviations suggest the presence of outliers or extreme values in the residuals, which could impact the validity of statistical inferences

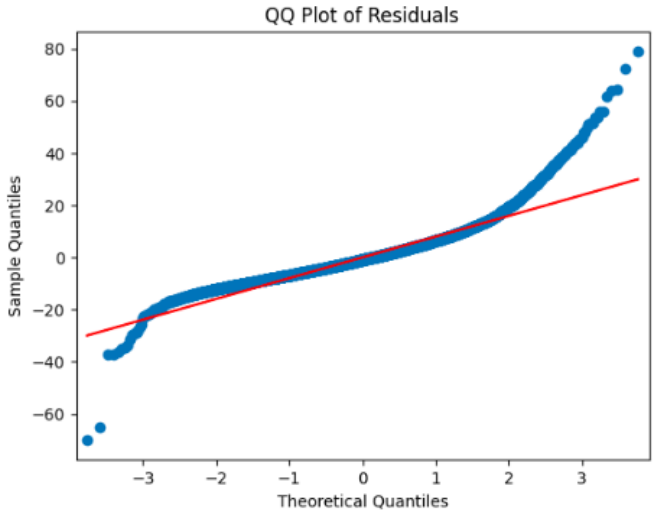

Figure 5. QQ Plot for residuals

4. **Potential for non-normality**: The overall departure from the red line indicates a potential departure from normality, which is a crucial assumption in many statistical models, including linear regression and ANCOVA
5. **Model Fit Implications**: non normal residuals could indicate issues with model fit, potentially leading to biased parameter estimates and inaccurate statistical tests and confidence intervals

Conclusion

The analysis of kindergarten student performance trends provide valuable insights into the factors influencing academic outcomes, particularly in reading and math, while considering the influence of household income.

The ANCOVA models revealed significant effects of both fall scores and income groups on spring scores, indicating that higher fall scores and belonging to higher income groups are associated with higher spring scores in both reading and math. Specifically, for reading scores, every one-unit increase in fall reading score is associated with approximately a 1.13 unit increase in spring reading score, while belonging to a higher income group is linked to slightly higher spring reading scores. Similarly, for math scores, increases in fall math scores and belonging to higher income groups correspond to higher spring math scores.

The interaction plots provided visual confirmation of these findings, illustrating a positive correlation between fall and spring scores in both reading and math. Furthermore, the plots highlighted variations in academic growth rates across income groups, with higher income levels generally associated with steeper ascents from fall to spring.

Additionally, assumptions testing through QQ plots for residuals identified departures from normality, suggesting potential issues with model fit. The presence of heavy tails and outliers in the residuals could impact the validity of statistical inferences, emphasizing the need for further investigation.

In conclusion, while the ANCOVA results underscore the significant influence of fall scores and income groups on spring scores, the visualizations and assumptions testing indicate potential limitations in the modeling approach. Addressing these limitations through robust sensitivity analyses and exploring alternative modeling techniques will enhance the validity and reliability of the findings. Ultimately, this analysis revealed the complex interplay of academic performance and socioeconomic factors.