Understanding Academic Performance Disparities Among Kindergarten Students:
A One-Way ANCOVA Analysis

Danika Mariam (1004014880)

Introduction

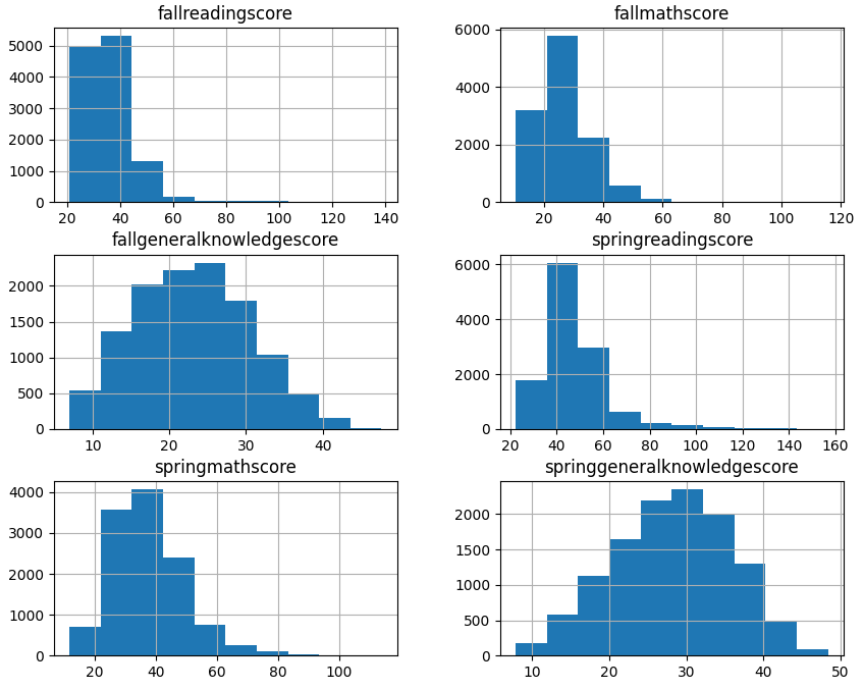
In this analysis, we explore the relationship between students' academic performance and their household income levels using data from a longitudinal study conducted during the 1998-1999 academic year. We aim to understand whether there are significant differences in fall and spring reading, math, and general knowledge scores among kindergarten students across different income groups, while controlling for total household income.

The research question we aim to answer in this paper is as follows: are there significant differences in fall and spring reading, math, and general knowledge scores among kindergarten students based on their household income levels?

Exploratory Data Analysis (EDA)

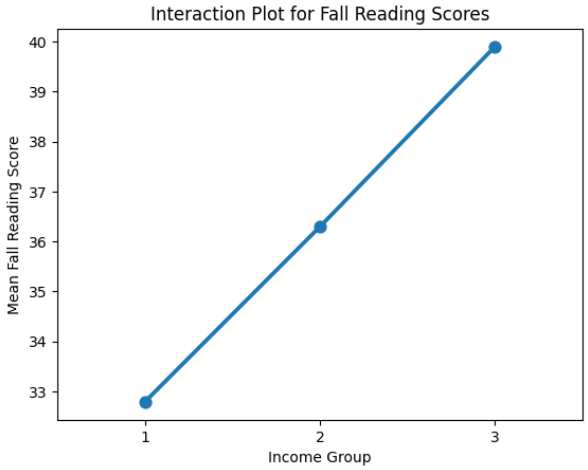
The dataset comprises six continuous variables representing fall and spring reading, math, and general knowledge scores, along with total household income and income groups categorized into three levels. Exploratory data analysis revealed that students' fall and spring scores exhibited a wide range of values, indicating variance in academic performance. Additionally, the distribution of income groups showed that there were more students in the lower income categories compared to higher ones. Histograms were made for each of the variables.

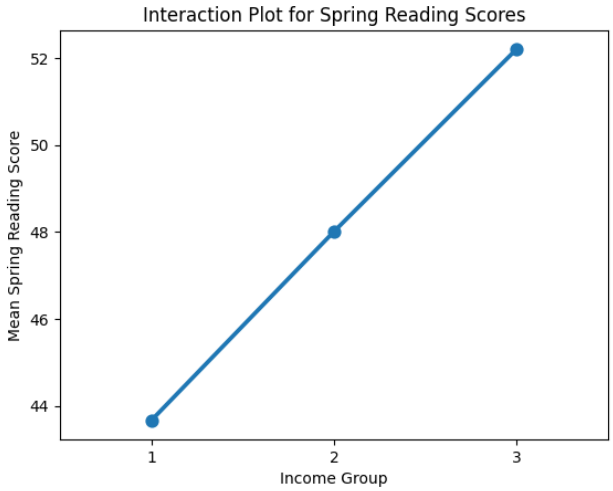
Figure 1: Histograms

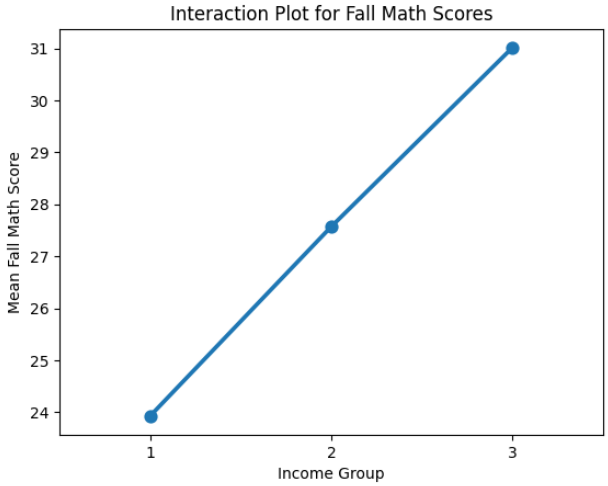


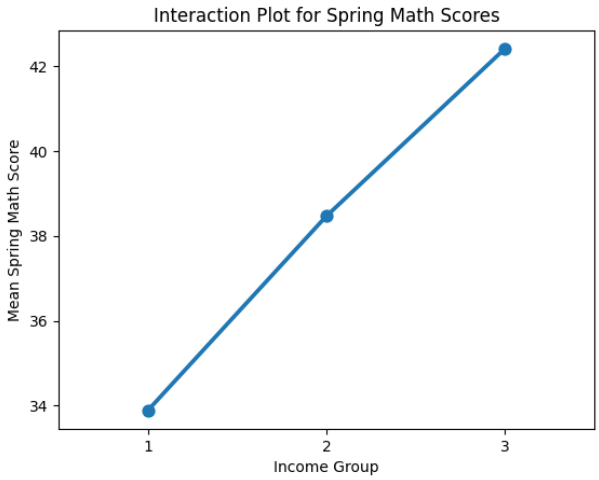
Interaction Plots

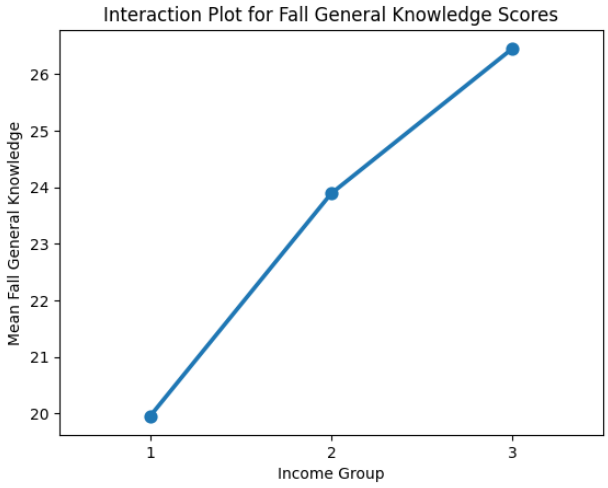
Interaction plots were created to visualize the relationship between income groups and fall/spring scores. These plots illustrate how the mean scores vary across different income groups, providing insights into potential disparities in academic performance.

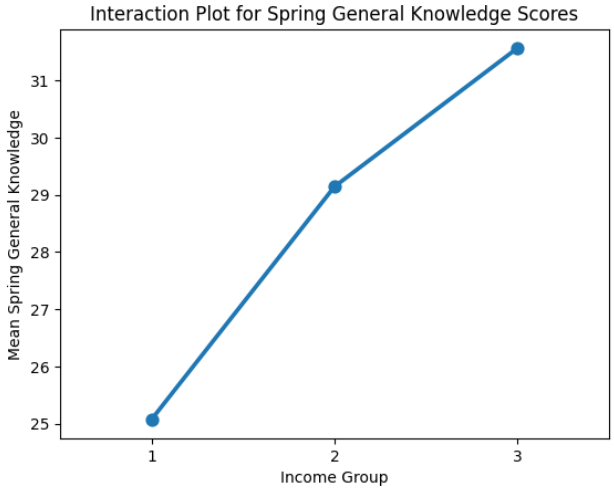












ANCOVA Results

Separate one-way ANCOVAs were conducted for combined reading, math, and general knowledge scores while controlling for total household income. The results revealed significant differences in combined academic performance among income groups, indicating that socioeconomic status influences students' overall academic achievements over time.

Model 1: Household Income effect on Reading Score

Dependant Variable:				combined_reading_score		
R-squared:				0.086		
Adj. R-squared:				0.086		
F-statistic:				374.6		
Prob (F-statistic):				1.66e-232		
Log-Likelihood:				-45927		
AIC:				9.186e+04		
BIC:				9.189e+04		
Covariance Type :				nonrobust		
	coef	std err	t	p	[0.025	0.975]
Intercept	36.5058	0.213	171.637	< 0.001	36.089	36.923
C(incomegroup)[T.2]	1.6031	0.308	5.211	< 0.001	1.000	2.206
C(incomegroup)[T.3]	1.6576	0.543	3.050	0.002	0.592	2.723
totalhouseholdincome	7.812e-05	6.09e-06	12.837	< 0.001	6.62e-05	9e-05
Omnibus:	7028.444		Durbin-Watson:		1.788	
Prob(Omnibus):	< 0.001		Jarque-Bera (JB):		84986.507	
Skew:	2.633		Prob(JB):		< 0.001	
Kurtosis:	14.966		Cond. No:		3.73e+05	

Model 2: Household Income effect on Math Score

Dependant Variable:				combined_math_score		
R-squared:				0.120		
Adj. R-squared:				0.119		
F-statistic:				539.9		
Prob (F-statistic):				< 0.001		
Log-Likelihood:				-43777		
AIC:				8.756e+04		
BIC:				8.759e+04		
Covariance Type :				nonrobust		
	coef	std err	t	p	[0.025	0.975]
Intercept	27.1578	0.178	152.892	< 0.001	26.810	27.506
C(incomegroup)[T.2]	1.7560	0.257	6.836	< 0.001	1.252	2.260
C(incomegroup)[T.3]	1.5469	0.454	3.409	0.001	0.657	2.437
totalhouseholdincome	7.929e-05	5.08e-06	15.602	< 0.001	6.93e-05	8.93e-05
Omnibus:	2687.117		Durbin-Watson:		1.774	

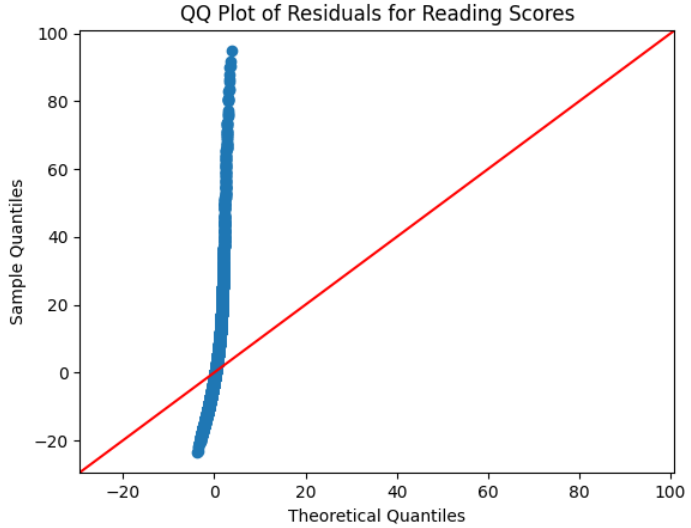
Prob(Omnibus):	< 0.001	Jarque-Bera (JB):	7597.212
Skew:	1.187	Prob(JB):	< 0.001
Kurtosis:	6.106	Cond. No:	3.73e+05

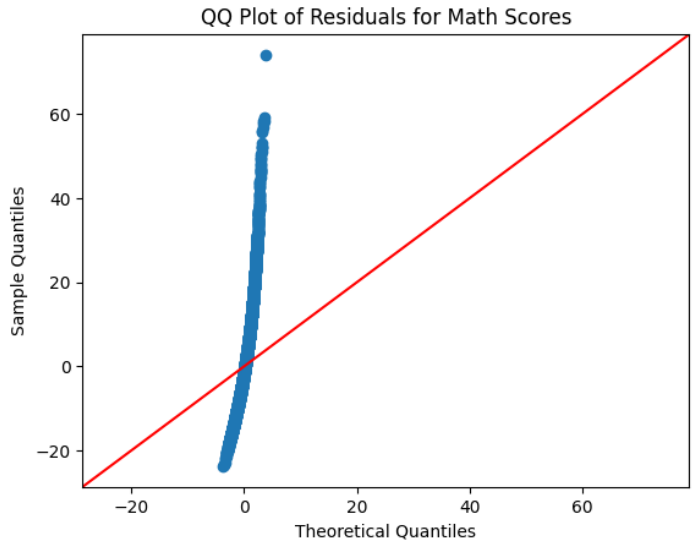
Model 3: Household Income effect on General Knowledge Score

Dependant Variable:				combined_general_knowledge_score		
R-squared:				0.161		
Adj. R-squared:				0.161		
F-statistic:				765.2		
Prob (F-statistic):				< 0.001		
Log-Likelihood:				-39452		
AIC:				7.891e+04		
BIC:				7.894e+04		
Covariance Type :				nonrobust		
	coef	std err	t	p	[0.025	0.975]
Intercept	21.2530	0.124	171.925	< 0.001	21.011	21.495
C(incomegroup)[T.2]	2.3123	0.179	12.933	< 0.001	1.962	2.663
C(incomegroup)[T.3]	1.9982	0.316	6.326	< 0.001	1.379	2.617
totalhouseholdincome	5.702e-05	3.54e-06	16.122	< 0.001	5.01e-05	6.4e-05
Omnibus:	100.903		Durbin-Watson:		1.678	
Prob(Omnibus):	< 0.001		Jarque-Bera (JB):		67.235	
Skew:	0.025		Prob(JB):		2.51e-15	
Kurtosis:	2.636		Cond. No:		3.73e+05	

Assumptions Testing

Using a QQ plot (Quantile-Quantile plot) of residuals is a common method to visually assess the normality assumption of residuals in regression analysis, including ANCOVA. We created QQ plots for the residuals of each ANCOVA model to evaluate the normality assumption.





While attempting to assess the normality of residuals using QQ plots, we observed that the residuals did not precisely align with the diagonal line, indicating a departure from normality. This deviation from normality raises concerns about the reliability of the ANCOVA results, as statistical inferences drawn from the analysis may be affected. Additionally, the presence of skewness, kurtosis, or outliers in the distribution of residuals could further complicate the situation, potentially undermining the validity of the statistical tests conducted.

Conclusion

In conclusion, our analysis highlights the existence of academic performance disparities among kindergarten students based on their household income levels. These findings indicate that socioeconomic inequalities can lead to different educational scores/outcomes. Further research could delve deeper into exploring the underlying

factors contributing to these disparities and developing targeted interventions to mitigate them. The ANCOVA findings suggest that children form higher income households tend to perform better on reading, math, and general knowledge tests.

In considering the limitations of this analysis, it is crucial to address the observed deviation of residuals from the diagonal line in the QQ plot, suggesting a departure from the normality assumption. This departure poses a significant limitation as it casts uncertainty on the reliability of the ANCOVA findings. Non-normality in the residuals may introduce bias into the estimates and affect the validity of statistical inferences drawn from the analysis. Furthermore, the presence of skewness, kurtosis, or outliers in the distribution of residuals could further compound the limitations, potentially confounding the interpretation of results. Thus, future research should prioritize addressing the non-normality of residuals through alternative modeling approaches or data transformations to ensure the robustness and validity of the findings. Additionally, employing longitudinal designs and qualitative methodologies could offer complementary insights into the dynamics of academic performance disparities and contribute to the development of more comprehensive interventions to mitigate socioeconomic inequalities in educational outcomes.