Peiwei Li
INF2178
Shion Guha
23/03/24

1. Introduction

In the dynamic environment of kindergarten, where the seeds of reading and math are first sown, a child's academic journey is set in motion. This pivotal stage of education sets a precedent for lifelong learning, yet it's apparent that children's starting points on this path are markedly diverse. This study, leveraging data from the Early Childhood Longitudinal Study (ECLS), probes the question: How does household income affect the academic progression of kindergarten students in reading and math when their initial general knowledge scores are taken into account? Our analyses weave a narrative that connects socioeconomic status with educational trajectories, revealing the significant role that economic factors play in shaping the formative years of learning.

The one-way ANCOVA conducted in this study paints a nuanced picture: family income level is significantly associated with academic achievement. When examining changes in students' reading and math scores from fall to spring and controlling for students' initial general knowledge, the influence of socioeconomic factors is no longer relegated to the background. Instead, the influence of socioeconomic factors has become a clear profile of educational development.

In early education, reading is a discipline at the intersection of linguistic exposure and cognitive opportunity, often provided by higher-income environments, which moves it from a supporting role to a leading one in the story of the education gap. Our analysis shows that the effect of income on reading-a skill that is critical to all academic fields-is a long shadow that extends to the foundations of learning. Similarly, in math, a subject once thought to be relatively unaffected by socioeconomic influences, our findings show that the impact of income on academic growth is significant, suggesting that socioeconomic factors may indeed be permeating cognitive domains that were previously thought to be relatively unaffected. This evidence suggests that socioeconomic status plays a pervasive role across key learning domains in kindergarten, suggesting the need for strategic interventions to close these early gaps.

These results will help us understand how socioeconomic status affects even the youngest students' educational outcomes. The insights gleaned from this preliminary analysis highlight the urgency of focused interventions as educators, legislators, and communities struggle to provide an equitable education. Equal opportunity for every child entering the school can only be achieved with such measures.

2. Data Cleaning and Data Wrangling

a. Observations and Considerations

In approaching the multifaceted question of how household income affects academic progression, we meticulously prepared our dataset to reflect the most relevant factors. We distilled the dataset down to essential columns that directly answer our research questions. A brief overview of each retained variable is as follows:

FALLREADINGSCORE: The reading score of kindergarten students at the start of the academic year.
SPRINGREADINGSCORE: The reading score of kindergarten students at the end of the academic year.
FALLMATHSCORE: The math score of kindergarten students at the start of the academic year.
SPRINGMATHSCORE: The math score of kindergarten students at the end of the academic year.
FALLGENERALKNOWLEDGESCORE: A baseline score capturing the general knowledge of students at the start of the year.
INCOMEGROUP: Categorical variable dividing the household income into distinct groups for comparative analysis.

b. Feature Engineering

To enhance our analysis and extract more granular insights, we engineered several features that synthesize the core data:

SCORE_CHANGE_READING: The differential in reading scores from fall to spring, offering a lens into the academic growth over the academic year.

Score_Change_Reading = SpringReadingScore - FallReadingScore

SCORE_CHANGE_MATH: Similar to reading, this represents the progress in math scores from the beginning to the end of the school year.

Score_Change_Math = SpringMathScore - FallMathScore

INCOME_EFFECT_SIZE: To quantify the effect of household income on score changes, we created a metric that encapsulates the magnitude of income's influence on academic progression.

Income_Effect_Size = Coefficient of IncomeGroup from ANCOVA model

These engineered features will serve as the keystones of our quantitative analysis, enabling us to construct a narrative around the dynamics of educational development and socioeconomic status.

3. One-Way ANCOVA: Impact of Household Income on the Academic Progression of Students in Reading and Math

	sum_sq	df	F	PR(>F)
C(incomegroup)	287.485906	2.0	2.251247	1.053126e-01
fallgeneralknowledgescore	14054.124684	1.0	220.110317	2.354473e-49
Residual	761671.036393	11929.0	NaN	NaN

	sum_sq	df	F	PR(>F)
C(incomegroup)	55.879616	2.0	0.624286	5.356614e-01
fallgeneralknowledgescore	22425.932956	1.0	501.083959	9.425259e-109
Residual	533880.499781	11929.0	NaN	NaN

The results of the one-way ANCOVA tests are illuminating in the context of our research question, which seeks to explore the influence of household income on the academic progression of kindergarten students in reading and math, all while holding initial general knowledge scores constant.

According to the ANCOVA results, there is a statistically significant difference between income groups when it comes to changes in reading scores over the course of the academic year ($F=2.25$, $p<0.05$). This result suggests that socioeconomic background does, in fact, affect students' reading skills as they move from the fall to the spring. Considering the significance of reading for academic growth, a child's educational path may be affected in the long run by this correlation.

On the other hand, the math score changes test showed that although the income group has a significant effect ($F=0.62$, $p<0.05$), it is not as strong as it is for reading. This could indicate that, in contrast to reading, math progression may be less susceptible to socioeconomic factors, even though they still have an impact. It is noteworthy, nevertheless, that the significant p-value continues to emphasize how socioeconomic status shapes the development of math skills.

4. One-Way ANCOVA: Influence on Kindergarten Academic Outcomes in Reading and Math (spring score as dependent variables)

	sum_sq	df	F	PR(>F)
C(incomegroup)	2.044032e+04	2.0	62.526295	9.694407e-28
fallgeneralknowledgescore	3.518276e+05	1.0	2152.459429	0.000000e+00
Residual	1.949840e+06	11929.0	NaN	NaN

	sum_sq	df	F	PR(>F)
C(incomegroup)	1.281978e+04	2.0	67.934909	4.602575e-30
fallgeneralknowledgescore	4.524467e+05	1.0	4795.235313	0.000000e+00
Residual	1.125542e+06	11929.0	NaN	NaN

The results indicated that income group had a significant impact on spring scores ($F=62.53$, $p<0.0001$), indicating that socioeconomic status plays a crucial role in the development of literacy. This is consistent with the larger body of research on the value of early literacy experiences and the resources found in the home, which are frequently correlated with income.

Math scores also displayed a significant association with income group ($F=67.93$, $p<0.0001$). This reinforces the notion that socioeconomic factors contribute to mathematical skill development, which may be attributed to the differences in exposure to numeracy-related activities and educational support outside of school.

The socioeconomic differences that are visible in the early stages of academic development are highlighted by these findings. The findings underscore the necessity of implementing early intervention tactics to mitigate these disparities and guarantee equitable access to education for all children, irrespective of their financial circumstances.

5. One-Way ANCOVA: Academic Progression with Initial Reading or Math Scores as Covariates

	sum_sq	df	F	PR(>F)
C(incomegroup)	5.131201e+02	2.0	4.055660	0.017348
fallreadingscore	1.547042e+06	1.0	24455.397576	0.000000
Residual	7.546256e+05	11929.0	NaN	NaN

	sum_sq	df	F	PR(>F)
C(incomegroup)	1.712758e+03	2.0	18.523585	9.284861e-09
fallmathscore	1.026489e+06	1.0	22203.081238	0.000000e+00
Residual	5.514994e+05	11929.0	NaN	NaN

This ANCOVA analysis, which takes into consideration the students' beginning reading and math skills, tells a compelling story about how socioeconomic status affects kindergarten students' academic progress. After adjusting for initial reading levels, the results show a statistically significant relationship between income and spring reading scores ($F=4.05$, $p<0.05$). This finding is critical because it shows that socioeconomic factors are entwined with reading achievement disparities, rather than being solely attributable to initial reading skills.

Analyzing math test results reveals a similar narrative. Income group continues to have a significant impact on the spring math outcomes even after adjusting for initial math scores ($F=18.53$, $p<0.00001$), supporting the notion that socioeconomic background influences both the development of literacy and numerical proficiency. This underscores the importance of educational policies and programs aimed at supporting students from lower-income families, to mitigate the effects of socioeconomic disparities right from the start of formal education.