Predicting School Performance from Socioeconomic Factors

Abstract:

This study was conducted to examine whether average ACT scores can be predicted based on socioeconomic factors. The data used in this research were obtained from EdGap.org and the National Center for Education Statistics (NCES). Several models were developed to explore the relationship between academic performance and socioeconomic indicators such as income, unemployment rate, educational attainment, and family structure.

The results show that socioeconomic variables can explain about 63% of the variation in ACT scores, with the percentage of students receiving free or reduced-price lunch contributing the most and emerging as the strongest predictor, having a much higher predictive power than the other variables. To improve model accuracy, data on neighborhood poverty levels surrounding schools were incorporated. Although the improvement in accuracy was minimal, the results suggest that the economic environment around schools also influences educational outcomes, adding value to the model beyond the factors observed within the schools themselves.

Introduction:

Socioeconomic disparities have long been recognized as influencing academic outcomes. Schools located in areas with lower income levels or higher poverty rates often face major challenges such as limited resources, lower education quality, and reduced student engagement in learning programs. These factors can directly affect how well students are prepared for standardized tests such as the ACT. This study was conducted to examine whether socioeconomic indicators can be used to predict average ACT scores.

The analysis uses data from EdGap.org, which includes statistics related to socioeconomic factors, combined with datasets from the National Center for Education Statistics (NCES) that provide additional information about each school. To better reflect broader community conditions, the study also incorporates neighborhood poverty rates from the NCES School Neighborhood Poverty Estimates. The objective of this research is to identify how socioeconomic variables strongly predict ACT scores and to evaluate whether including neighborhood-level information can improve the model's accuracy.

Methodology:

This study merged data using school identifiers from EdGap and NCES school information. The completed dataset includes variables such as median income, unemployment

rate, educational attainment, percentage of students living in married households, and percentage of students receiving free or reduced-price lunch, along with a charter school. In addition, data from the NCES School Neighborhood Poverty Estimates provide the poverty ratio of the community surrounding each school.

Before analysis, the data were reviewed to identify missing or invalid values. Unreasonable ACT scores (less than 1) and impossible free-lunch percentages (less than 0) were removed. Missing values in the ACT variable were deleted because the missing count was very small and would not significantly affect the analysis. The remaining missing values in other variables were imputed using the Iterative Imputer method. The data were then analyzed using multiple linear regression models to identify which socioeconomic variables best explain differences when predicting average ACT scores.

Several models were developed to provide multiple perspectives for comparison. The first was a simple model that used only one predictor at a time to compare each socioeconomic variable with the average ACT score. The second was a multiple model that included all socioeconomic factors simultaneously to capture their combined effects. Finally, an extended model was constructed by adding the neighborhood poverty rate to evaluate its additional contribution to predicting ACT scores. This extended model also tested whether neighborhood poverty could potentially replace median income as a predictor, since both variables represent similar aspects of socioeconomic conditions.

Result:

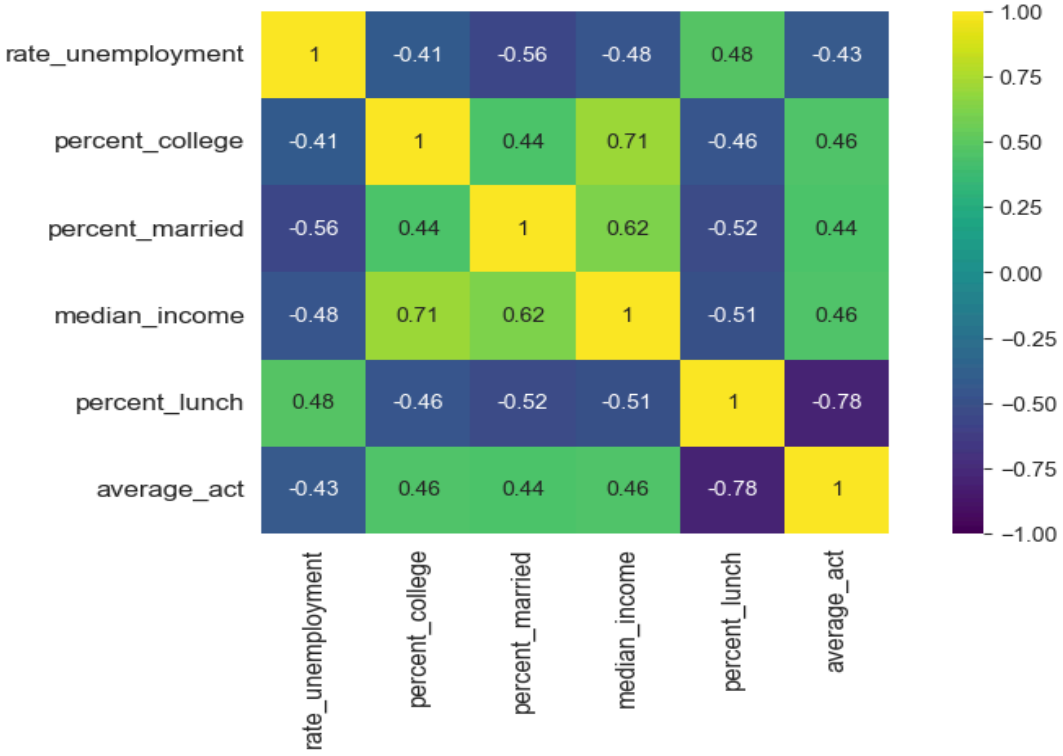

Figure 1. Correlation matrix of socioeconomic factors and average ACT scores.

The analysis shows that the percentage of students receiving free or reduced-price lunch varies widely across schools and exhibits the strongest negative relationship with ACT scores, with a correlation coefficient of –0.78. Schools with lower socioeconomic conditions generally produce lower average ACT scores.

Regression analysis confirms that socioeconomic variables collectively explain part of the variation in academic performance among schools. A model including five variables including unemployment rate, household income, percentage of adults with college degrees, percentage of students living in married households, and percentage of students receiving free or reduced-price lunch, explains about 63% of the variation in ACT scores.

Adding the neighborhood poverty rate slightly but meaningfully improves the model's fit when used as a substitute for household income, indicating that economic hardship in the surrounding community provides better predictive value for ACT performance.

Conclusion:

This study shows that socioeconomic factors are influential and strong predictors of schools' average ACT scores. Schools with higher proportions of economically disadvantaged students tend to have lower scores, and models combining multiple socioeconomic indicators can explain about 63% of the variation in academic outcomes. Adding a measure of neighborhood-level poverty slightly improves the model's predictive power, suggesting that the surrounding environment may also affect academic performance.

These results emphasize that the causes and impacts on academic achievement extend beyond the classroom and should involve addressing broader socioeconomic issues. Future research could expand on these findings by incorporating additional factors such as school funding, in order to build a more comprehensive understanding of how economic context influences students' learning outcomes.

References:

EdGap.org. (2017). *Educational Opportunity Project Data Portal.*

National Center for Education Statistics (NCES). (2017). *Common Core of Data: Public Elementary/Secondary School Universe Survey.*

NCES. (2017). *School Neighborhood Poverty Estimates.*