Predicting School Performance from Socioeconomic Factors

1. 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.

Using a multiple regression model, 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, this variable was included because community-level economic hardship often affects school funding, access to learning resources, and overall student readiness. Using a multiple regression model, the results show that socioeconomic variables explain about 63% of the variation in ACT scores, with the percentage of students receiving free or reduced-price lunch emerging as the strongest predictor. To improve model accuracy, data on neighborhood poverty levels were added because community-level economic hardship can influence school funding, learning resources, and student readiness. Although the improvement in accuracy was small, this finding suggests that the broader economic environment also affects educational outcomes.

2. 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 conditions may lead to fewer academic opportunities, larger class sizes, and reduced teacher retention, which together can hinder student preparation for standardized assessments such as the ACT. These factors can directly affect how well students are prepared for standardized tests such as the ACT. Socioeconomic inequality has long been recognized as a key driver of academic outcomes. Schools located in lower-income areas often face challenges such as limited funding, larger class sizes, and higher teacher turnover, which collectively reduce students' readiness for standardized assessments like the ACT.

Building on prior research linking poverty and family structure to test performance, this study examines how measurable socioeconomic indicators, both within schools and in their surrounding neighborhoods, can predict average ACT scores. The data combine socioeconomic statistics from EdGap.org with school-level and neighborhood-level information from the National Center for Education Statistics (NCES). This approach allows us to assess how community-level poverty interacts with school-level variables, providing a more complete view of how economic conditions shape academic performance. The next section outlines the data and modeling methods used in this analysis.

The analysis uses data from EdGap.org, which provides socioeconomic statistics, and from the National Center for Education Statistics (NCES), which includes detailed information about public schools. To capture broader community conditions, the study also incorporates neighborhood poverty rates from the NCES School Neighborhood Poverty Estimates. This combination of school-level and community-level data allows a more comprehensive analysis of how economic inequality shapes ACT performance, building on prior research linking poverty and family structure to student outcomes.

3. Methodology:

3.1. Data Preparation:

The dataset was created by merging records from EdGap.org and the National Center for Education Statistics (NCES) using unique school identifiers. It 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. To reflect broader community conditions, neighborhood poverty ratios were added from the NCES School Neighborhood Poverty Estimates dataset.

Before analysis, the data were reviewed to identify invalid or missing values. After filtering out extreme outliers, remaining gaps were filled using the Iterative Imputer to preserve multivariate relationships among variables and minimize bias compared to simple mean substitution.

3.2. Model Development:

Three regression models were developed for comparison. The first model used a simple linear regression approach that tested each socioeconomic variable individually to measure its relationship with average ACT scores. The second model employed multiple regression to include all socioeconomic factors simultaneously, allowing assessment of their combined explanatory power. Finally, an extended model replaced median income with neighborhood poverty to determine whether community-level hardship provided stronger predictive value for ACT performance.

3.3. Analytical Rationale:

Linear regression was selected because of its interpretability and its ability to estimate the relative strength of each socioeconomic factor. Preliminary inspection (see Figure 1) revealed near-linear trends between ACT scores and socioeconomic indicators, supporting this modeling choice. However, since linear models assume additive relationships, they may not capture more complex interactions among variables.

The following section presents the results from these models and the key statistical relationships observed.

4. Result:

4.1. Correlation Analysis

Figure 1 shows the correlation matrix illustrating relationships among socioeconomic factors and average ACT scores. Darker blue cells indicate stronger negative correlations, while lighter yellow cells indicate positive relationships. The percentage of students receiving free or reduced-price lunch shows the strongest negative correlation with ACT scores (r = –0.78), confirming that higher economic disadvantage corresponds with lower academic performance.

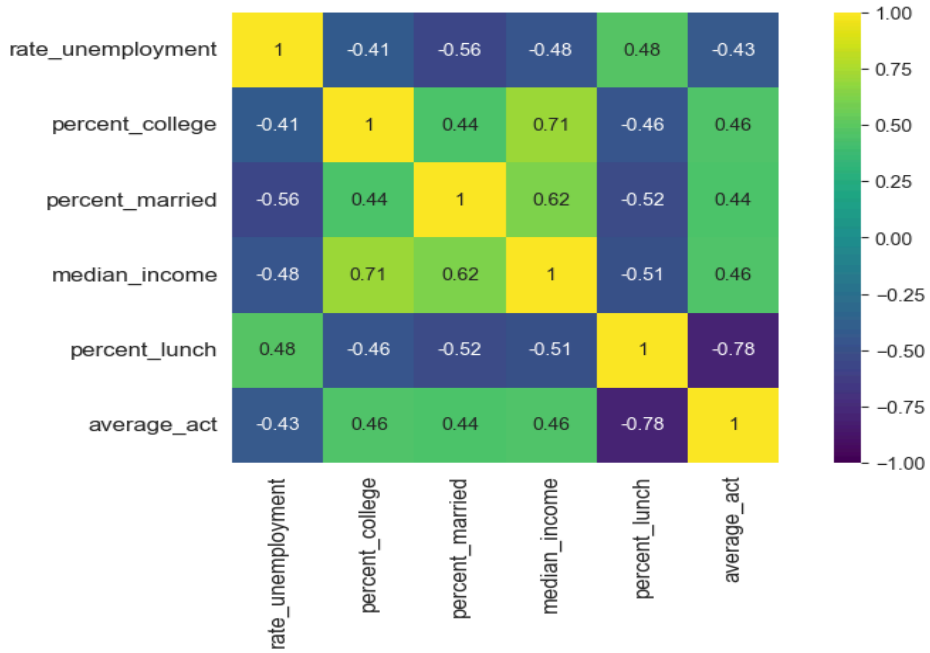

Figure 1. Correlation matrix of socioeconomic factors and average ACT scores.

This strong inverse relationship supports the hypothesis that economic hardship is a major driver of academic disparities, which the regression analysis quantifies below.

4.2. Multiple Regression Model

Regression analysis indicates that socioeconomic variables collectively explain 63% of the variation in average ACT scores ($R^2 = 0.63$). Schools with lower income levels, higher unemployment, and larger proportions of students receiving free or reduced-price lunch tend to report lower ACT performance. This suggests that socioeconomic factors are dominant—but not exclusive—predictors of academic outcomes.

4.3. Extended Model with Neighborhood Poverty

When neighborhood poverty replaced household income, the model's explanatory power increased slightly from 0.63 to 0.65 R^2, a 2% improvement. Although modest, this finding implies that community-level conditions capture aspects of educational inequality not visible through school-level data alone. Neighborhood context therefore adds incremental predictive value beyond the socioeconomic characteristics of the school population.

Conclusion:

While this study cannot establish causality, the strong associations observed imply that policies reducing economic hardship may indirectly enhance academic outcomes. These findings suggest that addressing socioeconomic inequality, through community investment, equitable funding, or targeted support, could improve academic performance. 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. This implies that addressing socioeconomic inequality through community investment, equitable resource allocation, and family support programs could positively influence student achievement.

Economic inequality may influence achievement through both resource disparities and psychosocial stressors, suggesting that interventions need to target not only schools but the broader community environment. Future studies could examine how school funding levels and teacher-student ratios interact with community poverty to shape outcomes over time, 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.*