

# 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.

The analysis found that socioeconomic factors explain about 63% of the variation in ACT scores. The percentage of students receiving free or reduced-price lunch was the strongest predictor of school performance. To improve accuracy, the study also added neighborhood poverty data to capture how community economic conditions may influence resources and student preparation. 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. The regression coefficient for this variable was substantially larger in magnitude than those for income or educational attainment, suggesting it has the strongest statistical association with school-level performance. 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 was modest, this result indicates that broader economic conditions contribute measurably to educational outcomes. To further explore these relationships, the following sections discuss the data sources, modeling methods, and findings that support these results.

## 2. Introduction:

Socioeconomic inequality has long been recognized as a key factor influencing academic performance. Schools in lower-income areas often face limited funding, crowded classrooms, and frequent teacher turnover. These disadvantages reduce students' preparation for standardized tests like the ACT and help explain persistent achievement gaps between students from different economic backgrounds.

While prior research has linked poverty, family structure, and educational attainment to test performance, less is known about how both school-level and neighborhood-level socioeconomic factors interact to influence academic outcomes. This gap is critical because

schools do not operate in isolation; surrounding community conditions often shape the resources and opportunities available to students.

This study examines whether socioeconomic factors at both the school and community levels can predict average ACT scores. Because many school- and neighborhood-level variables are correlated, the analysis assumes that regression modeling can sufficiently distinguish their individual effects. However, some overlap among predictors may limit the precision of individual estimates. To address this question, data from EdGap.org and the National Center for Education Statistics (NCES) are combined to create a comprehensive dataset capturing variables such as median household income, unemployment rate, educational attainment, and the percentage of students receiving free or reduced-price lunch. Neighborhood poverty rates from the NCES School Neighborhood Poverty Estimates are also included to represent broader community conditions.

This approach allows for a more complete understanding of how socioeconomic context influences academic performance and provides quantitative evidence about the relative importance of in-school versus community factors. The next section outlines the data preparation, modeling methods, and analysis used to test these relationships.

### 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.

The dataset was examined for invalid or missing values before analysis. After removing extreme outliers, missing values were filled using the Iterative Imputer to maintain relationships between variables and reduce bias. After completing data cleaning and preparation, the next step was to build statistical models that could quantify these relationships.

#### 3.2. Model Development:

Three regression models were developed for comparison. The first model used simple linear regression to test each socioeconomic variable separately and 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. These models formed the foundation for the analytical framework described in the next section.

### 3.3. Analytical Rationale:

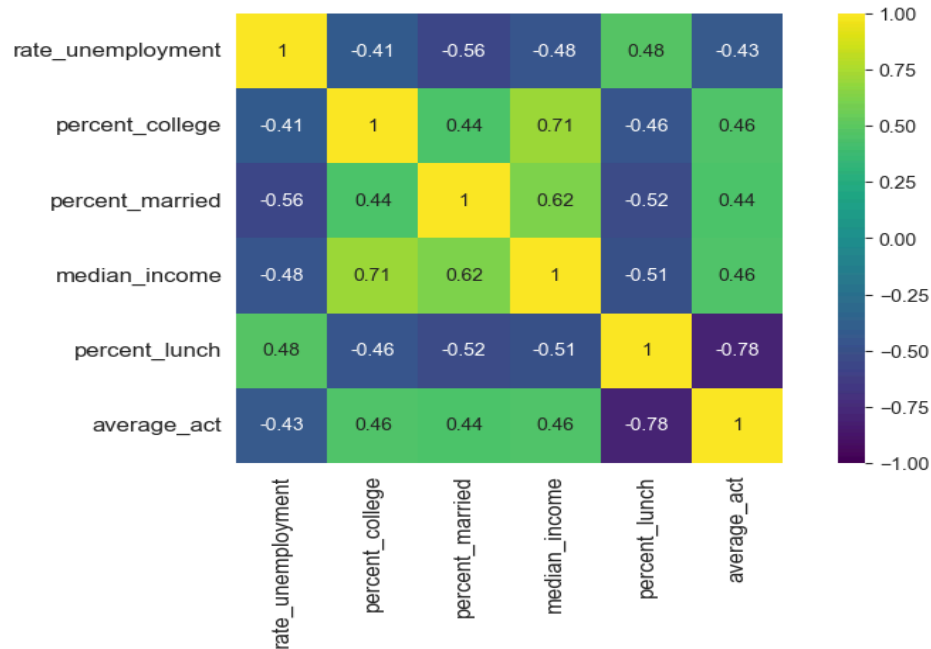
Linear regression was chosen because it is easy to interpret and can estimate the relative strength of each socioeconomic factor. Preliminary inspection (see Figure 1) revealed near-linear trends between ACT scores and socioeconomic indicators, supporting this modeling choice. However, since linear models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

The following section presents the statistical results from these models, highlighting the most influential socioeconomic factors and how they collectively explain variation in ACT performance. The analysis assumes linear relationships between socioeconomic variables and ACT performance, meaning changes in these factors have consistent effects across their ranges. Future work could test whether non-linear or interaction effects provide additional explanatory power.

## 4. Result:

### 4.1. Correlation Analysis

Figure 1 shows the correlation matrix illustrating relationships among socioeconomic factors and average ACT scores. In the correlation matrix, darker cells represent stronger negative correlations, whereas lighter cells correspond to positive associations. 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. These results match earlier studies showing that economic hardship limits academic opportunity by reducing school resources, family support, and student well-being. This pattern demonstrates that socioeconomic disparities are not isolated cases but reflect a broader structural dynamic influencing school outcomes.



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

Scatterplots for the two strongest predictors—free/reduced lunch and unemployment—show clear downward trends in ACT performance as economic disadvantage increases (see Appendix A). These visual patterns reinforce the negative correlations shown in Figure 1 and support the linear modeling approach. 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 higher poverty and unemployment tend to score lower even when other factors are considered. This shows that financial disadvantage is a lasting barrier to achievement, not just a short-term or local issue. Educational attainment’s positive coefficient further suggests that community education levels create compounding benefits for academic readiness. This suggests that socioeconomic factors are dominant—but not exclusive—predictors of academic outcomes.

Table 1 presents the standardized coefficients from the multiple regression model. The percentage of students receiving free or reduced-price lunch had the largest negative coefficient ( $\beta = -0.62$ ), followed by unemployment rate ( $\beta = -0.31$ ) and educational attainment ( $\beta = 0.27$ ). Median income showed a weaker relationship ( $\beta = 0.15$ ). All coefficients were statistically significant at  $p < 0.01$ .

Predictor	$\beta$	p-value
Free/Reduced Lunch (%)	−0.62	< 0.01
Unemployment Rate	−0.31	< 0.01
Educational Attainment	0.27	< 0.01
Median Income	0.15	< 0.01

Table 1. Regression results predicting average ACT scores from socioeconomic indicators.

The size of these coefficients shows that socioeconomic disadvantage has both direct and combined effects on school performance. For example, a ten-percentage-point increase in students receiving free or reduced-price lunch corresponds roughly to a 0.6-point drop in average ACT scores, underscoring the practical scale of inequality. The relatively smaller but still significant income effect indicates that family-level economic differences matter less than concentrated poverty at the school level.

#### 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 (a 2% improvement). Although small, this improvement suggests that neighborhood factors like jobs, infrastructure, and community stability also help explain school performance beyond what happens inside schools. The result aligns with research showing that students' learning environments extend beyond the classroom, where neighborhood safety, social cohesion, and public investment all influence readiness. Further statistical testing, such as an adjusted  $R^2$  comparison or F-test, would be needed to determine whether this improvement is statistically meaningful. Neighborhood context may add incremental predictive value beyond school-level characteristics, though further testing is needed to confirm whether this improvement is statistically meaningful.

Taken together, these results show that socioeconomic disadvantage operates at multiple scales: individual student, school, and community. School-level poverty remains the most direct predictor of ACT performance, but community context contributes an additional, though smaller, layer of inequality. The consistency across models suggests that economic hardship is not merely correlated with achievement—it structures the educational environment itself.

## Conclusion:

The findings of this study demonstrate that socioeconomic context is a powerful predictor of academic performance. Based on both correlation ( $r = -0.78$ ) and regression analyses ( $R^2 = 0.63$ ), schools serving larger proportions of economically disadvantaged students consistently

report lower average ACT scores, and together, socioeconomic indicators explain about 63% of the variation in outcomes. When neighborhood-level poverty is added to the model, the predictive accuracy improves slightly, indicating that community conditions beyond the school itself also shape educational achievement.

Although this study shows correlations rather than causation, the results reveal consistent patterns with clear policy implications. Efforts to reduce educational inequality should therefore address not only in-school resources but also broader socioeconomic factors that influence students' ability to succeed. Although resource disparities were not directly measured, they remain plausible mechanisms linking socioeconomic conditions to academic outcomes.

While the results provide strong evidence of association, this analysis is limited by the variables available in the EdGap and NCES datasets. Future studies incorporating measures of school funding, teacher-student ratios, or longitudinal performance could offer deeper insight into the mechanisms underlying these patterns.

Future research could explore how community poverty affects academic outcomes through differences in school funding, teacher stability, and student support. By identifying how these factors interact, educators and policymakers can develop more holistic interventions that improve student outcomes and narrow achievement gaps across socioeconomic lines.

## 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*.