

Predicting High School ACT Performance from Socioeconomic Factors

Abstract

This study investigates the impact of community-level socioeconomic factors on high school ACT scores. Using EdGap data, including unemployment, adult education, household income, family structure, and the percentage of students receiving free or reduced-price lunch, merged with NCES school information, we created a cleaned dataset of high schools. Exploratory analysis revealed that higher percentages of low-income students and elevated unemployment rates were associated with lower ACT scores, whereas higher household income, adult education, and family stability corresponded to higher scores. A simple linear regression using the percentage of students receiving free or reduced-price lunch explained 61% of the variation in ACT scores, with residual analysis confirming reasonable predictive accuracy. These findings highlight the substantial influence of socioeconomic disparities on educational outcomes and emphasize the importance of targeted support for low-income students.

Introduction

High school performance is shaped by both school-level characteristics and the socioeconomic context of students' communities. Understanding these relationships helps educators and policymakers design interventions to improve academic outcomes. This study examines average ACT scores and explores how community- and school-level socioeconomic factors relate to performance.

We use two primary datasets:

- **EdGap**, which provides community-level socioeconomic variables such as household income, unemployment, adult education, family structure, and the percentage of students receiving free or reduced-price lunch.
- **School Information**, which includes school identifiers, type, location, and charter status. Merging and cleaning these datasets for high schools allows for a focused analysis of how socioeconomic factors relate to ACT performance.

Theoretical Background

High school student performance, as measured by standardized tests like the ACT, is influenced by individual, school, and community factors. Community socioeconomic characteristics have consistently been shown to correlate with academic outcomes. Key factors include:

Median Household Income: Provides access to resources and extracurricular opportunities.

Adult Educational Attainment: Fosters a culture that emphasizes academic achievement.

Unemployment Rate: High rates may contribute to financial stress and lower student engagement.

Family Structure: Greater family stability supports learning.

Percent of Students Receiving Free or Reduced-Price Lunch: Serves as a proxy for low-income status and is typically negatively correlated with performance.

Methodology

The analysis followed a structured, reproducible approach to examine how socioeconomic factors relate to high school ACT scores. Two primary datasets were used:

EdGap data: School-level socioeconomic information.

NCES school information: School identifiers, type, and location.

Data cleaning included converting column types, filtering for high schools, standardizing column names, and handling missing or invalid values using iterative regression-based imputation. Datasets were merged on school ID, producing a final dataset saved as `education_clean.csv`.

Exploratory data analysis included generating a *correlation heatmap* to examine relationships between ACT scores and socioeconomic predictors. A **simple linear regression model** was then built using the percentage of students receiving free or reduced-price lunch as the predictor. Model performance was assessed with R-squared, mean absolute error, and residual analysis.

Computational Results

1. Exploratory Data Analysis (EDA)

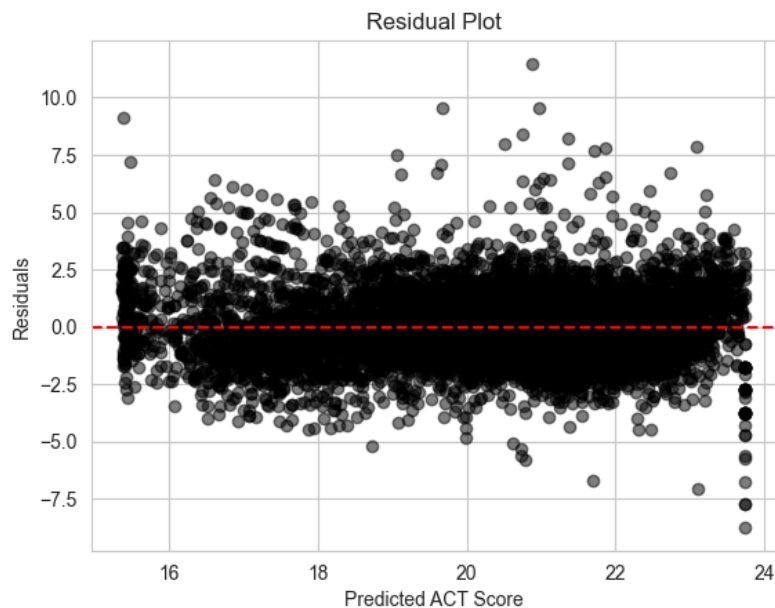
A correlation heatmap (*figure 1*) was generated to evaluate relationships between average ACT scores and socioeconomic predictors. Key findings include:

- Percent of students receiving free or reduced-price lunch (`percent_lunch`) exhibited the strongest negative correlation with average ACT scores ($r \approx -0.78$).
- Median household income, percent of adults with college degrees, and percent of married adults showed moderate positive correlations with ACT scores ($r \approx 0.44$ – 0.46).
- Unemployment rate had a modest negative correlation with ACT scores ($r \approx -0.43$).

These results suggest that schools in communities with higher economic resources and educational attainment tend to achieve higher ACT scores, whereas higher rates of low-income students and unemployment are associated with lower ACT performance.

Figure 1: Correlation heatmap of average ACT scores and socioeconomic factors.

Figure 3: Residual plot for the simple linear regression model.



The linear regression model confirmed that percent_lunch alone explained approximately 61% of the variation in ACT scores ($R^2 = 0.614$), with a mean absolute error of 1.61. Residual plots (figure 3) indicated that model assumptions were reasonably satisfied.

Discussion

The analysis demonstrates a clear link between socioeconomic disadvantage and lower ACT performance. Percent_lunch was the strongest predictor, while household income, adult education, and family stability also showed meaningful associations. These findings are consistent with prior research showing that academic achievement depends on a combination of community, family, and school-level factors.

While percent_lunch is a dominant predictor, other unobserved factors likely contribute to performance variation. These insights suggest that targeted interventions, such as tutoring, nutrition programs, or community engagement for low-income students, could reduce performance gaps. Policymakers and educators can use these findings to allocate resources where they are most needed.

Conclusion

This study demonstrates that community-level socioeconomic factors, particularly low-income status, significantly influence high school ACT performance. Percent of students receiving free or

reduced-price lunch was the strongest predictor, explaining approximately 61% of score variation, while household income, adult education, and family stability also contributed. Targeted interventions for disadvantaged students could meaningfully improve ACT outcomes, providing actionable guidance for educators and policymakers seeking to address educational disparities.

References

EdGap.org (2016) *EdGap socioeconomic and ACT/SAT data by U.S. high school* [data set]. Available at: https://github.com/brian-fischer/DATA-5100/blob/main/EdGap_data.xlsx (Accessed: 21 October 2025).

Data originally sourced from the U.S. Census Bureau's *American Community Survey (2016)* and various state departments of education. Available at: <https://www.edgap.org>

National Center for Education Statistics (2017) *Common Core of Data (CCD): Public school data file (ccd_sch_029_1617_w_1a_11212017.csv)* [data set]. U.S. Department of Education. Available at: https://www.dropbox.com/s/lkl5nvcdmwyoban/ccd_sch_029_1617_w_1a_11212017.csv?dl=1 (Accessed: 21 October 2025).

Original source: <https://nces.ed.gov/ccd/pubschuniv.asp>