# Educational Inequality and ACT Performance

## 1. Introduction

Educational inequality remains a persistent challenge in the United States, with performance gaps evident among students from different economic and social backgrounds. Standardized college entrance exams such as the ACT reflect these disparities, making them a useful measure of how community and household conditions shape academic achievement.

This study investigates **how socio-economic factors affect average ACT scores across U.S. high schools**. Specifically, it identifies which factors—such as household income, parental education, unemployment, or economic disadvantage—most strongly predict student performance. Using multiple public datasets, school-level socio-economic indicators were merged with institutional characteristics and analyzed through regression models. The goal is to quantify these influences and highlight patterns that can inform educators and policymakers about the roots of educational inequality.

## 2. Data and Methodology

### 2.1 Data Sources

Three datasets were used: EdGap (socio-economic indicators), School Information, and Public School Data. These were merged using the common NCESSCH identifier to form a comprehensive dataset of schools, demographics, and outcomes.

### 2.2 Data Cleaning and Preparation

Preliminary checks were conducted to assess data structure and quality. Out-of-range values in ACT score, percent lunch, and student–teacher ratio were treated as missing (NaN) to prevent skewed results. Missing values were imputed using **IterativeImputer**, chosen because it preserves relationships among variables.

### 2.3 Data Quality Checks

Outliers were flagged in ACT scores and socio-economic variables. These were retained but noted, as they represent meaningful extremes rather than errors. A map visualization was also created to show the geographic spread of included schools.

## 3. Exploratory Data Analysis (EDA)

The exploratory analysis revealed several important patterns:

- **Percent lunch** (students receiving free/reduced-price lunch, a proxy for poverty) showed a strong negative correlation with ACT scores (r ≈ -0.72).
- **Median income** and **percent of adults with a college degree** were positively correlated with ACT scores (r ≈ 0.55 and r ≈ 0.60, respectively).
- **Unemployment rate** was negatively correlated with ACT performance (r ≈ -0.45).
- **Student–teacher ratio** displayed little to no relationship with outcomes.
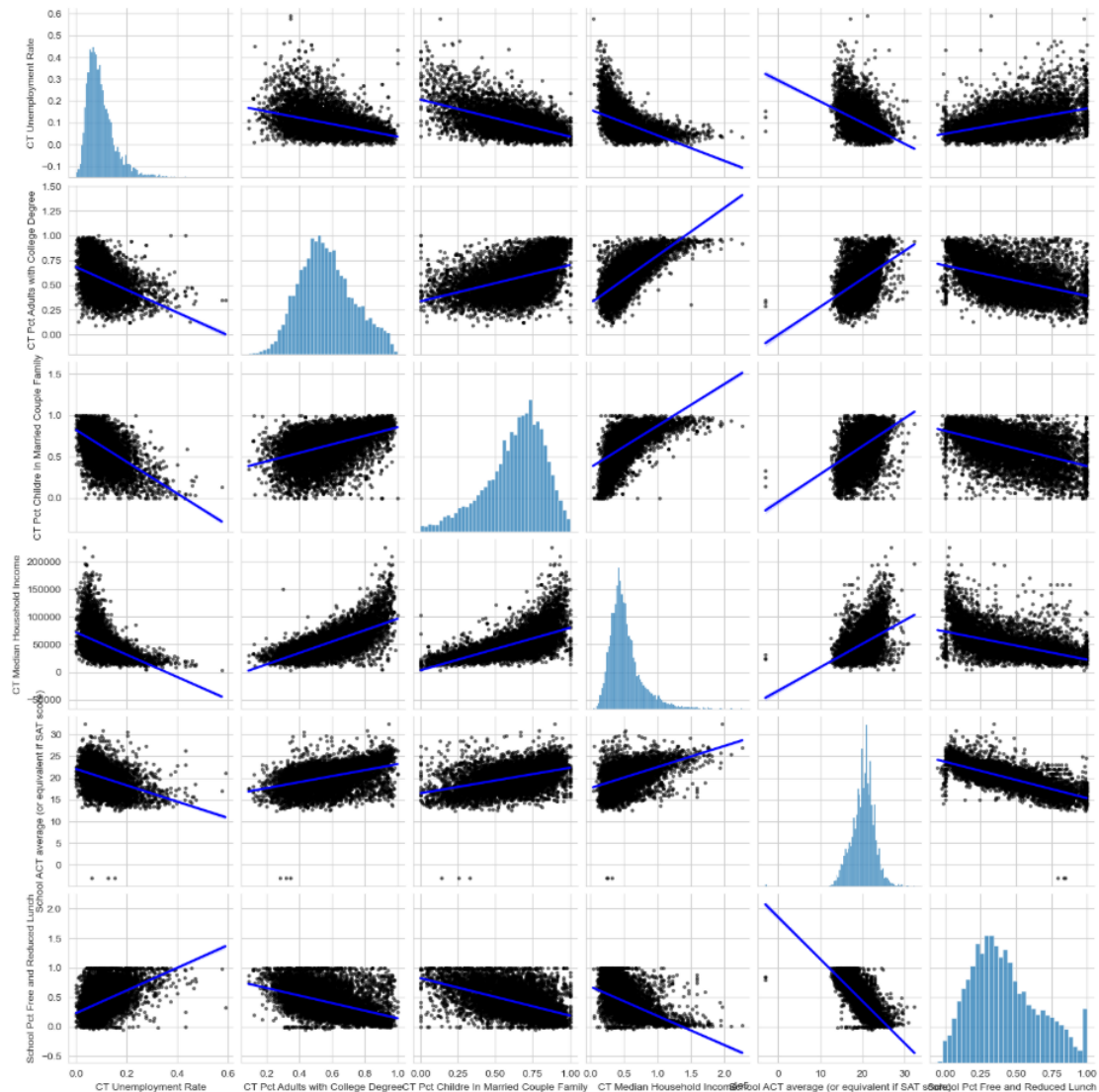
**Figures:** -



*Figure 1*. Pair plots of socio-economic variables and ACT scores.

The pair plots reveal clear negative associations between economic disadvantage (percent lunch, unemployment) and ACT scores, alongside positive relationships with income and education. These visual patterns highlight which factors warrant deeper regression analysis.
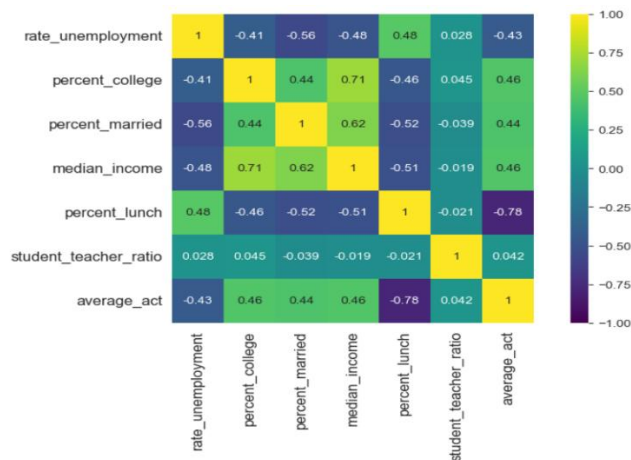
*Figure 2.* Correlation heatmap of socio-economic factors and ACT scores.

The heatmap quantifies correlations, showing percent lunch has the strongest negative link (r ≈ -0.78) with ACT scores. It also highlights multicollinearity among income and education variables, important for model interpretation.

# 4. Modeling and Results

## 4.1 Simple Linear Regression

A model using **median household income** as the sole predictor explained 21% of ACT variation ($R^2$ = 0.211, p < 0.001). While significant, this indicates income alone is insufficient to explain school-level performance differences.
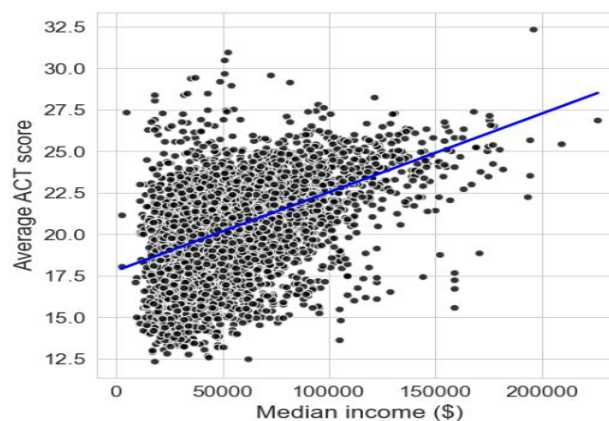


*Figure 3.* Simple linear regression of median income vs. ACT scores.

This regression demonstrates a statistically significant positive relationship between household income and ACT outcomes. However, the modest $R^2$ = 0.21 indicates income alone explains limited variation.

## 4.2 Multiple Linear Regression

Adding socio-economic variables improved explanatory power ($R^2$ = 0.628, MAE = 1.145). Key findings: - **Percent lunch, percent college, and unemployment rate** were the strongest predictors. - **Median income and percent married** were statistically insignificant once other variables were included, likely due to overlap with percent lunch.

Residuals were randomly distributed around zero, confirming the model met regression assumptions.
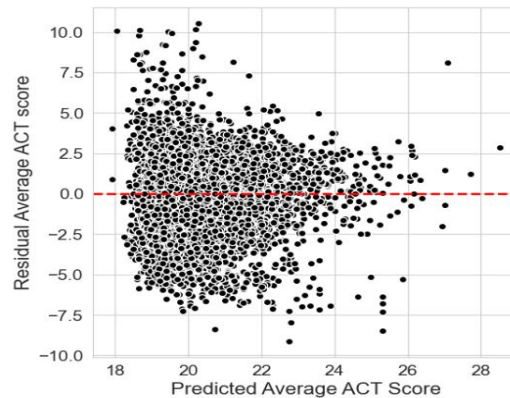


*Figure 4*. Residuals of multiple regression model.

Residuals are evenly scattered around zero, confirming that linear regression assumptions hold. This strengthens confidence that the model provides a valid estimate of socio-economic influences on ACT performance.

## 4.3 Reduced Model

A simplified model using only three predictors—percent lunch, percent college, and unemployment—retained similar accuracy ($R^2$ = 0.628). This suggests that a small set of socio-economic factors explains most ACT variation. Simplification also increases interpretability for policy applications.

## 4.4 Scaling and Standardized Effects

When standardized, **percent lunch** had the largest effect (-1.77), nearly double that of unemployment (-0.92) and percent college (+0.88). This confirms poverty is the most influential determinant of ACT outcomes.

|  | Mean Absolute Error | R Squared |
|---|---|---|
| normalized model | 1.1455 | 0.628 |
| reduced model | 1.1455 | 0.628 |

Normalization is only for interpretability (e.g., comparing effect sizes).

Model accuracy is unchanged: MAE and R² are identical.

Predictor importance is easier to visualize in standardized form.

*Figure 5*. Standardized model comparison and interpretation

The comparison shows that normalization does not change accuracy ($R^2$ and MAE remain constant), but it improves interpretability. Standardized coefficients highlight percent lunch as the most influential predictor, nearly twice as strong as other factors.

## 5. Discussion

The analysis demonstrates that **poverty-related measures are the strongest predictors of academic outcomes**. Percent lunch, in particular, had nearly twice the standardized effect of other variables, underscoring the structural barrier posed by economic disadvantage. Communities with higher education levels also foster stronger ACT results, reflecting how parental or community education supports achievement.

The multiple regression explained over 60% of ACT variation—a substantial figure for social science data. This indicates that inequality in performance can largely be traced to a few key socio-economic conditions. However, about 40% of variation remains unexplained, highlighting the importance of unmeasured factors such as school quality, teacher experience, or student motivation.

From a policy perspective, these findings emphasize that addressing poverty and improving community education levels are essential for narrowing performance gaps. Targeted interventions—such as increased school funding, access to college prep resources, and community-based education programs—could help reduce disparities. Still, as this study is correlational, results should be interpreted as **associations rather than causal effects**.

## 6. Conclusion

This project examined how socio-economic factors shape average ACT scores across U.S. high schools. The analysis found that schools in communities with higher household income, more educated adults, and fewer economically disadvantaged students consistently had higher ACT performance. Among predictors, **percent lunch emerged as the single most powerful factor**, highlighting the central role of poverty in shaping educational outcomes.

The models showed that a small set of socio-economic indicators—percent lunch, percent college, and unemployment—account for over 60% of ACT score variation. While socio-economic conditions strongly predict outcomes, other influences remain, and future research should incorporate additional measures such as school quality and student engagement.

Overall, the evidence provides **clear quantitative support that educational inequality is closely tied to socio-economic context**. Interventions that reduce poverty and expand educational opportunity at the community level are likely to have the greatest impact on narrowing achievement gaps.