

# **Inequality in Education: Analyzing how socioeconomic conditions and Student–Teacher ratios affect student achievement**

## **Abstract**

This study investigates educational inequality across the United States high schools by examining how socioeconomic factors and school level resources like student teacher ratio affects student performance on college entrance exams like ACT and SAT. This analysis combines Edgap data with National Center for Education Statistics (NCES) dataset like school information, staff information and student membership information. Using correlation analysis, simple linear regression and multiple linear regression, results show that students eligible for free or reduced-price lunch is the strongest predictor of ACT performance. Although the student–teacher ratio was statistically significant, it explained weak relationship in ACT scores. The findings suggest that economic disadvantages like students eligible for free or reduced-price lunch has a much stronger relationship with student achievement.

## **Introduction**

Unequal access to education in the United States leads to unequal outcomes for students. Disparities in educational opportunity arise from multiple factors, including government policies, school funding and choice, family wealth, parental education, implicit bias related to race or ethnicity, and unequal access to school resources. These inequities not only influence academic performance but also contribute to broader societal challenges such as income inequality and higher incarceration rates.

This project explores inequality in educational opportunities across U.S. high schools by examining how socioeconomic factors influence student achievement. Specifically, it analyzes the relationship between median household income, unemployment rate, percentage of students eligible for free or reduced-price lunch, proportion of married parents, and parental education levels and average student performance on standardized college entrance exams like the ACT and SAT. The goal is to identify patterns and disparities that reveal how varying socioeconomic conditions contribute to differences in academic outcomes, offering insights into the broader issue of educational inequality in the United States.

In addition, the project investigates whether the student-to-teacher ratio impacts average ACT scores. It evaluates if schools with lower student-to-teacher ratios which typically indicate smaller class sizes and greater individual attention are associated with higher student performance, thereby extending the understanding of how school-level resources affect educational outcomes.

## **Data Description**

### **Data Source**

-[National Center for Education Statistics](#)

-[Edgap.org](#)

Four data sets were combined for analysis:

1. [EdGap dataset](#) (2016): includes information about average ACT or SAT scores for schools and several socioeconomic characteristics of the school district. All socioeconomic data (household income, unemployment, adult educational attainment, and family structure) are from the Census Bureau's American Community Survey.
2. [NCES School Information](#) (2016–2017): is basic information about each school from the NCES.
3. [NCES Staff Count](#) (2016–2017): includes basic information of school and reports the number of teachers per school.
4. [NCES Student Count](#) (2016–2017): includes basic information of school and reports the number of students per grade level.

## Data Cleaning and Preparation

- The data was processed to subset the datasets, remove unnecessary columns, and rename the remaining columns for clarity.
- Only high schools (grades 9–12) were retained from NCES Student count.
- NCES Staff count and NCES Student count were combined to calculate the student teacher ratio.
- Schools with extreme student–teacher ratios ( $> 100$ ) were excluded.
- EdGap, school information and student-teacher dataset were merged into one dataframe.
- Out-of-range ACT scores ( $< 1$ ) and lunch percentages ( $< 0$ ) were replaced with missing values.
- Missing numerical values were imputed using an iterative imputer.
- This cleaned and consolidated dataset was used for the Exploratory Data analysis and Modeling.

## Theoretical Background

In the United States, a family's socioeconomic status (SES) has a significant impact on the child's education. The parents' level of education, income, and career attainment combine to determine the level of difficulty their children will face in school. This environment creates an inequality of learning between children from high-SES families and children from low-SES families. High-SES families can ensure their children receive a beneficial education while low-SES families are not usually able to ensure the same educational quality for their children. This results in children of less wealthy families performing less well in schools than children of wealthier families. There are several factors that contribute to this disparity; these factors narrow into two main subjects: resources and environment.

The student–teacher ratio reflects the level of instructional attention students receive. While smaller class sizes are often associated with improved learning, evidence on their impact to socioeconomic context is mixed. This study situates both factors within a single analytical framework to evaluate their relative contributions to student achievement.

# Methodology

## Exploratory Data analysis and Modeling

- Correlation analysis assessed linear relationships between predictors and ACT scores.
- Explored the relationship between the variables using correlation matrix of the numerical variables.
- Simple linear regressions were fit for each socioeconomic variable and student-teacher ratio to evaluate independent effects.
- Multiple linear regression combined predictors to estimate overall contribution.
- Reduced model was fit to include those predictor variables that has statistically significant coefficients.
- Scaled the predictor variables in the reduced model to have a mean of 0 and standard deviation of 1.
- Model comparison using ANOVA determined whether reduced models with only significant predictors performed better or worse the full model.
- Model performance was evaluated with  $R^2$ , mean absolute error (MAE), and p-values for significance.

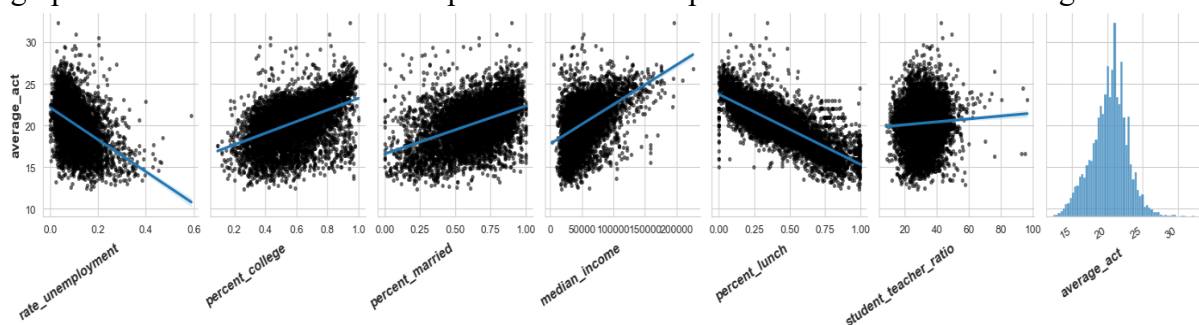
## Computational Results

### Correlation analysis:

ACT scores showed strong relationships with several socioeconomic factors.

- Median income ( $r = 0.46$ ), percentage of adults with college degree ( $r = 0.46$ ), and percentage of children in a married couple family ( $r = 0.44$ ) were positively correlated with ACT scores.
- Unemployment rate ( $r = -0.43$ ) and percent of students receiving free lunch ( $r = -0.78$ ) were negatively correlated.
- Student-teacher ratio showed a very weak positive correlation ( $r = 0.057$ ).

The graph below shows the relationship between various predictor variables and average ACT.



### Simple Linear Regression (one predictor)

- Median income: significant predictor ( $R^2 = 0.21$ ).
- Unemployment rate: negative relationship ( $R^2 = 0.19$ ).
- percentage of adults with college degree: significant predictor ( $R^2 = 0.21$ )
- percentage of children in a married couple family ( $R^2 = 0.19$ ).

- percent of students receiving free lunch: significant predictor ( $R^2 = 0.61$ ).
- Student-teacher ratio: significant but weak predictor ( $R^2 = 0.003$ ).

The summary table below shows the regression results for percentage of children receiving free/reduced lunch price as single input predictor.

OLS Regression Results						
Dep. Variable:	average_act	R-squared:	0.614			
Model:	OLS	Adj. R-squared:	0.614			
Method:	Least Squares	F-statistic:	1.149e+04			
Date:	Wed, 22 Oct 2025	Prob (F-statistic):	0.00			
Time:	10:21:05	Log-Likelihood:	-13461.			
No. Observations:	7227	AIC:	2.693e+04			
Df Residuals:	7225	BIC:	2.694e+04			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	23.7429	0.037	641.745	0.000	23.670	23.815
percent_lunch	-8.3902	0.078	-107.186	0.000	-8.544	-8.237
Omnibus:	842.388	Durbin-Watson:	1.472			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2845.328			
Skew:	0.582	Prob(JB):	0.00			
Kurtosis:	5.845	Cond. No.	5.02			

## Multiple Linear Regression:

The full model, including all socioeconomic variables and student-teacher ratio, resulted in  $R^2=0.63$ . However, median income and percentage of children in a married couple family were not statistically significant since there were correlations among these two predictors.

A reduced model including unemployment rate, percent college, percent free lunch, and student-teacher ratio performed equivalently i.e.  $R^2$  of 0.63, indicating that these four predictors capture most of the variation in ACT performance. Scaling the predictors showed that percent free lunch had the largest standardized coefficient, confirming it as the dominant factor. The student-teacher ratio remained statistically significant but contributed minimally with  $R^2$  of 0.003.

The summary table below shows the regression results for reduced normalized model.

OLS Regression Results						
Dep. Variable:	average_act	R-squared:	0.629			
Model:	OLS	Adj. R-squared:	0.629			
Method:	Least Squares	F-statistic:	3066.			
Date:	Wed, 22 Oct 2025	Prob (F-statistic):	0.00			
Time:	10:21:06	Log-Likelihood:	-13314.			
No. Observations:	7227	AIC:	2.664e+04			
Df Residuals:	7222	BIC:	2.667e+04			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	20.2986	0.018	1129.703	0.000	20.263	20.334
rate_unemployment_normalized	-0.1364	0.021	-6.414	0.000	-0.178	-0.095
percent_college_normalized	0.2621	0.021	12.337	0.000	0.220	0.304
percent_lunch_normalized	-1.7790	0.022	-82.225	0.000	-1.821	-1.737
student_teacher_ratio_normalized	0.0973	0.018	5.306	0.000	0.061	0.133
Omnibus:	863.435	Durbin-Watson:	1.481			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3016.351			
Skew:	0.587	Prob(JB):	0.00			
Kurtosis:	5.939	Cond. No.	1.93			

## Discussion

The analysis showcased that the socioeconomic variables, particularly the percentage of students receiving free lunch are the strongest predictors of the regression model. This relationship highlights how a family's socioeconomic status (SES) has a significant impact on the child's education. Median income and college attainment show positive associations but do not add substantial predictive power once poverty levels are included in the model. Although the student–teacher ratio was statistically significant, its explanatory value was minimal. The weak correlation suggests that while smaller classes might support learning, they do not offset broader socioeconomic challenges. Among the models that were tested, single input model with percentage of children receiving free lunch had  $R^2$  of 0.614, the multiple linear model with all the 6 predictors had  $R^2$  of 0.629, reduced model with only 4 of the significant predictors has a  $R^2$  0.629. This means that the simple model performs nearly as well as the multiple linear model and the reduced model.

However, the analysis is limited to only 20 states. The Edgap data set has approximately 7900 schools only. Including all 50 states or including all schools and additional school-resource variables like per student spending, access to internet access could provide a more detailed view.

## Conclusion

This study looked at how family income, education level, unemployment, and student–teacher ratio affects student achievement in U.S. high schools. The results show that economic factors, especially the percentage of students getting free lunch, have the biggest impact on ACT scores. Schools with higher household income, more college-educated parents, and lower unemployment rates usually have higher ACT scores.

Although the student–teacher ratio was statistically important, it explained only a small part of the difference in scores. This means that community and family conditions affect student success more than classroom size does.

## References

- National Center for Education Statistics: *Public Elementary/Secondary School Universe Survey Data (2016–2017)*. <https://nces.ed.gov/ccd/pubschuniv.asp>
- *Education gap Data Portal*: <https://www.edgap.org>
- Wikipedia: [Educational inequality in the United States](#)