EDUCATION PROJECT

Abstract

This Project investigates the predictive performance of socioeconomic status and school expenditures on academic performance across schools, as measured by mean ACT scores. Utilizing data from EdGap.org and school spending records, correlations between median family income, parent educational attainment, employment status, and salary-based school spending from federal and non-federal funds are examined. The Data is ready, merged, and examined with correlation and regression analysis techniques to assess the predictive power. Outcomes indicate that socioeconomic variables, namely income and education of parents are the strongest predictors of school achievement, with salary outlays indicating relatively weaker effects. The findings present evidence that investment in economics is not sufficient to close socioeconomic-based achievement gaps. The findings are evidence-based recommendations intended to guide education policy and resource allocation interventions to improve student performance.

Introduction

Academic achievement is the most reliable metric of student and education quality. However, achievement levels are extremely divergent among schools, and concerns are raised about the grounds on which these differences exist. Though policymakers prefer to cite school finance as the reform handle, the prevailing evidence identifies the intersecting roles played by school resources and socioeconomic context in determining outcomes.

This study explores whether socioeconomic status or school spending are more accurate predictors of school performance, as defined by average ACT score. The analysis uses three primary data sources socioeconomic data obtained from EdGap.org, fiscal data at the school level with federal and non-federal funding expenditures, and student performance metrics. This research is important in the sense that it provides insight into whether or not increasing school expenditure more precisely employee salary investment can minimize the performance gap caused by socioeconomic disparity. The empirical goal of this study is therefore to evaluate the relative explanatory power of measures of socioeconomic factors and school spending to forecast educational performance, and to examine whether sources of school finance (federal vs. non-federal) influence these findings differently. Answering these questions gives the analysis a fact-based discovery of how combined determinants of economic and institutional sources affect educational achievement.

Theoretical Background

Academic achievement has been a multidimensional outcome that is affected by institutional and individual factors. Socioeconomic status is one of the most established predictors of academic achievement, representing the combined influence of family income, parental education, and occupational stability. The empirical literature consistently demonstrates that students from higher socioeconomic status achieve more on standard measures as a result of increased access to educational resources, stable home environments, and conducive learning climates.

At the institutional level, school funding and expenditure patterns are believed to influence educational outcomes through mechanisms such as teacher quality, classroom resources, and student support services. The literature indicates, however, that the impact of funding on performance is more affected by how financial resources are used than by the total amount spent. Allocations that focus on hiring high-quality staff and providing instructional supports tend to have a greater effect on learning results than general increases in funding. Consequently, the theory shifts to a comparative analysis of socioeconomic factors versus fiscal policy in shaping school-level academic success.

Methodology

1. Data sources and description:

The analysis integrates three primary datasets:

1. Socioeconomic data from EdGap.org, including indicators such as median household income, parental education level and employment.
2. School expenditure data provide details on employee salary expenditure from both federal and non-federal funding sources.
3. Student achievement data, represented by each school's average ACT score.

The datasets were merged using consistent school identifiers and geographic information. After merging, redundant columns were removed, missing values were treated using mean imputation, and numerical variables were normalized to ensure comparability across scales.

2. Data preparation and cleaning

Data preprocessing was done in Python using libraries such as Pandas, NumPy and Matplotlib.
The main cleaning steps include:

1. Addressing missing socio-economic characteristics through stigma.
2. Rename variables for consistency across datasets.
3. Remove outliers that deviate more than three standard deviations from the mean.
4. Verification of data alignment after merging to maintain integrity across observations.

3. Analytical approach

The study used a combination of exploratory data analysis , correlational analysis, and linear regression modeling to examine how socioeconomic factors and school wage expenditures predict academic achievement. EDA involved visualizing variable distributions and relationships through scatter plots and heat maps to identify potential predictors. Correlation analysis determined the strength of the relationship between predictors (eg, income, education, financing type) and ACT scores. Multiple linear regression models were used to evaluate the combined effects of socioeconomic and expenditure variables on performance outcomes.

4. Evaluation Metrics

Model performance was primarily evaluated based on coefficient interpretability and prediction accuracy. The goal was not only to identify statistically significant predictors, but also to assess the relative importance of socioeconomic versus economic factors in explaining variation in school performance.

Computational Results

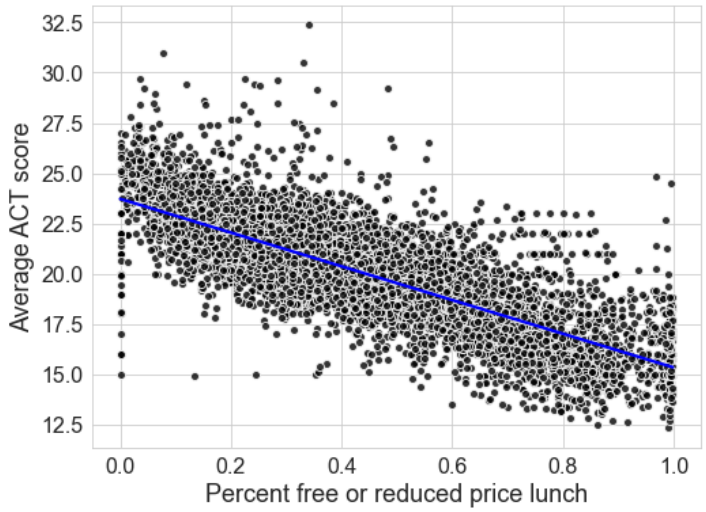

Figure 1.1: Relationship between socioeconomic status (represented by percent of students receiving free/reduced-price lunch) and academic achievement (average ACT score).

Scatter plot shows the relationship between the percentage of students receiving free or reduced price lunch and the average ACT score at various educational institutions. Each point represents data from one district, while the blue line shows the fitted regression trend. The visual trend shows a negative linear relationship, indicating that as the percentage of students who qualify for free or reduced-price lunch increases, the average ACT score decreases. The plot is densely populated in the bottom and middle ranges, suggesting that districts with moderate to high proportions of economically disadvantaged students generally show lower ACT performance levels.

OLS Regression Results

Dep. Variable:	average_act	R-squared:	0.626
Model:	OLS	Adj. R-squared:	0.626
Method:	Least Squares	F-statistic:	1908.
Date:	Wed, 22 Oct 2025	Prob (F-statistic):	0.00
Time:	21:51:35	Log-Likelihood:	-14900.
No. Observations:	7986	AIC:	2.982e+04
Df Residuals:	7978	BIC:	2.987e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	22.7160	0.132	172.266	0.000	22.458	22.975
rate_unemployment	-2.2512	0.380	-5.921	0.000	-2.996	-1.506
percent_college	1.6230	0.152	10.693	0.000	1.325	1.921
percent_married	-0.0167	0.127	-0.131	0.896	-0.266	0.233
median_income	-6.078e-07	1.2e-06	-0.506	0.613	-2.96e-06	1.75e-06
percent_lunch	-7.6376	0.092	-82.847	0.000	-7.818	-7.457
Tot_salary_exp	2.976e-09	2.92e-09	1.018	0.309	-2.76e-09	8.71e-09
Tot_salary_exp_federal	8.127e-09	3.83e-09	2.125	0.034	6.29e-10	1.56e-08

Omnibus:	935.156	Durbin-Watson:	1.520
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3083.746
Skew:	0.593	Prob(JB):	0.00
Kurtosis:	5.804	Cond. No.	1.73e+08

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.73e+08. This might indicate that there are
strong multicollinearity or other numerical problems.

Figure 1.2: Multiple Regression Model Coefficients

The ordinary least squares regression model with seven predictors explains 62.6percent of variance in average ACT scores across 7,986 schools. Poverty concentration percent free/reduced lunch is the dominant predictor with a coefficient of less p value indicating that each 1 percent increase in poverty concentration corresponds to a 7.676 point decrease in average ACT scores. Adult college education percentage shows a positive coefficient of, while unemployment rate displays a negative coefficient of. School salary expenditure variables both total and federal-funded yield weak, non-significant coefficients, confirming minimal predictive power. The model produces predictions with mean absolute error of 1.17 ACT points across the predicted range of sixteen to twenty-four points. Diagnostic assessment shows adequate satisfaction of regression assumptions with residuals randomly distributed around zero and approximately constant variance across predicted values.

Figure 1.2: Relationship between Total expenditures on salary with federal funds and academic achievement (average ACT score).

The scatterplot above presents the relationship between total expenditures for salaries with federal funds and average ACT scores at various educational institutions. Each dot represents a state or district, and the blue line shows the fitted regression trend. From the visualization, it is clear that there is a positive correlation between the two variables indicating that greater spending on teacher salaries (supported by federal funds) is associated with higher average ACT performance. However, the data points show significant dispersion near low spending levels, suggesting variation in performance outcomes despite similar spending levels. This figure effectively summarizes the underlying computational analysis, showing that investments in academic staff salaries can contribute to improved academic performance, although other relevant factors may influence the overall variation observed in ACT scores.

Discussion

The results of this analysis show a clean and regular relationship between socioeconomic elements and academic performance, confirming the theoretical expectation that family historical past and community context exert great have an effect on on instructional results. The regression consequences indicated that median family income and parental education stage had been the most extensive predictors of common ACT rankings. These findings align with previous literature emphasizing that social and monetary assets shape each get admission to to getting to know opportunities and scholar readiness for standardized testing. While faculty spending on employee salaries become positively associated with pupil performance, the energy of this courting become modest whilst socioeconomic variables had been included inside the version. This suggests that simply increasing profits costs does no longer robotically translate to improved instructional effects if underlying socioeconomic disparities persist. The analysis additionally as compared federal and non-federal investment assets, revealing that non-federal income fees confirmed a slightly stronger correlation with overall performance, probably because of greater flexibility in how neighborhood budget are allotted. However, these effects have been secondary to the dominant function of socioeconomic elements.
Importantly, the findings did now not display important proof of overfitting or records distortion, even though version in the regression residuals indicates the presence of unmeasured elements such as teacher best or faculty leadership. These ability sources of mistakes highlight that educational effects are motivated by way of complex interactions now not absolutely captured by means of economic or demographic indicators by myself.

Conclusions

This study was set out to determine whether socioeconomic status or school spending are more reliable indicators of attainment between schools. Combined socioeconomic, budgetary, and student performance data analysis revealed socioeconomic indicators i.e., parental income and education levels continue to be the strongest determinants of student attainment. While salary spending variances and funding origin did find some association with achievement, these budgetary indicators could not overcome disparities that were assigned to broader socioeconomic conditions.

The findings have policy and planning implications for education. They suggest that budget reforms alone cannot close achievement gaps unless they are accompanied by initiatives to directly confront inequality in family and community resources. Programs of targeted social support, early intervention, and equal access to school resources might be more lasting in their contribution to student gains than growth in spending alone. Briefly, the results provide concise evidence that scholastic success depends less on the level of school budgets than on the environment in which youngsters grow and learn. Addressing these underlying socioeconomic problems is the solution to fostering long-term fairness and efficacy in the education system.

References

https://www.act.org/content/dam/act/unsecured/documents/R2405-Equity-in-Education-2024-05.pdf

https://www.proquest.com/openview/bff78e66da4b3523be3393b7602550a9/1?pq-origsite=gscholar&cbl=18750&diss=y