Impact of Student Teacher Ratios and Socioeconomic Factors on Average ACT Scores

Abstract:

This study investigates whether the student–teacher ratio has as strong an influence on average ACT performance as socioeconomic factors. The analysis examined relationships among socioeconomic factors, school staffing, and academic outcomes. Correlation and regression analyses revealed that socioeconomic variables were far stronger predictors of average ACT scores within the data.

Introduction:

A student's performance on standardized tests like the ACT and SAT can have a significant influence on their future educational opportunities, such as college admissions or scholarships. Two factors often talked about in relation to educational quality are socioeconomic factors and student teacher ratios. This project will explore whether the student teacher ratio is as important to the quality of education (measured by average ACT score) as socioeconomic factors.
To answer this question and identify the most significant predictors, five datasets were used: Edgap.org's school level data, and four datasets from the National Center for Educational Statistics (NCES) Common Core of Data (CCD) which provided detailed school information, student body composition, and staffing counts.

Theoretical Background:

A student's performance on standardized tests can be influenced by a wide range of factors both inside and outside the classroom. Socioeconomic factors and school level resources are two topics that are often talked about as having an influence on educational quality. From a socioeconomic perspective, students from higher income communities often have access to more educational support and resources.

Within school, the student teacher ratio is frequently used as a measure of resource allocation and instructional quality. Smaller class sizes are associated with more attention on individuals, which is thought to improve academic performance. This

project aims to compare the relative impact of student teacher ratio and socioeconomic factors on average ACT scores.

Methodology:

Data Sources

Datasets from Edgap.org were used for school level performance measures and community demographic indicators including Average ACT score, Median household income, Unemployment Rate, Percentage of adults with a college degree, percentage of students receiving free or reduced lunch. The datasets from NCES CCD contained the student count, staffing count, school level, charter status, and location information.

Data Preparation

All datasets were cleaned and merged into a single dataset using a Unique School Identifier (NCESSCH), a numeric code unique to each school in the United States. The merged dataset included 7,986 schools with ~99% matches between the Edgap and CCD datasets.

Categorical fields like school level or charter status were standardized to make sure there was consistent labeling. Finally, missing numerical values for average ACT were removed, and the rest were imputed by modeling each variable as a function of the others.

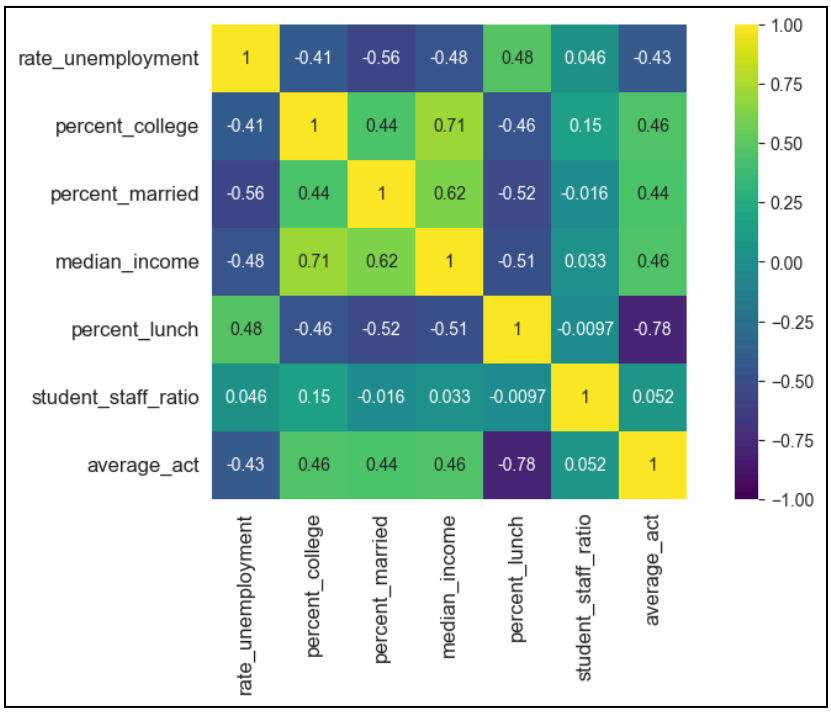
Variables and Analytical Approach

Average ACT score served as a measure of academic performance and was the dependent variable for the project. Two sets of independent variables were looked at: socioeconomic factors and student staff ratio. The socioeconomic factors included median household income, percent of adults with college degrees, unemployment rate, percent of households married, and percent of students receiving free or reduced lunch. The student staff ratio was found by dividing the number of students by the number of staff.

Exploratory data analysis (EDA) was used to visualize variable distributions and identify relationships. Correlation analyses were conducted to test the strength and direction of relationships between ACT scores and the independent variables.
Finally, statistical testing was used to compare the influence of student staff ratio and socioeconomic factors on academic performance. These analyses provided a base for determining if student staff ratio is as important as the socioeconomic conditions when trying to predict average ACT scores.

Computational Results:

A correlation matrix created in the initial EDA gave a clear picture of the strength and direction of the relationships among the variables. The results were fairly clear and were able to provide a loose answer to the project's question. The matrix showed that the strongest predictors were socioeconomic factors, in particular the percentage of students receiving free or reduced lunches (r = -0.78). On the other hand, student teacher ratio showed a near zero correlation (0.052). These results suggest that the staffing ratios play much less of a role in average ACT scores than socioeconomic factors.



Regression Models

In order to explore these relationships further, a series of single variable regression models were fitted. The model performance was evaluated using the R-squared value, and reinforced the patterns that were seen in the correlation matrix.The strongest individual predictor was the percentage of students receiving free or reduced lunch. This variable was able to account for about 61% of the total variance in the average ACT score on its own. Household median income was able to account for roughly 21% on its own, and the unemployment rate was able to account for roughly 18%. Median income and unemployment rate were also fit into quadratic models, which showed statistically significant, though small improvement.

Two multiple regression models were also fit, one using all five of the socioeconomic factors, and a reduced model with only three of the variables. The first model accounted for roughly 63% of the total variance in average ACT scores, and could predict a score within 1.15 points. However, there were variables that were not statistically significant, so a reduced model was fit using only significant variables. This reduced model accounted for the same ~63% of variance and was able to predict scores as accurately, additionally, statistical testing showed that adding the extra variables provided no statistically significant improvements to the model.

Discussion:

Variables such as the percentage of students receiving free or reduced-price lunch, median household income, and percentage of adults with college degrees showed the strongest correlations with ACT scores. In contrast, the student–teacher ratio displayed almost no linear relationship, suggesting that differences in staffing levels alone do not explain variation in academic performance across schools within these datasets.

However, two important data considerations emerged during the analysis. The first concerns the definition of staff used in the NCES Common Core of Data (CCD). The student–teacher ratio in the CCD may be derived from total school staff counts rather than strictly classroom teachers. This would inflate the denominator and discord

the true ratio of students to teachers, as a result the student staff ratio used may not be an accurate representation of the actual instructional ratio.

The second consideration involved the level of measurement for socioeconomic indicators, the median household income is based on census tract data, and may not be representative of the individual schools due to reflecting the broader community. In contrast, the percentage of students receiving free or reduced lunch is directly representative of the schools, and is likely the main reason why it was the strongest predictor.

These observations highlight an important lesson learned: when integrating datasets from multiple sources, variable definitions and geographic levels of aggregation must be carefully considered. Differences in how data are collected and who they represent can meaningfully influence analytical outcomes.

Conclusion:

This project was set up to determine whether student teacher ratio has as strong an influence on average ACT score as socioeconomic factors. Using school-level data from EdGap.org and the National Center for Education Statistics (NCES) Common Core of Data (CCD), the analysis compared several predictors of academic outcomes, including median household income, unemployment rate, educational attainment, and the percentage of students receiving free or reduced-price lunch.

Based on the datasets analyzed the results were clear: socioeconomic factors, in particular the proportion of students receiving free or reduced lunch, were far more predictive of average ACT Score than the student teacher ratio. While schools with smaller ratios may provide more individualized attention, the data showed almost no linear relationship between staffing levels and test scores. In contrast, socioeconomic factors explained over 60% of the variance in average ACT scores.

During the analysis two possible issues came up, first it was learned that the staff count used to calculate student staff ratio may include more than just teachers. This could make the ratio inaccurate and not representative. The second was how the data was reported, there are some variables such as median household income that are more reflective of the general community as opposed to the individual school. Taken together, these findings show the importance of context when interpreting data.

References:

EdGap. (n.d.). *EdGap.org – Data on school performance, demographics, and community context.*

National Center for Education Statistics. (2016–17). *Common Core of Data (CCD) — School Staffing and Student Counts*.