TO : Uncle Fred

FROM : Scott Emerson

DATE : February 14, 1996

SUBJECT : funding for schools

Here, finally, is the analysis I said I would do to try to settle our argument. I am not sure that this particular analysis succeeded in that vein. You can read through this, and we can pick up the debate again at our next get together. I note that I wrote this up somewhat formally, so that we can establish some common ground. To that end, I included a brief description of the points being contested. Please read that background section to make sure I am addressing our real point of contention. If there are errors there, the results of my analysis might not be too germane.

Summary

In an effort to assess the association between public school funding and educational achievement of students, I analyzed data from 21 states regarding their expenditures on education (per pupil expenditures and teachers' salaries) and the educational achievement of their students (average SAT scores for the state). In addition, I had available information on the per capita income for the state and the median years of education for adults over age 25. These data were analyzed in a linear regression model relating average SAT scores as a function of per pupil expenditure for each state. In that analysis, I found a statistically significant trend toward increased performance on SAT scores in states providing higher levels of public school funding ($P = .009$). Though this analysis suggested an association between higher funding and higher average SAT scores, it did not suggest a particularly large return for the money, in that a state spending an extra $300 per pupil (roughly a 10% increase for the average state) is only predicted to average SAT scores 9 points higher than an otherwise comparable state (95% CI: 2.5 to 15.2 points higher). Furthermore, the association found in this analysis is driven primarily by the relatively low average SAT scores observed in three Southern states (Georgia, South Carolina, and North Carolina) which also had very low expenditures on public education. Because those states share a lot of common features beyond their funding of public schools, I think we should exercise extreme caution in generalizing the results of this analysis to all states.

Background

Three weeks ago last Sunday, you and I got into a debate about whether the school levy for the Seattle Public Schools was justified. I claimed that more funding was necessary if our educational system was to improve. You felt that throwing money at the problem did not always help and that there were many other factors (such as parental influence) which were more important in determining student achievement. I promised to try to find some data that we could use to shed some light on the subject.

In scanning through some of my statistics textbooks, I did find a data set that was promising. This memo describes the results of my analysis of that data.

Questions of Interest

The overall goal of this analysis is to determine whether increased funding for public education is associated with better student achievement. In an attempt to address this question, at least in part, I performed a data analysis to test whether states spending more on education (on a per pupil basis) during 1982 tended to have higher average SAT scores.

Description of the Data

The data set consists of measurements on each of 21 states. The states with available data, as shown in Figure 1, are primarily Atlantic coast states, though California, Texas, Indiana, Oregon, and Hawaii are also included. For each state, we have measurements of the per pupil expenditure on public education in 1982, the average teacher salaries in 1982, the median years of education attained by the adult population over age 25 years, the average scores on the Scholastic Aptitude Test (SAT- a college entrance examination), and the per capita income in 1984.

In using these data to answer our questions, I will use the average SAT scores as a measure of student achievement, and I will use the per pupil expenditure on public education and the average teachers' salaries as measures of public school funding. The per capita income for the state is chiefly of interest as a measure of the cost of living within each state, although it is conceivable that the relatively 'richer' states might also be able to afford to spend more on public schools. In any case, we consider the possibility of wanting to make comparisons adjusting for the per capita income of the state.

The variable measuring median education of adults is a bit more difficult to address. On one hand, it is reasonable to assume that parents with more education would tend to encourage their children to take their education seriously. Thus we might expect states with higher median adult education to have higher average SAT scores. On the other hand, in order for adults to have higher levels of education, they must have done well on college entrance examinations like the SAT. Thus, in some sense the median adult education is in itself a measure of educational achievement in the state. Because of this latter aspect, I elected not to adjust my analyses of average SAT scores for this variable. In essence, the median adult education is a surrogate measure of educational achievement, though it is clearly a less direct measure than the SAT scores.

One additional note is in order regarding the variables measured. Although the per pupil expenditure and teachers' salaries pertain to public schools, the average SAT scores may reflect achievement of private school students as well. I do not have any way to assess the impact that this might have on the analyses, so I will assume that any such impact will be similar across states.

Statistical Methods

I analyzed the data using a linear regression model in which the expected average SAT scores for a state was assumed to be linearly dependent on the variables measuring per pupil expenditures, average teachers' salaries, and per capita income. Decisions regarding the variables to be included in the final model were based on statistical significance of the effects and the presence of confounding in the data. The adequacy of such a model was evaluated using analysis of the residuals. Statistical tests are summarized using two-sided P values and 95% confidence intervals.

Linear regression is a statistical analysis technique in which we assume that the expected average SAT scores for a state will depend on one or more of the variables measuring the level of state funding of education and the per capita income. That is, we assume that the expected SAT scores for a state can be described by the formula

$$E(SAT) = \beta_0 + EXPEND \times \beta_1 + SALARY \times \beta_2 + INCOME \times \beta_3,$$

where $E(SAT)$ represents the expected average SAT scores for a state having per pupil expenditures of $EXPEND$, average teachers' salaries of $SALARY$, and per capita income of $INCOME$. The purpose of linear regression is to find estimates for the unknown parameters β_0, β_1, β_2, and β_3.

Parameter β_0 is a mathematical construct that is not of particular interest in this analysis. Parameters β_1 and β_2 are more direct measures of the effect that funding of public schools might have on educational achievement. If β_1 is a positive number, the above model would predict that states spending more money on public education would tend to have higher average SAT scores. Conversely, if β_1 is negative, the model would predict that states with higher per pupil expenditure on public education would have lower average SAT scores. The exact value of β_1 indicates the expected difference in average SAT scores between two states which differ in their funding for public education by $1 per pupil. A value of zero for β_1 suggests that there is no association between level of per pupil expenditures and average SAT scores.

Similar interpretations apply for β_2 with respect to average teachers' salaries and β_3 with respect to per capita income. In each case, the value of the parameter indicates the expected difference in average SAT scores between two states differing in their measurements of the corresponding variable by $1.

An alternative measure of association among the variables is the correlation. Correlations are values between -1 and 1 which measure the tendency of states with higher values for one variable to also have higher values of another variable. Two variables with a correlation near 1 tend to 'move together' in that states will have either higher values for both variables or lower values for both variables. Two variables with a correlation near -1 tend to have lower values for one variable and higher values for the other. Two variables that are independent of each other will have a correlation near zero.

When the values of the parameters are unknown (as is generally the case), we can use a technique called least squares estimation to guess the true values. As in any estimation technique, however, our guesses are not precise. Hence it is often of interest to assess whether we can be reasonably confident that there is in fact an association between, say, average SAT scores and per pupil expenditures for public education. In this case, an association exists if β_1 is nonzero. A common method of assessing the presence of an association between two variables is the P value. This measures the probability that our results are due to random chance in the absence of a true association. Low P values are interpreted as results that are unlikely to be due to random chance. It is quite common for P values less than 0.05 to be regarded as evidence of a true association between the variables ('statistically significant').

A related method of assessing our confidence in an association between the variable is to compute a 95% confidence interval (CI) for, say, β_1. Such an interval gives a range of values that are in some sense reasonable guesses for the true measure of the effect of per pupil expenditures for public education on average SAT scores. If the 95% CI includes zero, then we do not have sufficient evidence to declare an association between average SAT scores and per pupil expenditures for public schools. If the 95% CI does not contain 0, then we can state with 95% confidence that there is an association between the variables beyond that that could be explained by coincidence.

The interpretation of P values and confidence intervals is based on some technical assumptions about the nature of the data. The appropriateness of these assumptions can be assessed through statistical techniques called residual analysis. Our only concern with such analyses is that they not demonstrate clear evidence that the technical assumptions are not valid. Generally, if a residual analysis demonstrates no particular problems, the results of the linear regression analysis can reasonably be trusted.

Results

Table 1 presents descriptive statistics for the variables measured on each state. Several variables were found to be highly correlated. Per pupil expenditures for public schools and average teachers' salaries had a fairly high correlation of .76, not a surprising result given the similar nature of the variables. Similarly, the correlations between per pupil expenditures and per capita income (.64) and between average teachers' salaries and per capita income (.60) were quite high. The highest correlation among the variables in the data set was between average SAT scores and median adult education (.77), thus confirming my suspicion that median adult education is a surrogate measure of educational achievement.

Table 1

Variable	Mean	Std Dev	Min	Median	Max
EXPEND	2773.5	598.5	1862	2763	4280
SALARY	19991.5	3144.6	13040	20625	25000
INCOME	12790.1	1640.0	10075	12729	16370
ADULTED	12.47	.17	12.1	12.5	12.7
SAT	877.9	31.8	790	888	925

Figure 2 displays a graph of average SAT scores versus per pupil expenditures for public schools (Figure 2a) and versus average teachers' salaries (Figure 2b). In both cases I display a line tracking the trends in the relationship between average SAT scores and the measures of funding of public education. The data points are labeled by the state abbreviation. In each graph, a tendency for increased average SAT scores in states with higher levels of funding is evident, although the trend is lessened as the level of funding gets very high.

A linear regression model predicting average SAT scores according to level of per pupil expenditures for public schools finds a statistically significant trend toward higher SAT scores in states with higher levels of funding (P = .009). The estimate for β_1 is .0296, suggesting that a state spending $1 more per pupil on public schools will tend to average SAT scores .0296 points higher than an otherwise comparable state. This same model would suggest that a state spending $300 more per pupil on public schools will tend to average SAT scores 8.88 points higher (95% confidence interval 2.5 to 15.2 points higher) than an otherwise comparable state. There was no evidence to suggest that there were any violations of the assumptions necessary for such an analysis to be valid.

From Figure 2, it is apparent that the increase of average SAT scores associated with increased per pupil expenditures is not consistent across all states. In fact, if the three states Georgia, South Carolina, and North Carolina are omitted, we have no evidence for an association between average SAT scores and funding of public schools (P = .57). Those three states are geographically clustered, and there may be other, unmeasured variables which explain their relatively low average SAT scores.

When modeled alone, average teachers' salaries were not found to be significantly associated with average SAT scores (P = .118). The best estimate for a trend was that a state having average teachers' salaries $1,000 more than another would be expected to average SAT scores 3.6 points higher than the other. It should be noted, however, that in the absence of statistical significance, this observed effect could merely represent a coincidental observation when no association truly existed between average SAT scores and average teachers' salaries.

In attempting to fit models assessing the effect of funding of public schools on average SAT scores while adjusting for per capita income, I found no improvement in the models. Adjusting for per capita income did not substantially change the estimated effect of per pupil expenditure on public schools, although it did affect the statistical significance. Such a result is to be expected in such a small data set having a high correlation among the variables.

Similarly, no advantage was afforded by simultaneously exploring associations between average SAT scores and per pupil expenditure while adjusting for average teachers' salaries. With a data set having only 21 observations, we are severely limited in our ability to make detailed inference. I do note, however, that when modeling both per pupil expenditure and average teachers' salaries, the estimated effect of increased average teachers' salaries when holding total per pupil expenditures constant is to decrease the average SAT scores (that is, β_2 is negative). If such a trend were statistically significant (it is not: P = .57), one interpretation might be that average SAT scores are increased by increasing per pupil expenditures on items other than teachers' salaries. I stress the speculative nature of this analysis.

Discussion

There are a great many problems with this data analysis. First, it represents what is called an 'ecological study'. That is, we examine statistics on a statewide basis, when our true goal is to assess effects on individuals. Within any state there are undoubtedly variations among the school districts with respect to levels of funding. It may well be that the students with the higher SAT scores were schooled in districts with lower levels of funding. In this case, we have no data to suggest that the students doing better on the SAT were those who were at schools that received higher levels of funding. Such a situation is called an 'ecological fallacy', and in this case it would include the possibility that students at private schools are overly represented in our statistics.

An additional problem may be due to differences among the states with respect to the students taking the SAT. While SAT scores are required by many colleges, such a requirement is not universal. Some state university systems do not require the SAT of their applicants. Hence some states may only have their best students (those good enough to get into out of state or private colleges and universities) take the SAT tests, while other states have a higher proportion of their students take the SAT. Again, we have no data available to assess this problem.

The trend that we observed in increased average SAT scores among states spending more money on public schools was not a strong one in the sense that a lot more money has to be spent in order to raise the average SAT scores a very few points. It is also disturbing that the evidence of a trend in average SAT by funding level is based entirely on the data from three neighboring states. Clearly the generalizability of our results is suspect in such a small sample size with so much regional correlation among our cases.

Lastly, I find it somewhat interesting (but distressing for someone in my profession) that we estimate a negative effect of teachers' salaries on average SAT scores when adjusted for total per pupil expenditures. Such adjustment means that if we compare two states having the same total per pupil expenditure on public schools, the state with the higher average teachers' salaries will be expected to have (very slightly) lower SAT scores. The lack of statistical significance suggests this observation can be explained very well by random chance, but it certainly does not present strong evidence in favor of raising teachers' salaries.

In conclusion, then, I think we would have to regard the results of this analysis as equivocal at best. If there is evidence that increased funding leads to better SAT performance, it would only seem to hold at the lowest levels of funding. I do not know if Seattle is currently below such a threshold. I will have to continue my search to find a better data set to settle our argument. In the meantime, I will say that giving schools less money is quite unlikely to improve the achievement of the students, and that this analysis did provide some meager evidence that it might help.

EXPEND vs SAT by State