A Primer on Statistical Methods in Education Research

In order to understand the importance of teachers’ impact on students, one must first understand the statistical methodologies used to measure the impact of various factors, such as teachers, on student achievement. There are numerous influences on the academic achievement of students; innate abilities may predispose students to achieve in particular subject areas; family situations can create home environments that either foster or discourage learning; and neighborhood and peers can set a tone that determines the social status of learning. These influences, as well as many others, are outside the direct control of educators. Thus, when assessing how education factors, such as teacher experience or class-size, actually impact student achievement, researchers attempt to use methodologies that isolate the effects of these education factors from family or personal factors that also influence student achievement.

Data and the Measurement of Student Achievement

To properly analyze the relationship between student achievement (the dependent variable which is the behavior of interest and varies according to its influences), and education factors (the independent or explanatory variables that influence the behavior of the dependent variable), one must have detailed and accurate data. Some of the most common teacher-level variables in datasets are teacher degree-level, type of teacher licensure, years of teaching experience, and the performance of teachers on various tests administered to them.

If our dependent variable is student achievement, datasets should have student achievement information from at least two different points in time so that one can
measure how much that student has learned over a defined period of time, measured as the difference between the first or "baseline" achievement score on a standardized test to the later, second score on the same standardized test. Growth in student achievement is a more accurate way to understand the affect that current education factors have on student achievement, rather than looking only at scores from one period in time.

Another important feature of data is that the variables included in any models (which are explained below) are robust and measured accurately. Unfortunately, most datasets have variables that are to some degree mis-measured. For example, studies often determine a student's class-size by dividing total number of students in the school by the total number of teachers (or professionals) in the school. But clearly not every student has the same size classroom and not every professional in the school has their own classroom or the same number of students. The variation in class-size within schools means that we may not have properly characterized the true class-size for a particular student or group of students. Often, researchers create proxy variables that attempt to measure something that is difficult to quantify. The effects of such variables can be difficult to interpret because the variables themselves are not reflective of actual values but instead are attempts to reflect non-numeric information. For example, teachers who have advanced degrees may be more effective in producing growth in student achievement (this point is discussed in detail below), however, it is not the degree itself that causes them to be more effective, rather the degree serves as a proxy for their knowledge as a teacher. Teacher knowledge is the truly important factor in making these teachers effective, not the knowledge conveyed by the degrees themselves. This concept may be illustrated by examining the estimated effects that parental income has on student
achievement. As not all parenting practices are easy to quantify, they are typically omitted from datasets and models so that their influence on academic achievement is attributed instead to one variable like family income. Thus these models tend to over-predict the importance of this quantified variable. For example, parents’ education level tends to correlate with their income, and both have a relationship to student achievement. If a variable measuring parents’ education level is absent from a data set, but a variable for parents’ income is available and used in a model, then the omitted parents’ education variable may bias the parents’ income variable’s relationship with student achievement, making it appear as though parents’ income has a greater relationship with student achievement than it actually does.

Most studies show that parental income has a strong positive impact on a variety of student outcomes including student achievement. However, Mayer (1997) who found such an effect, makes a strong case that the relationship between parental income and students’ achievement is *correlational* rather than *causal*. Statistical analysis allows researchers to understand correlations, but not causes; statistics cannot determine the cause of an event. However, statistics can help researchers to predict how well certain conditions do or do not correlate to the outcome of an event (e.g., a child takes standardized tests over time, and the growth in the child’s scores can be correlated with certain factors that relate to the outcome of the growth in test scores).

**Methodology**

Regression analysis, which uses mathematical models to make predictions about outcomes, is the typical statistical tool researchers use to estimate relationships between
education factors and student achievement or other dependent variables.\textsuperscript{1} A key feature of regression analysis is that, if the model that is being analyzed is correctly constructed by a researcher, then the regression will accurately establish the relationship between the dependent and independent variables, while simultaneously accounting for other variables that could influence the dependent variable. Accounting for the other variables involves holding their influence in the model as constant; a condition referred to as ceteris paribus.\textsuperscript{ii} Thus, it is possible to assess how school factors, such as class size, influence student achievement while other factors, such as family income, are held constant. This is an important step if one wants to isolate the impact of school factors on student outcomes from other influences.

The most common regression model used in education studies is referred to as the "educational production function.\textsuperscript{1}" The educational production function dates back to the Equality of Educational Opportunity report in 1966 (also known as the “Coleman Report”). Since then, literally thousands of studies have utilize the educational production function to try to better understand how various aspects of education, such as per pupil expenditure, class and school size, and teacher degree and experience level affect an outcome like student achievement. The production function is made up of inputs that predict the outputs or outcomes. Some inputs will correlate to the outcome in a positive way—as the input increases, one can predict that an increase in the outcome is

\textsuperscript{1} It is important to note that the production function formulation calls for identifying the maximum possible level of output from each conceivable combination of inputs. This is a significant stipulation and is what makes it possible to derive causal inferences from a true production function. The fact that it is difficult to identify the maximum possible part of the formulation in education applications is what ultimately undermines many causal claims.
more likely to occur. Likewise, some inputs will correlate negatively—as the input increases, the outcome decreases or is less likely to occur and visa versa.

The relationships between various education inputs and an output can be understood through the coefficient estimates that regressions predict. Coefficient estimates are what the mathematical model (production function) of the regression analysis uses to show the relationship between the inputs (independent variables) and the outcome (dependent variable). Coefficients can be understood by imagining a study that looks at the effect of years of teaching experience on student test scores. A positive coefficient of 10 for the variable years of teaching experience would indicate that for every one unit increase in the number of years a teacher has taught, that teacher's student test scores are likely to be 10 points higher than students with a teacher who has one year less of experience, holding all other factors constant. A positive coefficient signifies a positive relationship between the independent variable and the outcome, meaning that they change in the same direction. A negative coefficient reflects the negative relationship between the explanatory variable and the outcome, reflecting that the variables change in opposite directions from each other. A coefficient of -10 for years of teaching experience would mean that students that have teachers with one additional year of teaching experience are likely to score 10 points below students with teachers with one year less of teaching experience.

Understanding Results and Drawing Conclusions

As discussed above, the coefficients that are generated from regression analysis allow us to predict the relationships and effects of the independent variables on the
dependent variable. The magnitude of the coefficients alone, however, is not enough to draw conclusions about the relationships between various education inputs and student outcomes; the accuracy of the derived relationships must also be considered. The accuracy of an observed relationship is understood in the context of statistical significance, which is a measure of how confident we are that the coefficients we estimate are actually different from a specified value. Typically this specified value is a zero relationship between an independent and dependent variable (a coefficient of zero implies that a change in an independent variable has practically no effect on the dependent variable). While there is no objective or universally agreed upon criteria for what constitutes a variable having statistically or educationally significant effects, it is common practice among statisticians to use the 95 percent level as a cutoff for statistical significance. This means that there is only a 5 percent chance of committing what is known as a type I error. Type I errors wrongly reject the null hypothesis (the hypothesis that an estimated relationship – coefficient – is different than our specified value) and the true effect of a variable on the output is actually zero.

What does all this mean for measuring teacher quality? First, to accurately measure the impact of teachers (or any educational resources), it is necessary to have detailed data. At a minimum, data sets should have student achievement information that covers at least two points in time so that value-added models may be estimated. Second, the results from statistical models are quite often sensitive to the type of different regression models used in the analysis and the construction of those models with the variables, and thus one should be careful about drawing too strong conclusions from single (or a handful of) studies. Third, there is no objective cutoff of what constitutes
statistically or educationally significant results. For these reasons, it is appropriate to look for broad patterns of results from a variety of different studies when investigating the impact of teachers and the correlates of teacher quality. These complications help explain why two well-conducted studies on similar topics can have different results and conclusions.

---

1 Experimental studies, where individuals are randomly assigned to either a treatment or control group, represent the most rigorous research design for controlling for the direction of causality and unobservables, however, for many reasons, educational experiments are rare. Regression methodologies likely represent the next best option to assess the impact of educational resources on students.

2 In practice, many of these characteristics cannot so easily be held constant. It is rare, for instance, to see students from very affluent families in schools with low per pupil spending because affluent families tend to live in areas that support high levels of school spending. Still, this is a useful framework for public policy purposes because policymakers need to have a sense of how different educational investments are likely to affect students.