SOME CONSIDERATIONS ON THE USE OF ANALYSIS OF COVARIANCE IN EDUCATIONAL RESEARCH

Daniel D. Blaine

In the classical experimental approach to the investigation of complex phenomena, the behavior of objects of interest are observed under conditions that are equated with respect to all variables—except the variable whose effect is of interest. Differences in observed behavior are assumed to be due to the variable which has varied from one condition to the next. The control of extraneous variables is critical to the inferential logic which leads to lawful statements about the relationship between an independent variable (the variable being manipulated) and a dependent variable (the variable assumed to vary as a function of the independent variable).

As the phenomena of interest become more and more complex, the control of all the relevant, but extraneous, variables becomes a functional impossibility. To circumvent this problem, laboratory experimenters have traditionally resorted to random assignment of the objects of interest (the objects of observation) to the various conditions defined by the independent variable. It is then assumed that the effects of any extraneous variable will be "balanced out." Any systematic differences among the groups of objects is then attributed to the independent variable.

Educational researchers have attempted to apply this logic of experimentation within the educational setting to evaluate instructional programs, but have often had to deal with the problem of being unable to randomly assign the objects of interest to the treatment groups. In the case of educational research, the objects of interest are, more often than not, students in a classroom. Problems—administrative, logistical, social, etc.—are associated with randomly assigning students to classrooms so that an educational experiment can be conducted are overwhelming. So overwhelming, in fact, that the highly desirable (from an experimental point-of-view) random assignment rarely occurs. The educational researcher, then, is left with making inferences based on "intact" groups or classrooms of students. The problem, then, becomes one of whether systematic differences between groups are due to the experimental treatment under investigation (e.g., methods of the teaching of reading), or to some pre-existing difference (e.g., differing intelligence levels) resulting from the nonrandom procedure of assigning students to classrooms.

As educational researchers became more aware of a statistical technique known as analysis of covariance, many felt that a statistical answer to a logistical problem had been found. Essentially, analysis of covariance is a procedure by which groups are "statistically" equated on one or more extraneous variables. (In laboratory experimentation, it is assumed that the groups are equated on these variables through the random assignment procedure.) For example, two methods of reading instruction might be compared by teaching two different classrooms of students, each with a different instructional method. At the end of a term, the two groups would be observed and compared with respect to reading achievement. Suppose that one of the groups has a higher reading achievement. Suppose, also, that the group with higher reading achievement scores also has higher IQ scores. Assuming for a moment that there are no other differences between the groups, the inferential dilemma becomes one of whether the difference in reading achievement is attributable to the difference in teaching method, the difference in intelligence, or both.

This problem is approached analytically by the analysis of covariance technique. Essentially, the technique starts by assuming that the expected change in reading achievement for a unit of increase in intelligence is the same for both groups. The expected change can easily be estimated by a least squares regression line, the slope of which is the estimate of expected change. For example, if the regression line had a slope of 2.0, that would mean that for each increase of one IQ score point the reading achievement score would be expected to
increase by two points. Using this information as the basis for an adjustment procedure, the analysis of covariance technique addresses the question of how the groups would compare on reading achievement if they had been equal on intelligence. The slope of the regression line provides the “handicapping” system for adjusting the groups to a common value on the extraneous variable (covariable). More technically, the dependent variable is residualized with respect to the covariable and the residuals which are uncorrelated with the covariable are analyzed for systematic covariation with the independent variable.

Since the logic of analysis of covariance can be extended to more than one covariable, it is especially attractive to the researcher who must make inferences based on intact groups. On the surface, analysis of covariance would appear to make it possible to deal with situations in which a variety of extraneous variables are correlated with the groups defined by the independent variable. By “partialling out” all of the differences due to all possible covariates, the researcher would eventually reduce the dependent variable down to the point where the differences remaining would most likely be due to the independent variable. The procedure of analysis of covariance, however, has certain characteristics and is based on certain assumptions which make it less than desirable for its most commonly explained purpose, which has just been described.

It is now more widely recognized that analysis of covariance is not appropriate as a method for controlling the effects of any and all extraneous variables that might influence the results of comparative observations. It is tempting to adjust the dependent variable for all possible sources of
influence other than that of the independent variable of interest. Such a procedure has an extreme potential for capitalization on chance so that differences in the dependent variable which should be attributed to the independent variable are “removed” because of a chance correlation with one of the variables.

Although the “use-everything-as-covariables” strategy has mostly disappeared, it is still common to find researchers using the analysis of covariance technique when one or maybe two extraneous variables are correlated with the independent variable. The decision to treat an extraneous variable as a covariable in an analysis is often based on the degree of correlation between the covariable and the independent variable (i.e., the degree to which the treatment groups differ with respect to the covariable). Although the discussion of the analysis of covariance technique by many authors would suggest that this is a very legitimate approach to the decision as to which, if any, extraneous variables should be used as covariables, the analysis of covariance model becomes more questionable as the correlation between the potential covariable and the independent variable increases. Technically, the analysis of covariance assumes that the treatment groups were sampled from populations having equal means on the covariable. This means that analysis of covariance is not appropriate when the groups exhibit large differences in covariable means, which is the situation in which analysis of covariance is most attractive given the typical description of the “logic” of the technique. Thus, an analytic paradox: the more useful analysis of covariance would seem, the less appropriate it becomes as a statistical model. Generally, caution should be exercised in the use of the technique when dealing with groups for which there are large pre-existing differences.

An additional assumption underlying the technique of analysis of covariance is that the expected change in the dependent variable for a unit of increase in the covariable is the same for all treatment groups. If this is not the case, then the adjustment procedure described earlier is not appropriate. Textbook treatments of analysis of covariance usually mention the assumption and often will describe how the assumption can be tested; however, little discussion is given to the implications of the rejection of the assumption. It is within the realm of the homogeneity of regression or the lack thereof that some of the more interesting results with respect to education may lie. If the expected change in the dependent variable for a unit of increase on the covariable is not the same for the groups, then the covariable and the independent variable are said to interact. That is, the effect of the independent variable cannot be discussed without the levels of the covariable being taken into account.

Returning to the example of the effect of methods of reading instruction using intelligence as a covariable, suppose that the unexpected change in reading achievement for a unit of increase in IQ was +2.0 for a group taught by one method and −1.0 for the other group. In such a case, “homogeneity of regression” is violated and an adjustment procedure such as the analysis of covariance would not be appropriate.

However, a result much more interesting than the failure to meet an assumption underlying a statistical model has occurred. The expected changes (+2.0 for one group and −1.0 for the other) can be interpreted geometrically as the slopes of the regression lines (reading achievement regressed on IQ) for the respective groups. When the slopes are not equal, this implies that the regression lines are not parallel. For any level of intelligence, the group with the “higher” regression line is higher in reading achievement than the other. If the lines are not parallel then, for certain levels of intelligence, one group will be higher in reading achievement, while the other group will be superior on all other levels. That is, the difference between the methods of instruction changes as a function of changes in intelligence. With regard to educational decisions, such a result would imply that one method of reading instruction would be more appropriate for students of a given range of intelligence, while the other method would be more appropriate for a group with a higher (or lower) range of intelligence. This type of prescription would seem to be one which would be highly desirable if educational procedures are to be adapted to the individual in such a way as to optimize his chances of success.

In summary, a statistical model which has been commonly used in evaluating research results from intact groups was shown to have serious limitations in its most commonly-touted form. It does, however, have the potential for being a viable analytic model (under certain circumstances) capable of orienting the researcher toward the investigation of educational phenomena which, when more fully understood, will provide educators with more prescriptive formulations for intervention in the learning process.

Daniel D. Blaine is Associate Professor, Department of Educational Psychology. The series of statistic courses which he teaches is part of the quantitative core for graduate study in educational psychology, psychology and social work.