Discussion of AERA 1986 Session 21.25 Applications of Multiple Linear Regression

Bruce G. Rogers
University of Northern Iowa

Comments on the Paper by Joe Ward:

Since I have not seen the full paper, I will need to base my comments on the short draft I received. It proved to be an innovating application of both a utilitarian philosophical viewpoint and interaction in a simple 2-way ANOVA.

The model is based upon the criteria of maximizing the learning when summed across all students. This is reminiscent of one of the 19th century philosophical discussions on ethics. Jeremy Bentham (b 1748) developed the concept that the criteria of the goodness of a policy was determined by calculating the good for each individual and then summing up the individual goods. Sometimes it is called the calculus approach, in reference to integration as the summing of the values. And that is what is done in the table entitled "Optimality Index Values." For every possible way of assigning the four students to the four teachers, the sum of the Optimality values is computed. Then that particular assignment of pupils with teachers which yielded the maximum sum is chosen as the desired assignment.

Bentham was aware that sometimes the principle of the "greatest good for the greatest number," when applied to public policy, could come in conflict with what a particular individual perceived as their own greatest good. I told Joe that many principals might be hesitant about applying this model for fear of confronting irate parents who wanted another choice. For example, if the parents see the table of Predicted Values, it is likely that all of them will request that their child be placed with Teacher 1. Trying to convince any one parent to allow their child to be put with a less-than-best teacher in order to maximize some abstract "Optimality Index" may prove to be very challenging. Indeed, it is my understanding that many principals randomly assign pupils to the teachers, when several teachers are teaching the same grade, in order to avoid possible charges of favoritism toward teachers and pupils. But Joe assured me that in some districts (including the one in which his wife taught) the principal and the teachers do consult on how to best assign the students. Given that such decisions are to be made, the Ward procedure has the definite virtue of providing an unbiased approach.

The procedure uses a two-way ANOVA interaction design. It is a variation of the aptitude-treatment interaction, where aptitude is past performance and treatment is the teacher. Richard Snow, Lee Cronbach, and others, have worked extensively to find such interactions, with limited success. However, since the teacher is such an important variable in the classroom, it is possible that this approach will prove to be an efficient method of detecting such interactions.

I like the term "catalytic" variable. In chemistry, we take two compounds which react very slowly or not at all. However, when we add a catalyst, the reaction is speeded up, but the catalyst is not affected. In Figure 1, only a weak interaction is present, but when the catalytic variable is added, a strong interaction is observed, as seen in Figure 2. And the resulting "Optimal Sum of Payoff Values" is increased fourfold, as a result of this interaction.

Let me conclude by making a practical suggestion to the authors. Special computer programs were written to compute the tables. Is it ossible to do this with regular routines in MINITAB, SPSS, SAS, BMDP, etc.? If so, it would be useful to describe how that is done, thus making the procedures easily available to a large number of readers.

Comments on the papers by Jerome Thayer

In the paper on Model Building, attention is given to a set of widely used approaches to variable selection in multiple regression. It is pointed out that no technique should be used indiscriminantly, but rather, that user judgment should be used to determine that set of predictor variables which will be most interpretable.

These techniques were applied to a variety of data sets, ranging from real world data to contrived data. The results in Table 1 suggest that, in general, the Stepwise method is a desirable procedure, but that exceptions do exist. Therefore, the general consensus does seem to support the author's conclusions.

A suggestion might be made for this paper. The "Best Subsets" program was obtained from BMDP, but is not available in SPSS. What are users to do if only SPSS is available to them? A look at Figure 1 suggests that if the Stepwise and Backward procedures were run, and the highest R^2 selected, the results would not be substantially different from using the Best Subsets procedure. While this point is implied in the paper, perhaps it could be made more explicit.

Thayer's paper on Dichotomous Variables shows an empirical example of the mathematical equivalency of several least squares statistics. The paper first points out that a number of writers in the behavioral sciences have argued that regression is inappropriate for data in which the dependent variable is dichotomous. Thayer chose not to attack the critics directly, but used that well-known proof model from geometry, reductio ad absurdum. A set of data is analyzed twice, using the dichotomous variable first as the dependent variable and then as the independent variable. The results are shown to be identical. It is then concluded that if the reasoning of the critics was followed to its logical conclusion, it would be necessary to discard t-test, ANOVA, ANCOVA, discriminant analysis, and multiple regression. It would be interesting to hear how the critics would respond to this argument.

But let me suggest a reason why one might prefer a computer program specifically written for each of the above routines, rather than using a regression program only. While it is possible to show that, on a two-group comparison, the t-test, F-test, and simple correlation are mathematically equivalent, the computer output for each is not in the same form. Thus, the square root of F must be taken to get t, and a more complicated transformation must be made to get r to t. It is also true that a 2 group discriminant analysis is the same as multiple regression on a dichotomous dependent variable, but again the computer output looks different. And for more than two groups, the output is much different. If the transformations are not made correctly, then serious differences can result. While that is not the situation that the critics had in mind, it is a legitimate reason why a person might use a technique other than regression.

But I digress. This does not detract from Thayer's basic conclusion that the underlying theory of the various least squares techniques is the same, and therefore all of them can be considered as special cases of multiple regression, canonical correlation, or multivariate analysis of variance (SPSSX uses the latter procedure as an umbrella). Conceptually, this is a powerful tool for helping the student to see classical statistics as variations on a major theme rather than as a "bag of tricks."

My only suggestion for this paper is that the layout of the tables and the use of the t values may prove difficult for the reader to follow. Perhaps the author will submit the paper to a colleague or a student, and if they have similar difficulties, revise the layout to strengthen the presentation.

Comments on the paper by John Morris

The Morris paper begins by stating that the primary concern in regression is the predicting of accurate criterion scores, rather than the estimating of population regression weights. While it is true that, in the theoretical sense, these two criteria are comparable (i.e., you cannot have accurate criterion prediction without accurate regression weights), it is also true that the beta weights may change if a different type of regression is used (e.g., ridge regression). But in both cases, the ultimate focus is upon the accuracy of the criterion scores.

The PRESS Algorithm was designed to select a multiple regression model variable subset that would minimize the Sum of Squares on Cross Validation. This is somewhat akin to the "best set" selection of which Thayer spoke. The philosophy of cross validating the total choice process (p. 13) by omitting one subject at a time is akin to the "Jackknife" procedure.

In the computer runs, "real" data was used instead of data from Monte Carlo simulations. That definitely has the advantages that are mentioned (p. 15) but also has the disadvantage that one does not know a priori which assumptions are violated and why, whereas with Monte Carlo data we can specify and create the violations. Perhaps in a revision of this paper it would be useful to discuss both the strengths and weaknesses of these procedures.

The results show that, for most cases, the OLS is sufficient and even better than the other methods. I like this conclusion. It is compatible with my own philosophy of techniques. Some people complain that we use statistics without carefully analyzing the data to see if it meets all the assumptions. But I suggest that if the data even vaguely looks appropriate, we can submit it for computer analysis. Thus, we can examine the results. Do they make sense? If not, what violations might account for it? And how might the data be transformed or the procedure modified to make better interpretable results? The results of this study seem consistent with that. Ridge regression and the techniques have an important place, but for most data we should first look at OLS, and then try other techniques where appropriate. The PRESS algorithm, available on a microcomputer, can then provide an effective way to address this selection problem.