MULTIPLE LINEAR REGRESSION MODELS WHICH MORE CLOSELY REFLECT BAYESIAN CONCERNS[1]

ISADORE NEWMAN, Educational Foundations, The University of Akron, Akron, OH 44325
ERNEST L. LEWIS, Guidance and Educational Psychology, Southern Illinois University, Carbondale, IL 62901
KEITH A. McNEIL, National Testing Service, Durham, North Carolina
STEPHEN S. FUGITA, Department of Psychology, University of Akron, Akron, OH 44325

The purpose of this paper is to discuss relationships between the Bayesian and the multiple linear regression approaches of hypotheses testing. In a frequently cited article, Novick and Jackson (1970) suggest that Bayesian inference can increase the efficiency of test utilization in educational guidance services. They argue that Bayesian regression equations provide more accurate predictions than traditional regression equations when the same information is available to both methods for the purpose of calculating regression weights. Some of the problems which Novick and Jackson associate with traditional regression equations, however, can be eliminated through the use of appropriate multiple regression models.

Consider the question of how one might predict the success of an individual in a given college department. If information was available on previous individuals from the particular department in question, the multiple regression equation would be:

Model 1: $Y_1 = a_0 U + a_1 X_1 + E_1$
where: a_0 and a_1 = partial regression weights
Y_1 = criterion established to indicate success in the college department
X_1 = predictor variable
U = unit vector
E_1 = error in prediction

If a second department, for which infor-

mation was available, was under consideration the multiple regression equation would be:

Model 2: $Y_2 = b_0 U + b_2 X_2 + E_2$
where: b_0 and b_2 = partial regression weights
Y_2 = criterion established to indicate success in the college department
X_2 = predictor variable
U = unit vector
E_2 = error in prediction

However, if a third department, which is thought to be similar to the first two, were under consideration but little criterion information was available with respect to that department, Novick and Jackson (1970) suggest that the multiple regression equation would have to be:

Model 3: $Y_3 = c_0 U + c_1 X_3 + E_3$
(there would be very weak power because of the small N size)
where: Y_3 = criterion score
X_3 = predictor variable
U = unit vector
c_0 = the Y-intercept for all students
c_1 = the slope for all students
E_3 = error in prediction

According to Novick and Jackson (1970) the Bayesian prediction equations for a department on which no information is available would be:

$$\hat{Y}_4 = \hat{\alpha} + \hat{\beta} X$$

where: α and β are determined by pooling the data from all departments similar to the one in question and then correcting the covariance and variance terms in some way. The exact nature of these corrections is not explained in their article.

It seems that Novick and Jackson are correct in concluding that the Bayesian equation presented above can predict more accurately than the traditional equation. A multiple regression model may, however, be written which incorporates more information than does Model 3. The following model would allow for two regression lines, each having different Y intercepts (a_1 and a_2) and different slopes (a_3 and a_4):

Model 4: $Y_4 = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + E_4$

where: a_1, a_2, a_3, a_4, = partial regression weights

Y_4 = success criterion in college departments 1 and 2

$X_3 = 1$ if subject in college department 1; 0 otherwise

$X_4 = 1$ if subject in college department 2; 0 otherwise

X_1 = predictor variable for subject if in college department 1; 0 otherwise

X_2 = predictor variable for subject if in college department 2; 0 otherwise

E_4 = error vector

In this equation we are essentially utilizing the information available from departments 1 and 2, which is what Novick and Jackson imply cannot be done with traditional regression analysis. Ultimately, whether or not multiple regression Model 4 can predict as well as the Bayesian model is an empirical question.

In the above example, the Bayesians assume that the functional relationship between the predictor and criterion are similar in the two departments and that the third department is similar to the first two. By combining the data from the two departments, the sample size is increased and more stable regression weights are established. When a researcher is using the multiple regression

approach, he is likely to test the assumption that departments 1 and 2 have the same functional relationship. It might be that the two intercepts are not the same. In this event, the following model would need to be tested:

Model 5: $Y_4 = a_0U + a_1X_1 + a_2X_2 + E_5$

where: a_1, a_2, a_0, = partial regression weights

Y_4, X_1, X_2, are defined as in Model 4

$U = X_3 + X_4$ (from Model 4)

E_5 = error in prediction

Empirically, it may also be the case that the two slopes are not similar, even though the initial starting points are the same:

Model 6: $Y_4 = a_0U + a_3X_3 + a_4X_4 + a_5X_5 + E_6$

where: a_0, a_3, a_4, = partial regression weights

Y_4, X_3, C_4, are as defined in Model 4

$X_5 = X_1 + X_2$ (from Model 4)

E_6 = error in prediction

Model 5 can be tested against Model 4 to determine if the intercepts are statistically different over and above slope differences. Model 6 can be tested against Model 4 to determine if there are any significant differences in slopes over and above differences due to intercepts.

Both of these models should be statistically tested against Model 4. If statistical significance results, Model 3 cannot be utilized (since it assumes no significant differences in sope or intercept for pooling information) and must be discarded in favor of the model which has generated significance (in this case Model 4). When this occurs, then the researcher can also decide which department (1 or 2) the third department resembles most closely. For example, suppose that the two departments had dissimilar slopes and the predicted scores were to be obtained from equation (1):

(1) $\hat{Y}_4 = a_0U + a_1X_1 + a_2X_2$

Assuming that the empirical weights were determined, the equation (a) might be as follows:

(2) $\hat{Y}_4 = 6U + 3.2X_1 + 4.1X_2$

If the third department looked more like department 1, then the predicted criterion would be obtained from the following simplification of equation (2):

(3) $\hat{Y}_4 = 6 + 3.2X_1$

On the other hand, if the third department more like the second department, then the simplified prediction equation would be:

(4) $\hat{Y}_4 = 6 + 4.1X_2$

Notice that both equations (3) and (4) have the same Y-intercept, indicative of the fact that data from *both* departments were used to develop the Y-intercept in Model 6.

Researchers should not limit themselves to investigating rectilinear relationships. High degrees of predictability may be lost by ignoring non-linear relationships. Testing similarities between departments can be done with curvilinear multiple regression models as well as with linear regression models (e. g., McNeil *et al* 1975).

It is conceivable that a researcher would begin with, for instance, six different schools, and develop groupings of schools to find out which schools are alike in their functional relationship.

Model 7: $Y_5 = a_1U_1 + a_2U_2 + a_3U_3 + a_4U_4 + a_5U_5 + a_6U_6 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + E_7$

where: $a_1, a_2, \ldots a_6$ and $b_1, b_2 \ldots b_6$ = partial regression weights

$U_1, U_2, \ldots U_6 = 1$ if subject from corresponding college; 0 otherwise

$X_1, X_2, \ldots X_6$ = predictor variable for subject from corresponding college department; 0 otherwise

E_7 = error in prediction

Suppose that the following outcomes resulted. With the use of various restrictions that are empirically reasonable, schools 1, 5, and 6 are determined to have slopes that are not statistically significantly different. Schools 1, 5, and 6 have intercepts that are not statistically different from schools 2 and 3 (2 and 3 have a common slope, although different

from schools 1, 5, and 6). School 4 is unique in both intercept and slope. The resulting prediction equation would be:

(5) $Y_5 + P(U_1 + U_2 + U_3 + U_5 + U_6) + QU_4 + R(X_1 + X_5 + X_6) + S(X_2 + X_3) + T(X_4)$

where: P = common intercept for schools 1, 2, 3, 5, and 6

Q = unique intercept for school 4

R = common slope for schools 1, 5, and 6

S = common slope for schools 2 and 3

T = unique slope for school 4

This process has led to no increase in predictability for school 4, because it is different from all the others. More stable regression weights can be determined for the other schools, however, because their empirical similarity permits the sharing of data. In this way, more stable predictions can be made for schools for which there is very little available data. One of the selling points of Bayesian statistics is the ability to capitalize on small amounts of data. If the appropriate regression models are utilized, the multiple regression approach can also effectively utilize small amounts of data.

In all of the previous models, the testing of interactions is what distinguished the regression procedure. In most prediction studies, too often even plausible interactions are ignored and all subjects are lumped together and treated as similar. Our conceptual theories have long ago distinguished distinct groupings, and hence our statistical procedures should reflect this empirical possibility, whether they be Bayesian or multiple linear regression. Until the Bayesian methodology, however, has been empirically shown to be more predictive than multiple regression analysis, the availability and relative mathematical simplicity of multiple regression analysis would seem to be the procedure of choice and we encourage Bayesian statisticians to present their methodologies in a more simplified and straightforward manner than has been done to date (Meyer and Collier 1970, Schmidt 1969). Additional ex-

amples of clear applications of the Bayesian approach, such as the one by Pietz (1968) to the testing of theories of perception of rotary motion, would be helpful to the research community.

LITERATURE CITED

McNeil, K. A., F. J. Kelly and J. T. McNeil 1975 Testing Research Using Multiple Linear Regression. Southern Illinois Univ. Press, Carbondalle, IL. 587 pp.

Meyer, D. L. and R. O. Collier (Eds.) 1970 Bayesian Statistics. F. E. Peacock Publishers, Inc. Itasca, IL.

Novick, M. R. and P. H. Jackson 1970 Bayesian guidance technology. Rev. Ed. Res. 40: 459–494.

Pitz, G. F. 1968 An example of Bayesian hypothesis testing: the perception of rotary motion in depth. Psych. Bull. 70: 252–255.

Schmidt, S. A. 1969 Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics. Addison-Wesley Publishing, Reading, MA. 500 pp.