MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 3, Number 3
February, 1973

A publication of the Special Interest Group on Multiple Linear Regression of the American Educational Research Association

Editor: John D. Williams, University of North Dakota

President of the SIG: Bill Connett, University of Northern Colorado

Secretary-President Elect of the SIG: Judy Lyons, CIRCE, St. Louis

Table of Contents

AERA Convention

Members of the Special Interest Group are reminded that the schedule regarding SIG: Multiple Linear Regression in New Orleans is as follows:

Wednesday, February 28, 1973

Terrace Suite #3 (Juno) 2:15-5:15

STATISTICAL INTERACTION VIEWED WITHIN THE MULTIPLE LINEAR REGRESSION APPROACH
(Special Address and Business Meeting, SIG/Multiple Linear Regression)

	Business Meeting
Chairman	William Connett, University of Northern Colorado, Greeley
	Special Address
Chairman	Gerald Schluck, The Florida State University, Tallahassee
Speaker	Statistical Interaction Viewed Within the Multiple Linear Regression Approach. Keith McNeil, Southern Illinois University, Carbondale
	Display
Chairman	Computer Regression Analyses. Gerald Schluck, The Florida State University, Tallahassee

Joe Ward and Earl Jennings' book is finally out. It is, Introduction to Linear Models, Prentice-Hall ($10.95).

MULTIPLE LINEAR REGRESSION MODELS WHICH MORE CLOSELY REFLECT BAYESIAN CONCERNS

Isadore Newman
The University of Akron

Ernest L. Lewis
and
Keith A. McNeil
Southern Illinois University

The purpose of this paper is to discuss several relationships between the Bayesian approach and the multiple linear regression approach. Having examined these relationships, both approaches may be better understood.

In a recent article, Novick and Jackson (3) suggests that Bayesian methodology can provide and increase in the efficiency of guidance services. They defend this position by indicating that Bayesian regression equations will provide more accurate predictions than traditional regression equations when the "same" information is available to both methods for the purpose of calculating regression weights. It seems, however, that some of the problems which Novick and Jackson identify with respect to traditional regression equations 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 were available on previous individuals from the particular department in question, the multiple regression equation would appear as follows:

Model 1 $Y_1 = a_0 + a_1X_1 + E$

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

Model 2 $\quad Y_1 = b_0U + b_1X_1 + E_2$

where: $\quad b_0$ and b_1 = partial regression weights

Y_1 = criterion established to indicate success in the college department

X_1 = 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 no criterion information was available with respect to that department, Novick and Jackson suggest that the multiple regression equation would have to be:

Model 3 $\quad Y_1 = c_0U + c_1X_1 + E_3$

where: $\quad Y_1$ = criterion score

X_1 = 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, the Bayesian prediction equation for a department on which no information is available would be:

$$\hat{Y}_1 = \hat{\hat{\alpha}} + \hat{\hat{\beta}}X$$

where: $\hat{\hat{\alpha}}$ and $\hat{\hat{\beta}}$ 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. However, a multiple regression model may 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 $\quad Y_1 = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + E_4$

where: $\quad a_1, a_2, a_3, a_4$ = partial regression weights

Y_1 = success criterion in college departments 1 and 2

X_1 = 1 if subject in college department 1; 0 otherwise

X_2 = 1 if subject in college department 2; 0 otherwise

X_3 = predictor variable for subject if in college department 1; 0 otherwise

X_4 = 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. 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 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, resulting in the necessary usage of the following model:

Model 5 $Y_1 = a_1X_1 + a_2X_2 + a_5X_5 + E_5$

where: a_1, a_2, a_5 = partial regression weights

Y_1, X_1, X_2 are defined as in Model 4

$X_5 = X_3 + X_4$ (from Model 4)

E_5 = error in prediction

It may empirically turn out that the two slopes are not similar, even though the initial starting points are the same:

Model 6 $Y_1 = a_0U + a_3X_3 + a_4X_4 + E_6$

where: a_0, a_3, a_4 = partial regression weights

Y_1, X_3, X_4 are as defined in Model 4

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

E_6 = error in prediction

Both of these models (as well as Model 4) can be statistically tested against Model 3. If statistical significance results, Model 3 cannot be utilized and must be discarded in favor of the model which has generated significance. When this occurs, then the researcher must decide which department (1 or 2) the third department looks <u>most</u> like. For example, suppose that the two departments had dissimilar slopes and the predicted scores were to be obtained from Eq. 1:

Eq 1 $\hat{Y}_1 = a_0U + a_3X_3 + a_4X_4$

Assuming that the empirical weights were determined, then Eq. 1 might be as follows:

Eq 2 $\hat{Y}_1 = 6U + 3.2X_3 + 4.1X_4$

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

in the necessary usage of the following model:

Model 5 $Y_1 = a_1X_1 + a_2X_2 + a_5X_5 + E_5$

where: a_1, a_2, a_5 = partial regression weights

Y_1, X_1, X_2 are defined as in Model 4

$X_5 = X_3 + X_4$ (from Model 4)

E_5 = error in prediction

It may empirically turn out that the two slopes are not similar, even though the initial starting points are the same:

Model 6 $Y_1 = a_0U + a_3X_3 + a_4X_4 + E_6$

where: a_0, a_3, a_4 = partial regression weights

Y_1, X_3, X_4 are as defined in Model 4

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

E_6 = error in prediction

Both of these models (as well as Model 4) can be statistically tested against Model 3. If statistical significance results, Model 3 cannot be utilized and must be discarded in favor of the model which has generated significance. When this occurs, then the researcher must decide which department (1 or 2) the third department looks <u>most</u> like. For example, suppose that the two departments had dissimilar slopes and the predicted scores were to be obtained from Eq. 1:

Eq 1 $\hat{Y}_1 = a_0U + a_3X_3 + a_4X_4$

Assuming that the empirical weights were determined, then Eq. 1 might be as follows:

Eq 2 $\hat{Y}_1 = 6U + 3.2X_3 + 4.1X_4$

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

Eq 3 $$\hat{Y}_1 = 6 + 3.2X_3$$

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

Eq 4 $$\hat{Y}_1 = 6 + 4.1X_4$$

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.

We take this opportunity to remind researchers not to 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 (See Kelly, Beggs, McNeil, Eichelberger, and Lyon (1)).

It is conceivable that a researcher would begin with, say, six different schools, and develop groupings of schools to find out which schools were alike in their functional relationship. For example:

Model 7 $$Y_1 = a_1U_1 + a_2U_2 + a_3U_3 + a_4U_4 + a_5U_5 + a_6U_6 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + E_7$$

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

Through various restrictions that were empirically reasonable, schools 1, 5, and 6 might be found to have the same slope. Schools 1, 5 and 6 all have the same intercept as schools 2 and 3 (who themselves have a common slope, although different from schools 1, 5 and 6). School 4 was found to be unique in both intercept and slope. The resulting prediction equation would have been:

$$\text{Eq 5} \qquad \hat{Y}_4 = 6(U_1 + U_2 + U_3 + U_5 + U_6) + 8\,U_4 + 3.1(X_1 + X_5 + X_6) + 3.4(X_2 + X_3) + 2.7\,X_4$$

This process has led to no increase in predictability for school 4, for it was found to be different from all the others. More stable regression weights can be determined for the other schools, though, because of the sharing of data between schools which are empirically similar. In this way, more stable predictions can be made for schools which have very little data available. One of the selling points of Bayesian statistics is its capitalization on small amounts of data. If the appropriate regression models are utilized, the multiple regression approach can also capitalize on small amounts of data.

In all of the above models, the testing of interaction is what distinguished the regression procedure outlined here from that commonly used in prediction studies. Too often, even plausible interactions are ignored and all subjects are lumped together and, hence, treated as similar. Our conceptual theories have long ago turned to distinct groupings, and it is about time that our statistical procedures reflect this empirical possibility, whether they be Bayesian or multiple linear regression. Until the Bayesian methodology has empirically been shown to be more predictive than multiple regression analysis, the availability and relative mathematical simplicity of multiple regression analysis would seem to indicate preference for its utilization rather than the Bayesian approach. However, we encourage Bayesian statisticians to present their methodologies in a more simplified and straightforward manner than has been done to date (Meyer and Collier, (2), Schmidt, (5)). More applications such as the one by Pitz (4) would be helpful to the research community.

REFERENCES

1. Kelly, F.J.; Beggs, D.L.; McNeil, K.A.; Eichelberger, T.; and Lyon, J. Design in the Behavioral Sciences: Multiple Regression Approach. Carbondale: Southern Illinois University Press, 1969.

2. Meyer, D.L. and Collier, R.O. (Ed.) Bayesian Statistics. Itasca, Ill.: F.E. Peacock Publishers, Inc., 1970.

3. Novick, M.R. and Jackson, P.H. Bayesian Guidance Technology. Review of Educational Research, 1970, 40, 459-494.

4. Pitz, G.F. An example of Bayesian hypothesis testing: The perception of rotary motion in depth. Psychological Bulletin, 1968, 70, 252-255.

5. Schmidt, S.A. Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics. Reading, Mass.: Addison-Wesley Publishing Co., 1969.