MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 17, NUMBER 2, FALL 1990

The Case Against Interpreting Regression Weights

Keith McNeil
University of New Mexico

Abstract

One of the major problems that has occurred in the use of the regression statistical procedure, is the tendency of individuals inappropriately interpreting regression weights. The purpose of this paper is to discuss and to clarify problems that can arise from such interpretation.

Introduction

Although most multiple regression texts argue against interpreting regression weights: ("shaky and dangerous") (Kerlinger and Pedhazer, 1973); "not very clear how these values are useful" (Ward and Jennings, 1973); "acquire more meaning than statistically appropriate" (McNeil, Kelly and McNeil, 1975)), some statistics text authors and researchers still want to place some sort of importance or meaning on the magnitude or relative magnitude of the regression weights. The purpose of this paper is to provide various reasons for why such interpretations are not appropriate. Two cases will be discussed in which the interpretations do not have to do with "importance."

Reasons for not interpreting regression weights include: 1) degree of predictability in the population is less than perfect, 2) regression weights fluctuate from sample to sample, 3) assignment of weight is arbitrary, 4) regression weights would probably be different in a manipulated situation as compared to a non-manipulated situation, 5) the purpose of the test of significance is unrelated to interpretation of weights, and 6) the purpose of using multiple predictors.

Orthogonal Predictors

In the situation where the predictor set is orthogonal, regression weights are indeed estimates of the population means. A subsequent sample would probably produce a different set of weights, but each set is an unbiased estimate of the population means. But in no case would one want to rank the regression weights to "find the most important variable." The variable with the highest regression weight has the highest sample mean but that highest mean doesn't make it "the most important."

Non-Orthogonal Predictors

R^2=1.0. If the R^2 is 1.00. in the population then the weights would be stable from sample to sample because there would be no sampling error. Newton's law of gravity $D = 1/2\ GT^2$ was shown to be derivable from regression technology (McNeil, 1970). But what does the weight's coefficient of 1/2 mean? Similarly, Circumference = Pi * Diameter, but what does Pi mean? Pi is simply the weight, which, when multiplied times the diameter, yields the circumference.

R^2 less than 1.0. When the R^2 is less than 1.0, successive samples from the same population, especially with correlated predictors, will yield quite different regression weights. Since these weights bounce around, the term "bouncing betas" has been coined (Kerlinger and Pedhazur, 1973). Furthermore, when attempting to increase R^2 or a particular sample, the addition of non-orthogonal (correleated) predictors will change the magnitude of the regression weights. When the population's functional relationship has been mapped the weights will be stable. Even when correlated predictors are used, weights may be stabilized even then.

An extreme case of perfectly correlated predictors.

One cannot use weights to assess the "importance of a variable", because when predictor variables are correlated both variables do not "get the weight" equally. In the extreme case when two variables are perfectly correlated, one would "get the weight" and the other would get a weight of zero. Certainly one would not want to attach "no importance" to the variable that got a weight of zero. It is the case that this variable does not provide any new information over and above the perfectly correlated variable, but the luck of the draw assigned the weight to the other variable.

Control, or Upsetting the Prediction

These applications where once a high R^2 is obtained that the goal then becomes one of "upsetting the prediction" (for example attendance predicting GPA). One tends to manipulate one or more predictor variables in an attempt to alter prediction.

But one must remember that until manipulation has occurred, one cannot know for certain the effect of such manipulation. Once variables are manipulated, other, correlated or uncorrelated, variables may have a different effect on the criterion. The magnitude of the beta weights do not give any clue as to what may happen. Some predictors will be more amenable to manipulation and some manipulated variables will have no differential effect on the criterion. Finally, manipulating one predictor will certainly have some possibly unknown effects on some of the other predictors.

Interpretation of Statistical Tests

When one tests a regression weight, one is usually testing the restriction that the weight is equal to zero. If significance is determined, then one can reject the null hypothesis weight ($a_i = 0$) and accept the research hypothesis that weight $a_i = 0$) (non-directional) or weight $a_i = 0$ or weight $a_i < 0$ (directional). In neither case is the conclusion "the regression weight is the sample value, say 1.34."

The virtue of testing non-zero restrictions such as weight $a_i = 1.34$ has been delineated (McNeil, in preparation). But if significance is found with this test, then one can only conclude that, say $a_i > 1.34$. If significance is not obtained, one cannot conclude that $a_i = 1.34$, but that we fail to reject the hypothesis that $a_i = 1.34$. We not only cannot interpret the weight, but we don't know the exact value of the population weight. (When R^2 equals 1.00 we may "know" the weight.)

Purpose of Using Multiple Predictors

The most compelling argument against the interpretation of regression weights is that when one utilizes MLR one is taking the stance that behavior is complexly determined (complex in terms of a large number of predictor variables). The goal then is to account for the variation in the criterion by obtaining as high an R^2 as possible by that set of predictors. To try to isolate the "most important variable" in that set is not related to the goal of maximizing the R^2 which is what MLR produces.

The Inverted U Example

Suppose data were obtained as in Figure 1, where there is a systematic second degree function between X and Y. The linear correleations are: $r_{xy} = .00$, $r_{xy} = .27$, $r_{x2x} = .96$ when both X and X^2 are used in a multiple regression model, the resulting R^2 is 1.00, and the function of best fit is $Y = 5 * U - 12 * x + 5 * X^2$. In no way is X^2 "more important" than X. It takes the unit vector, X and X^2 to account for the variation in Y. Each variable, X, U, and X^2, contributes "over and above" the other two variables.

Although the variable X illustrates the typical "suppressor variable", (correlating 0.0 with Y, correlating high with the other predictor, and having a negative weight) the fact remains that X is as necessary in the equation as X^2. Yet, the beta weight are similar, but opposite in sign!

The following Appendix A is presented for the purpose of identifying a sample of a large number of authors who have made statements related to problems and concerns with the interpretation of regression weights and prominent authors who actually interpreted beta weights. Let's hope that these examples will increase the sensitivity of individuals who read the interpretation of regression analysis results.

Appendix A

1) Draper and Smith (1981) p 117

If multiple samples of the same variable are obtained, b is an unbiased estimate of the population b only if the postulated model is the correct model (i.e. $R^2 = 1.00$). If it is not the correct model, then the estimates are biased. The extent of the bias depends... not only on the postulated and true models, but also on the values of the X variables...

2) Cooley and Lohnes (1962) p 40

"The beta weights... indicate that... is the most useful in the battery, followed by... and...

3) Williams (1959) p 31-32.

The significance tested is actually that of the additional amount of variation (in the criterion) accounted for by the (predictor) variable... above that accounted for by the remaining variables.

4) Ward and Jennings (1973) pg 271.

Some questions, however, that arise in natural language form almost defy translation. Examples are the questions:

1. Which predictor variable is the most important in explaining the criteria?

2. What are the relative contributions of the various predictors to the prediction of the criterion?

"articles by Darlington (1968) and Ward (1969) do describe ways of calculating values to reflect answers to these questions. Although it is usually not very clear exactly how these values are useful..."

5) Kerlinger and Pedhazur (1973) pg 63.

"The relative sizes of the b and beta weights seem to indicate that... and... contribute about equally, and that... contributes little, but such interpretations are shaky and dangerous..." pg 77.

Another difficulty is the instability of regression coefficients. When a variable is added to a regression equation, all the regression coefficients may change from sample to sample as a result of sampling fluctuations, especially when the independent variables are highly correlated, (Darlington, 1968). All this means, of course, that substantive interpretations of regression coefficients is difficult and dangerous, and it becomes more difficult and dangerous as predictors are more highly correlated with each other.

Bibliography

Cooley, W. W., & Lohnes, P. R. (1962). Multivariate procedures the behavioral science, New York: John Wiley and Sons.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Draper, N., & Smith, H. (1966). Applied regression analysis, second edition. New York: John Wiley and Sons.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Introduction to linear models. New York: Holt.

McNeil, K. A. (1970). Meeting the goals of research with multiple regression analysis. Multivariate Behavioral Research, 5, 375-386.

McNeil, K. A. The case for non-zero restrictions in statistical analyses. Paper in preparation.

McNeil, K. A., Kelly, F. J., & McNeil, J. T. (1975). Testing research hypotheses using multiple linear regression. Carbondale: Southern Illinois University Press.

Ward, J. H., & Jennings, E. J. (1973). Introduction to linear models. Englewood Cliffs: Prentice Hall.

Williams, E. J. (1959). Regression Analysis. New York: John Wiley and Sons.

Williams, J. D. (1974). Regression analysis in educational research. New York: MSS information Corporation.