MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 19, NUMBER 1, SUMMER 1992

# Response to Lunneborg: The Conditions for Interpretation of Regression Weights

Keith McNeil
New Mexico State University

ABSTRACT

In a reply to McNeil (1990), Lunneborg (1991) indicates his strong desire to interpret regression weights. While taking such a stand, he hints at several conditions, but does not explore them deeply. Unfortunately, these conditions are seldom obtained in applications of the General Linear Model. Although these conditions can be obtained, most researchers do not obtain them and are often too impatient to restrain their interpretations. These conditions are an $R^2$ close to 1.0 and predictor variables that have been manipulated.

## Deterministic behavior

Lunneborg concludes that "not all behavior in the behavioral sciences is deterministic," basing his conclusion on two senses. "First, there is the possibility of some inherent randomness," and second, "in practice we shall never identify all of the EVs needed to account fully for the variability in the" (criterion). While one may disagree with his conclusion, one could agree with his two senses. Indeed, if one uses regression weights to predict behavior then one is acting upon a deterministic model. For instance, if one reads a journal on regression, one is intending to learn more about regression.

The reason that the two senses might disturb one is that too many researchers take these as a rationale for conducting sloppy research, for using only a few variables, and for not considering any other relationship other than linear relationships. One must start off with the assumption in the behavioral sciences that behavior is complexly determined (caused, occurring, or whatever synonym that you choose) and therefore one must include enough (which may be many) variables in the regression model.

## Manipulation of predictor variables

Most regression applications are really in the data snooping category, attempting to find out what is happening. In the example that Lunneborg provides, predicting Annual Income from Years of Education, the regression weight for Years of Education is correctly interpreted as "The increase in expected Annual Income associated with an increase by one year in the number of Years of Education." Now this interpretation is valid with the static sample of data at hand. The data is static in the sense that the data was collected ad hoc and there was no attempt at random assignment to various Years of Education. That is, there was no manipulation of Years of Education. Now consider the case when the researcher decides to manipulate the predictor variable. (It is not clear that any researcher or any subjects would be willing to do such a study, but let us assume that there are such individuals.) Is it reasonable to assume that all subjects will react the same way to receiving their allotted Years of Education? Will not some subjects attempt to override the allotment, by requesting more years of education, while other subjects might even request fewer years of education? And even if the subjects were controlled enough to take the right amount of assigned Years of Education, is it reasonable to assume that these Years would have the same effect on the criterion that was observed in the non-manipulated situation? All of the internal and external validity issues discussed many years ago by Campbell and Stanley (1963) are still alive today. Only until the predictor variable(s) are manipulated will one be able to use the regression weights to make accurate "manipulation" predictions.

$R^2$ close to 1.00

If a researcher has not obtained an $R^2$ close to 1.00, then interpretation of regression weights can lead to very uninformative and in some cases totally false predictions. Lunneborg contends that "we often make scientific headway by considering, at one time, only a few of the many EVs which are known to be relevant....If I sample randomly I need not worry overmuch about what else I might have put into my model" (Lunneborg, 1991).

Figure 1 indicates (totally fictitious) data that directly contradicts the above thinking. The regression weight from the single straight line model is accurate in predicting the sample's Annual Income over the lower range of Years of Education, but not so at the upper ends. Indeed, the interaction between Gender and Years of Education nullifies the use of the regression weight from the single straight line model even in the static case of the sample data. The single regression line of best fit from the single straight line model is not applicable to either males or females, and indeed would lead to erroneous recommendations for females. That is, the single line of best fit would recommend additional Years of Education for both males and females, but the two interacting second degree curves recommends a plateau at about 11 Years of Education and no additional Years of Education after that.

Careful sampling to obtain as many males as females would not in any way alleviate the misinterpretation provided by the single straight line regression weight. Lest the reader argue that the data is "unusual," another example is provided. Many functional relationships are of a second-degree nature, either inverted U-shaped or U-shaped as in Figure 2.

If a researcher took the usual "easy way out," only the single line model of using X to predict Y would be investigated. Upon finding that the slope of the line is close to 0 and that the $R^2$ is close to 0, the researcher would conclude that there is no (linear) model is of no value in the prediction of Y at any point along the X axis. Most researchers would likely not again use X in the prediction of Y. Obviously, the data depicted in Figure 2 would

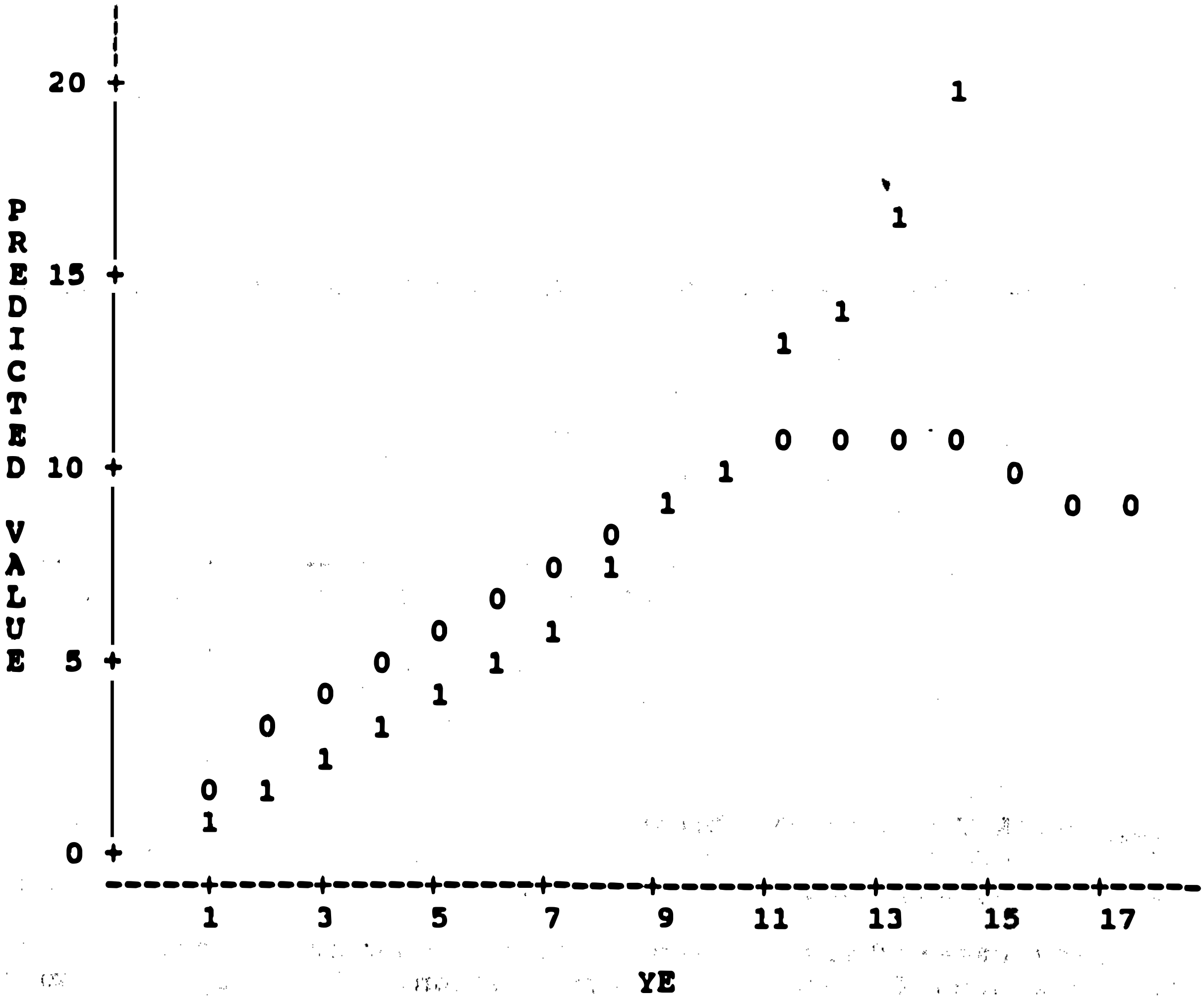


**Figure 1.** A case wherein linear prediction is high, but misleading for values above 11.

relationship between X and Y. Indeed, the single straight line lead one to investigate the second-degree relationship between X and Y.

The data in Figures 1 and 2 present cases wherein false predictions are made when the $R^2$ is less than 1.00. Too much research is limited to the Pearson product and the t-test single variable thinking. Even interaction is usually not investigated as a valuable component for increasing the $R^2$, but as problematic in the interpretation of the Main Effects.

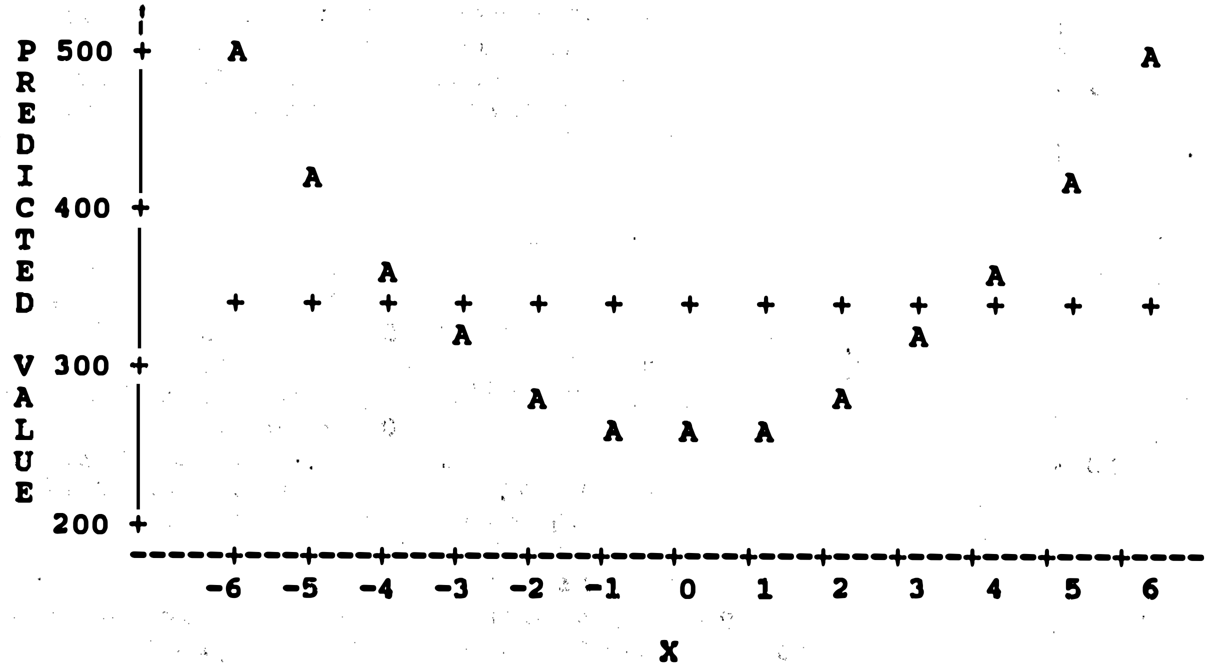


*Figure 2.* A U-shaped relatinship.

Interpreting Regression Weights in Terms of Relative Importance

Many researchers and some statistics authors provide support and procedures for such an interpretation. The original impetus for the McNeil (1990) paper was the concern with interpreting regression weights in terms of *relative importance*. The paper did not address this issue, and hence neither did Lunneborg. If one takes the multivariate stance, then one is stipulating that behavior is multiply determined, and that many predictor variables may need to appear in the regression model. All of the predictor variables are important, and the various predictor variables are almost certainly correlated with each other to some extent. Therefor, all of the "necessary" predictor variables are valuable in the determination of the criterion variable. Again, Figure 2 is a good example of the inadvisability of identifying the "most important" predictor. Which predictor has the highest weight will be a function of the data. It might well be that the linear term has the highest weight, yet we know that the linear term, by itself, has absolutely no relationship with the criterion. The tone of Lunneborg's comments would lead one to believe that he, too, would disavow interpretation of "the most important" predictor variable.

Lunneborg's caveat of conditional interpretation, although more appropriate than a non-conditioned interpretation, is inappropriate for a relationship that has an $R^2$ less than 1.00 as well as for a static relationship. Figures 1 and 2 again provide evidence to support this assertion. Since the predictor variables are correlated one cannot assume that one predictor variable can be held constant when another predictor variable is changed.

### Testing Non-zero Weights

Lunneborg concerns himself with the proper interpretation of the significance of the hypothesis test. He says that he is more interested in the interpretation of weights than in whether or not there is statistical significance. There is a way to accomplish his goal through hypothesis testing. If one is interested in making statements about the magnitude of the weighting coefficients, then one should be testing non-zero statistical hypotheses (McNeil, 1991). The testing of a weight equal to zero has become so automatic and common-place that often researchers fail to consider other alternatives. As Lunneborg states, the Research Hypothesis should guide the models tested, and that Research Hypothesis is guided by what the researcher wants to conclude from the research. If one is not going to be satisfied to conclude that "the regression weight is not zero," then one should be testing another Research Hypothesis. If one is not going to be satisfied to conclude that "the regression weight is not zero," but wants to conclude that "the regression weight is greater than zero," then that Directional Hypothesis is the Research Hypothesis that should be tested. If one is not going to be satisfied to conclude that "the regression weight is greater than zero," but wants to conclude that "the regression weight is, say greater than 500," then that Non-zero weight is the Research Hypothesis that should be tested.

### Stages of Modeling

Lunneborg (1991) refers to stages of modeling. "Interpretation comes in after final, or at least, promising, models have been identified". In other sections of the paper he talks about "accepted" or final models. And in another section he agrees that an observational study may be a poor guide to what happens when we attempt to manipulate. One could conclude that he would like to refrain from interpreting a weighting coefficient until he has obtained a model that has a high $R^2$ and that has been validated on manipulated data. But such a definite conclusion does not appear in the article. Any researcher should be aware of the stage of modeling that they are in, and since so many researchers jump from one content area to another, most should rightly find themselves in the very lowest stage. In an early regression text (McNeil, Kelly, & McNeil, 1975, p. 474), an argument was made for the relative value of probability and $R^2$ depending upon the stage of the research. Five stages were identified: 1) data snooping, 2) hypothesis testing, 3) replication, 4) manipulation with dynamic variables, and 5) replication with dynamic variables. An emphasis on low probability was seen as valuable in stages 2, 4 and 5, whereas an emphasis on high $R^2$ was seen as valuable in stages 1, 3, 4, and 5. The addition to those notions in light of the above discussion is that the emphasis on interpreting weights would be valuable only at stage 5 when one had successfully replicated, at a high $R^2$ value, manipulated data.

Note. I would like to acknowledge the assistance of Fred Lillibridge who provided comments on the rough draft.

## References


Campbell, D. T., & Stanley, J. C. (1963). Experimental and Quasi-Experimental Designs for Research on Teaching. In N. L. Gage (Ed.), Handbook of Research on Teaching (pp. 171-246). Chicago: Rand McNally.

Lunneborg, C. E. (1991). A case for interpreting regression weights. Multiple Linear Regression Viewpoints, 17(3).

McNeil, K. (1990). The case against interpreting regression weights. Multiple Linear Regression Viewpoints, 17(2), 40-47.

McNeil, K. (1991). The case for making non-zero restrictions in statistical analysis. Multiple Linear Regression Viewpoints, 17(3).

McNeil, K., Kelly, F., & McNeil, J. (1975). *Testing Research Hypotheses Using Multiple Linear Regression*. Carbondale, Il.: Southern Illinois University Press.