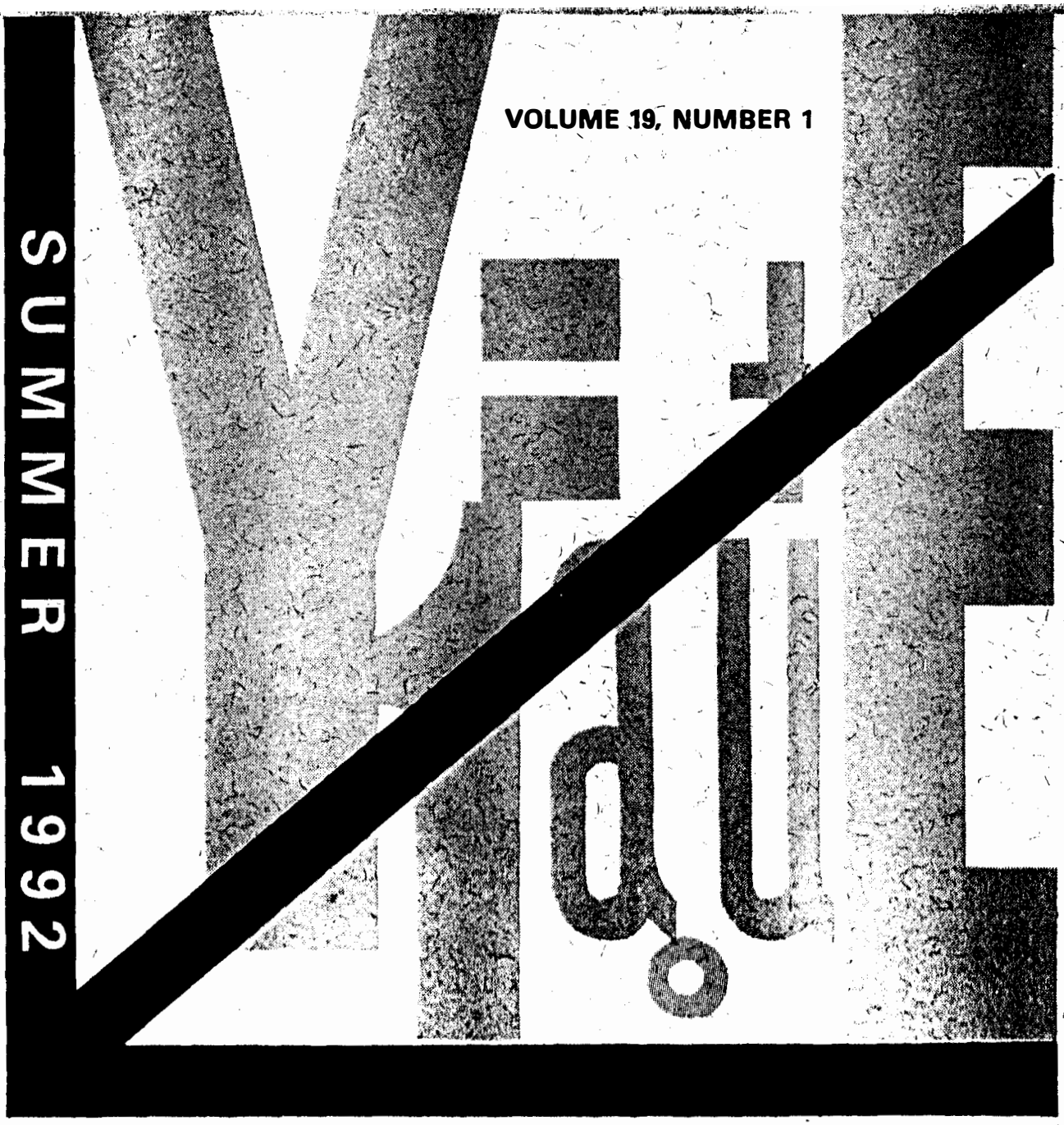


# MULTIPLE LINEAR REGRESSION VIEWPOINTS

**A publication of the Special Interest Group on Multiple Linear Regression**

**MLRV Abstracts appear in CIJE, the ERIC System, and microform copies are available from University Microfilms International**
**MLRV is listed in EBSCO Librarians Handbook.**

**ISSN 0195-7171**

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

# MULTIPLE LINEAR REGRESSION VIEWPOINTS



## EDITORIAL BOARD

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

# The Beta or Not the Beta; What is the Research Question?

**Rio Brown**
**California State University, Fresno**

## Abstract

The present article discusses the interpretability of beta weights in terms of their definition, technical aspects and the research philosophy guiding the use of multiple regression. The major conclusion is that variable importance and variable ordering can not be ascertained by examining beta weights. Additionally, it is recommended that discussion of variables as a group without identification of singular variable importance would more appropriately match the multivariate purpose of multiple regression.

## Introduction

Cohen and Cohen, (1975, p. 79) say that one important problem in multiple linear regression is not straight forward - that of defining the contribution of each independent variable. They suggest that substantive reasoning and precise formulation of the research question are critical in utilization of statistical methodology. Discussion of the inability to interpret a beta weight in terms of identifying best or most important variable(s) in a regression equation seems to center on 3 issues: their definition with respect to purpose in multiple regression, their stability as a parameter estimate, and an understanding of the research question posed when using multiple regression (Brown and Tracz, 1990). The first two issues will only be briefly highlighted since they are generally well covered in major textbooks on the topic.

---

An earlier version of this paper was presented at the 1991 American Educational Research Conference in Chicago, Illinois. The author thanks Dr. Isadore Newman for his helpful suggestions.

Definition

Beta is the partial regression coefficient when all variables are standardized. Its square is the proportion of variance shared with the dependent variable that is independent of the remaining independent variables (Cohen and Cohen, p. 92). (Thorndike 1978, p. 152) presents the equation:

(ß = Beta)

$$S^2_{zy} = \beta^2_{x_1} + \beta^2_{x_2} + 2\beta_{x_1}\beta_{x_2}r_{x_1x_2}$$

showing the variance predicted in standard score form and noting that the squared beta weights reflect the relative importance of the independent variables, pointing out that the $ß^2$ is not a proportion of variance, but relative contribution. However, Cohen and Cohen (1975, p. 95) show a similar formula:

(ß = beta)

$$R^2 = \Sigma\beta^2_i + 2\Sigma\beta_i\beta_j r_{ij}$$

saying that the above formula and its variations only appear to partition portions accounted for uniquely noting that any $ß_i$ and $r_{yi}$ may be of opposite sign (suppression) and that $ß_i\ ß_j\ r_{ij}$ may be negative precluding use of this equation as a variance partitioning procedure.

Edwards (1984, p. 107) says that if the test of a regression coefficient for a given variable is significant, then that variable when entered last in a regression would result in a significant increase in the regression sum of squares. A variation of that definition by Edwards suggests that if all other independent variables are held constant except X, the b (unstandardized) is the amount that the dependent variable increases with each unit of the independent variable.

Pedhazur (1982, p. 63) notes that testing a given beta weight is like testing incremental changes in $R^2$ for a given independent variable. Similarly, Huberty (1989) notes that the difference of incremental squared multiple correlations is precisely the square of the semi-partial correlation between the criterion and any predictor with the remaining predictors partialled. He states that it is clear from this relationship that a variable ordering cannot not be accomplished via the beta values.

## Technical

The instability of beta weights (bouncing betas as they are often called) is well documented. Stevens (1986, p. 98) indicates that the desirable property of least squares regression is the unbiased, minimum variance estimator of the population beta that will not be consistently high or low but will bounce above or below.

The test for the beta asks if it is different from zero while controlling for the effects of other variables (Pedhazer, 1986, p. 59), but because the denominator in the test reflects other variables, the higher the intercorrelations, the larger the standard error. Situations exist where a significant $R^2$ exists with no significant betas or a non-significant beta for a given variable, but a significant correlation between the variable and the dependent variable. Huberty (1989) notes that use of the squared values of the standardized regression coefficients to assess variable importance is generally eschewed by methodologists due to the unreliable effects of multicollinearity.

Huberty also notes that sample specificity is a major issue in beta interpretation and that although a large ratio of sample size to response variables is preferred, such does not ensure valid generalizations. For path analysis, Pedhazur (1982, p. 628) warns the coefficients are sample specific and cannot be used for comparisons or generalizations across populations.

Pedhazur (1982, p. 247) says that it is the scale free property of the beta that leads researchers to treat them as an indicator of the relative importance of the variables for which they are associated. However, the magnitude of the beta reflects not only the presumed effect of the variable in question, but also the variance/covariance of the other variables in the model.

## Research Philosophy

The research question being asked in multiple regression and the singular importance of variables presents an incongruity. Huberty (1989) says the idea of relative variable importance in a multivariate context is not clear and that there is little consensus of the meaning of relative variable importance existing among social and behavioral science methodologists. He goes on to say that the fundamental reason for conducting multivariate analysis is the study of a system of variables rather than univariate relations. For research, Huberty states that variable importance depends on the collection of variables studied, including all relevant variables while excluding irrelevant ones and that interdependence among variables makes the concept of variable importance very questionable and fruitless. Darlington (1968) echoes that thought in stating that independent contribution to variance makes little sense when variables are intercorrelated. Edwards (1984) finds no satisfactory method for determining the relative

contribution of independent variables to the regression sum of squares when intercorrelation exists. Stevens (1986, p. 99) indicates that discussing the unique contribution of a given independent variable is generally meaningless if the predictors are correlated. Multicollinearity is a problem that makes the importance of a given predictor difficult because of confounded effects among the variables.

The concept of control of variables enters the discussion of research design since partialling is considered in the definition of a beta weight. But, Pedhazur (1982) argues that controlling variables only has meaning if grounded in theory. With little theoretical consideration among the pattern of variables, controlling the variance of one variable to examine the effects of other variables may amount to distortion of reality and misleading results. He refers to the concept of studying the effect of one variable on another by holding one constant via regression analysis as an "air of fantasy" (p. 225). In experimental research, Pedhazur notes that if independent variables can be manipulated and control of extraneous variables is reasonably done, then conclusions of the direct effects of one variable on another can be made. In regression, the equations reflect the average relations between a dependent and independent variable and not necessarily the process by which the independent variable effects the dependent variable. He points to an example from the Coleman study that having versus not having a language lab in a school may be different from removing a lab from a school. To draw a similar type of conclusion from regression research based on a beta weight interpretation must be done with a much care. He notes that it has been argued that to find out what happens to a system when you interfere with it, you must interfere with it.

Summary

None of the major texts or papers reviewed suggested the use of beta weights for purposes of identifying the most important variables. In fact, there were little or no suggestions for interpreting beta weights at all for reasons of definition, instability, sample specificity, specification errors, and most importantly, the incongruity of the general purpose of multiple correlation and the singling out of individual variables.

It appears that interpretation has come about among some researchers by paralleling experimental designs' congruity with the ANOVA in the context of multiple regression. That is, if factorial design (with ANOVA, as a statistical tool), can isolate independent contribution to explaining variance in the outcome variable (even if independence is forced by equalizing cell sizes) then attempts are made to apply the same "logic" to correlational design and multiple regression. Typically, one reads a research statement such as, "the focus of this study was to see if variable Y (dependent) can be explained by a combination of variables X1, X2, . . . (independent)." However, after analysis, the discussion usually includes

statements such as ". . . the overall $R^2$ was .xx with X2 being the best predictor and X1 not being important because of its small beta weight." No caveats are expressed and very often the relationships among the variables are ignored.

The question now becomes what can be said with respect to beta weights following the completion of a multiple correlational analysis. For interpretive purposes, Pedhazur (1982, p. 247) suggests reporting the beta, the b weight and the standard deviation of all variables with discussion of issues that may be a factor. Huberty recommends data exploration including variable screening before inclusion in a model and cross validation. The discussion of the variables as a set with no speculation of univocal importance would more appropriately follow the multivariate purpose of multiple regression. Of course, specifying models to be tested based on theory to untangle complex relationships is preferred.

Pedhazur (1982, p. 65) points to the frustration of trying to identify the relative importance of variables since there is more than one answer to the question and the ambiguity of some problems is not entirely able to be solved. He notes that beta weights have "great appeal because they hold the promise for unraveling complex phenomena" (p. 221), but they are unstable and require many conditions for interpretation. He goes on to say that the absence of a model precludes any meaningful interpretation of coefficients. "No amount of fancy statistical acrobatics will undo the harm that may result using an ill conceived theory" (p. 230).

## References


Brown, R. & Tracz, S. (1990). Considerations and issues in variable selection and interpretation in multiple regression. Paper presented at the meeting of the American Educational Research Association, Boston.

Cohen, J. & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, New Jersey: John Wiley and Sons.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69(3), 161-182.

Edwards, A. L. (1984). An introduction to linear regression and correlation. (2nd Ed.). New York: W. H. Freeman and Company.

Huberty, C. J. (1989). Problems with stepwise methods -- better alternatives. Advances in Social Science Methodology, 1, 43 - 70.

Pedhazur, E. J. (1982). Multiple regression in behavioral research. (2nd Ed.) New York: Holt, Rinehart and Winston.

Stevens, J. (1986). Applied multivariate statistics for the social sciences. New Jersey: Lawrence Erlbaum Associates.

Thorndike, R. M. (1978). Correlational procedures for research. New York: Gardner Press, Inc.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 18, NUMBER 1, SUMMER 1992

# The Interpretation of the Beta Weights in Path Analysis

Susan M. Tracz
California State University, Fresno

A paper submitted to Multiple Linear Regression Viewpoints, April, 1991.

Path analysis is a method for determining "the direct and indirect effects of variables taken as causes of variables taken as effects" (Pedhazur, 1982, p. 580). Researchers who use path analysis attempt to arrive at models, often called causal models, showing the relationships between exogenous variables, those with variability explained by causes outside the model, and endogenous variables, those whose variability is explained by some constellation of exogenous and/or other endogenous variables in the model. Regosa (1987) calls path analysis "simple multiple regression with pictures" (p. 186).

## Causality

It is worthy of note that there is a heated debate concerning what actually constitutes causality. The consensus is that three criteria must be met:

1) a temporal sequence of variables (X precedes Y),
2) an association or relatedness among variables ($r_{xy}>0$), and
3) control ($X \rightarrow Y$).

While some authors (Biddle & Marlin, 1987; Kenny, 1979) believe that causal relationships can be established with regression and other related techniques, others believe such conclusions are unwarranted (Freedman, 1987; Regosa, 1987) and

are the result of faulty logic (Games, 1990). To underscore the fervor researchers exhibit on this issue, Ling (1982) in a review of a book entitled Correlation and Causation (Kenney, 1979) writes, "the serious limitations of this book lie not in its lack of mathematical rigor, but in its faulty logic as well as its faulty presentation and interpretation of certain statistical methodology.... I feel obligated to register my strongest protest against the type of malpractice fostered and promoted by the title and content of this book" (p. 491).

Despite the often repeated admonition that correlation does not imply causation (Games, 1990; Pedhazur, 1982), the literature is filled with examples of interpretations and conclusions erroneously made more broadly than was appropriate. As Hayduk (1987) noted, "causation may not be in the real world or in the equations, but it is definitely in our thinking" (p. XV).

Control

As a criterion in the definition of causality, control means that variation in Y is the direct result of X. Biddle and Marlin (1987) say that it is possible to control statistically for possible confounding effects of variables using partial correlations. Games (1990), on the other hand, believes that random assignment of subjects to groups provides control. He emphasizes that, "the experiment provides control; the correlation study does not" (p. 244). Pedhazur (1982) agrees with Games saying, "one of the most powerful methods of control is randomization. Being in a position to manipulate and randomize, the experimenter may feel reasonably confident in

making statements about the kinds of action that need to be taken in order to produce desired changes in the dependent variables" (p. 578).

Thus, there is a distinction drawn between experimental research and correlational research. In the former the independent variables can be manipulated so that instead of simply observing what occurs, researchers can effect change. In correlational research, this is not the case. This distinction has important implications for policy makers. While there are numerous examples of the mistaken belief that manipulating independent variables in correlational studies will change outcomes, the classic example is the Coleman Report. On the basis of correlational information the Coleman Report concluded that "if a minority pupil from a home without much educational strength is put with schoolmates with strong educational backgrounds, his achievement is likely to increase" (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld & York, 1966, p. 22). Many large scale busing programs were initiated on the basis of the Coleman Report, but increases in minority students' achievement never materialized.

Further, the widespread belief that a model is "'confirmed' if the correlations in the matrix correspond to those we would have predicted from our model" (Biddle & Marlin, 1987, p. 5) does not mean that there is proof for that model. "Consistency of the model with the data, however, does not constitute proof of a theory; at best it only lends support to it" (Pedhazur, 1982, p. 579).

## Path Analysis

Numerous authors (Cliff, 1983; Freedman, 1987; Mulaik, 1987; Regosa, 1987) complain that path analytic techniques are often misused and that this misuse is fostered by the availability of computer programs. To further complicate the issue and to underscore why causal modeling is unlikely to determine actual causes, it is possible that "very different causal structures may fit the same set of data equally well" (Stelzl, 1986, p. 309).

Misuse of a technique, however, does not mean that the technique is inappropriate, invalid or incorrect. Mulaik (1987), who states that "the rule of a causal connection is that of functional relation" (p. 23), also argues that the "concept of causality may be modified to have causes determine not specific outcomes but the probabilities of outcomes" (p. 18).

## Assumptions

In path analysis, the variables are generally expressed as standard scores, and the equation for an endogenous variable is formed by weighting each endogenous and exogenous variable presumed to have a causal effect and summing all these terms plus error. These weights are the path coefficients, and these equations are regression equations. A path analysis arrives at one or more regression equations. In addition, certain assumptions are made when performing such an analysis. A potent criticism of the use of path analysis, however, is that the assumptions required for this technique are frequently not met (Freedman, 1987). Pedhazur (1982) lists the assumptions for nonrecursive models as follows:

1. The relationships among the variables in the model, are linear, additive, and causal.

2. Each residual is not correlated with the variables that precede it in the model.

3. There is a one-way causal flow in the system. That is, reciprocal causation between variables is ruled out.

4. The variables are measured in an interval scale.

5. The variables are measured without error. (p. 582)

Under these assumptions, the path coefficients are the ordinary least squares, regression coefficients. The assumptions have been stated by other authors (Biddle & Marlin, 1987; Freedman, 1987), who also note that newer techniques such as LISREL have all the assumptions of regression plus additional assumptions. These assumptions are seldom tested and would rarely hold if they were tested.

Interpretation of Weights

Another criticism of path analysis is that the weights are not interpreted correctly. Despite the innovations and increasing sophistication of path analysis, including the use of LISREL and hierarchical modeling with their additional assumptions, path analyses generally use regression models for which beta weights are reported. Beta weights as scale-free indices reflecting the increase or decrease in the dependent variable with a unit increase in the independent variable allow for comparisons across variables of different metrics. The magnitude of the beta is a function of the correlation between the independent and dependent variable, the model's variance covariance matrix, and the error term which includes the variances of variables not included in the model. For these

reasons, beta weights are highly unstable from sample to sample (Freedman, 1987; Pedhazur, 1982). All the caveats regarding the interpretation of beta weights that apply to multiple regression also apply to path analysis. Problems that arise in explaining phenomena with regression are specification errors, measurement errors and multicollinearity. Consequently these affect the regression weights.

Unfortunately, many researchers believe betas can be interpreted like correlation coefficients. This error is common in published path analyses as well as regression analyses. Although in some cases the magnitude of the beta weights can give an indication of the importance of the variables in the model, the ever present danger of specification errors should lead researchers to be tentative in their interpretations of these weights. When there is high multicolinearity between independent variables in the model, statements about the importance of any one variable based on betas may be very misleading. When choosing variables to be included in or deleted from a path analysis model, theory especially and probably cost, must be considered along with beta weights.

Although, unstandardized regression coefficients depend on the metric of the variable, they tend to be quite stable from sample to sample. Therefore, their use for prediction purposes or making policy decisions is appropriate. However, the variable may not have been reliably measured or may be interval level, and the weights give no information on the relative importance of the variables in the model.

It has been argued that "when the theoretical model refers to one's standing on a variable, not in an absolute sense but relative to others in the group to which one belongs, standardized coefficients are the appropriate indices of the effects of the variables in the model" (Pedhazur, 1982, p. 249). On the other hand, due to their stability across samples, many authors believe "that the unstandardized coefficients come closest to statements of scientific laws" (p. 249).

It is quite possible, if not probable, to reach very different conclusions about the importance of different variables in regression model depending on whether one interprets standardized or unstandardized regression coefficients. Therefore, regression weights should be tested, and both standardized and unstandardized regression coefficients should be reported in all regression analyses. This applies to path analysis as well as to regression analysis.

Conclusions

Scientific laws are statements of cause and effect relationships among variables. If path analysis is to establish causality, a feat which numerous authors view as impossible (Freedman, 1987; Regosa, 1987), then even its appropriate use of beta weights alone will not accomplish that goal. In good path analysis, as in good regression, the following recommendations should be adhered to. First and foremost, a path analysis should be based on sound theory. It is not an exploratory data analysis technique. Second, despite the cost involved, large samples are desired. Third, tests of the assumptions should be conducted.

Fourth, both standardized and unstandardized regression coefficients and a test of those coefficients should be reported. Fifth, replication and cross validation are needed to confirm original conclusions. Finally, regression and path analyses are correlational techniques, and the results of these analyses should not be reported in the "as-if-by-experiment" mode (Freedman, 1987, p. 108).

## REFERENCES


Biddle, B. J., and Marlin, M. M. (1987). Causality, confirmation, credulity, and structural equation modeling. Child Development, 58, 4-17.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. Multivariate Behavioral Research, 18, 115-126.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of Educational Opportunity. Washington, D. C.: U. S. Department of Health, Education and Welfare, Office of Education.

Freedman, D. A. (1987). As others see us: A case study in path analysis. Journal of Educational Statistics, 12(2), 101-128.

Games, P. A. (1990). Correlation and causation: A logical snafu. Journal of Experimental Education, 58(3), 239-246.

Hayduk, L. A. (1987). Structural equation modeling with LISREL: Essentials and advances. Baltimore, MD: The Johns Hopkins University Press.

Kenny, D. A. (1979). Correlation and causality. New York: John Wiley and Sons.

Ling, R. F. (1982). Review of Kenny, D. A., Correlation and Causation. Journal of the American Statistical Association, 77, 489-491.

Mulaik, S. A. (1987). Toward a conception of causality applicable to experimentation and causal modeling. Child Development, 58, 18-32.

Pedhazur, E. J. (1982). Multiple regression in behavioral research. (2nd Ed.). New York: Holt, Rinehart and Winston.

Regosa, D. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. Journal of Educational Statistics, 12(2), 185-195.

Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. Multivariate Behavioral Research, 21, 309-331.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 18, NUMBER 1, SUMMER 1992

# A Graphical Method for Selecting the Best Sub-set Regression Model

Mark Alexander Constas
University of Northern Colorado

Joe D. Francis
Cornell University

Graphical Method

2

## Abstract

The purpose of the present paper is to provide an empirical example of a graphical procedure that may be used for parameter selection in multiple linear regression. An all possible sub-sets approach is utilized in order to accomplish this objective. Elements related to the model construction process, such as parsimony, and explanation value are explored in the context of this approach. A graphical presentation of the findings is discussed as a way of facilitating the understanding of how $R^2$, adjusted $R^2$ and the number of parameters in a model are related to one another.

# A Graphical Method for Selecting the Best Sub-set Regression Model

Stated in general terms, the purpose of many studies using regression analysis is to determine which variables or combination of variables offers the best prediction for a given dependent variable. Most will agree that there does not exist a single indicator or method of determining the absolute goodness of a given model. It is often suggested that researchers employ the tripartite criteria of theory, economy, and explanation, where theory is concerned with the set of theoretical rationalizations for proceeding with a selected group of variables, economy is centered around issues such as simplicity or efficiency of explanation (i.e. how many variables does it take to achieve some reasonable level of prediction) and explanation is concerned with the amount of variance explained by a given model.

The primary purpose of the present paper is to describe a graphical method for applying the criteria of economy and explanation in the process of constructing a prediction model using regression analysis. Of central importance here is illustrating the relationship between $R^2$, adjusted $R^2$ and the number of parameters in a model. After providing a brief description of the empirical context of the analysis, the paper will proceed to illustrate the way in which a graphical comparsion of $R^2$ and adjusted $R^2$ makes clear an important concept in the parameterization of multiple linear regression problems. It is argued that the graphical clarity of the method helps explain the utility of using adjusted $R^2$ in $k+1$ regression problems.

## Empirical Context

The data used to demonstrate the method being discussed here were derived from a study that sought to determine the relative importance of two types of cognitive variables in predicting the clinical skills of medical students. While one domain of cognitive functioning (cognitive preference) was composed of four variables, the second domain (knowledge competencies) was composed of two. Stated more specifically, the cognitive preferences variable was composed of four independent scores that represented an individual's preference for four different kinds of cognitive functioning (Recall, Principles, Application, and Questioning). The knowledge competency domain was composed of two grade point averages that reflected a given student's level of academic achievement for two distinct periods of his/her medical education. In all cases a total of 14 terms were included in the model construction process. The total of 14 was accumulated by having four terms from the cognitive preferences domain, two terms from the knowledge competencies domain and eight interaction terms that were products of the simple terms.

## Analytical Framework

While there are a number of different methods for generating prediction models in the context of multiple linear regression, the most comprehensive and obviously the most exhaustive method involves running regressions between the dependent variable and all possible subsets of the independent variables. With k regressors one may generate $2^{k-1}$ models. As one can see, the number of models to consider will grow to a large number when trying to construct a model with only a small number of variables. With k=14, as in the case for the present empirical example, the number of models generated exceeds 8,000. While the development of high

speed computers has almost trivialized calculation procedures, subsequent decisions about which variables to include in the initial runs and which models to select for further analysis are not simplified.

Within the framework of multiple regression, $R^2$ is often used as a general indicator of the power of a given model. Although $R^2$ exists as a convention for model selection and evaluation there are some rather fundamental limitations of relying on that statistic. For example, it is important to note that $R^2$ will continue to increase as a direct function of the number of parameters (k) in the model. It could be argued that a strong reliance on $R^2$ is inappropriate, given the illusory effects of increasing k. One can see the way in which $R^2$, being a partial artifact of k, may be misleading.

As an antidote to the problems associated with $R^2$, the adjusted $R^2$ has a built in discounting factor that counters this rather serious flaw in $R^2$ by attaching a penalty clause for increasing the value of k (see Darlington, 1968; Kerlinger and Pedhazur, 1982). The equation takes the following form:

$$R^2 (adj) = 1 - (1 - R^2) \frac{(N-1)}{(N-k-1)}$$

where
N = sample size
k = number of parameters

The presence of the "N-k-1" component in the equation has an attenuating effect that provides a correction for increments in $R^2$ that are associated with simply increasing the number of parameters in a given model.

## A Graphic Demonstration

A basic consideration in model construction often concerns the number of parameters to include in the model. When an a priori decision has not been made regarding parameterization one must proceed in an inductive fashion where empirical outcomes more actively determine the number of variables to include in a given model. In the present example $2^{14-1}$ equations of varying combinations and lengths were generated in order to find the maximal prediction equation.

The graphic approach for selecting the maximum value for k involves plotting the $R^2$ and adjusted $R^2$ values against k. This procedure gives one a visual display to help determine the point at which the incremental value of $R^2$ is insufficiently large to counter the unwanted effects of increasing k. In a typical plot of $R^2$ against k, the curve rises more steeply or less steeply, depending on the nature of model specification. After the addition of a certain number of parameters the curve will usually begin to flatten. The notion of using a flattening area as a termination point for adding additional predictors is often employed as a decision rule in model construction. To many, this decision rule may appear questionable, since the perception of flatten may seem subjective.

A plot of $R^2$ against k fails to reveal a definite turning point. By comparison the adjusted $R^2$ aginst k plot demonstrates a distinct point of descent. More than a mere perturbation, there is a very real turning point to be observed. This point may serve as a ceiling for the number of parameters to be used in model construction.

---

Insert figure 1 about here

---

Figure 1 provides an illustration of the relationship between $R^2$, adjusted $R^2$ and the number of parameters in the model. The point of descent mentioned above appears in a rather clear way. One should also notice the way in which $R^2$ continues to increase in relation to $R^2$ (adj.) This information suggest that one should not proceed beyond a certain level of k. Although the area between k=3 to k=7 should be considered more closely, the level of k selected should certainly not exceed seven.

## Subsequent Procedures

Having decided on the number of parameters to include in the model, issues such as simplicity, theoretical relevance, and ease of explanation may be considered more closely. The next step is may be to obtain the combinatoric options for k=1 to k=7. For purposes of illustration, permutations of variables, for only the top two candidates at each level of k are presented in Table 1.

---

Insert Table 1 about here

---

The primary criterion that one may apply at this point is often invoked under the term "parsimony." A "parsimonious" model is one that contains the parent terms of any interaction terms that may appear in the model while simultaneously using the fewest number of parameters to achieve the

greatest amount of explanation. Application of this criterion led to the selection of the model marked as "tested" in Table 1. Although there is a gain of approximately .4 when moving from three to seven parameters, it was decided that the value of this increment is dubious, given the cost. The necessity of using four more parameters does not support the notion of parsimony. The model selected could then be subject to more detailed statistical scrutiny such as tests of significance.

## Summary Statement

There are a wide variety of methods for constructing models in multiple regression. In the case where one has chosen to use the all possible regressions approach some defensible procedure is needed to help make decisions about the size and contents of a final model. Admittedly, The model construction procedure followed here was not informed by an incredibly strong theoretical base, hence the decicision to proceed with the all possible sub-sets approach. Such a situation is not uncommon in social and educational research. Results of the kind obtained here may provide one with enough empirical evidence to perform a replication or to forge an inductively derived theoretical base. Some progress may be realized.

The relation between $R^2$, adjusted $R^2$ and the number of parameters in the model is an important one to understand. Although a tabular display of these date will reveal the relationship, a graphical expression may make the association more explicit. In summary, one may argue that the present approach to generating a regression model is useful in at least two areas. Firstly, it provides one with a reasonably objective method for defining the upper limits for model construction. Secondly, the graphical method has proven to be quite useful in instructional settings for demonstrating the weaknesses associated with the $R^2$ selection method. The procedure is also

useful in that it provides a rather telling illustration of the relationship that exists between $R^2$ adjusted $R^2$, and the number of parameters in a model.

## References

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

Kerlinger, F., and Pedhazur, E. (1973). Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston.

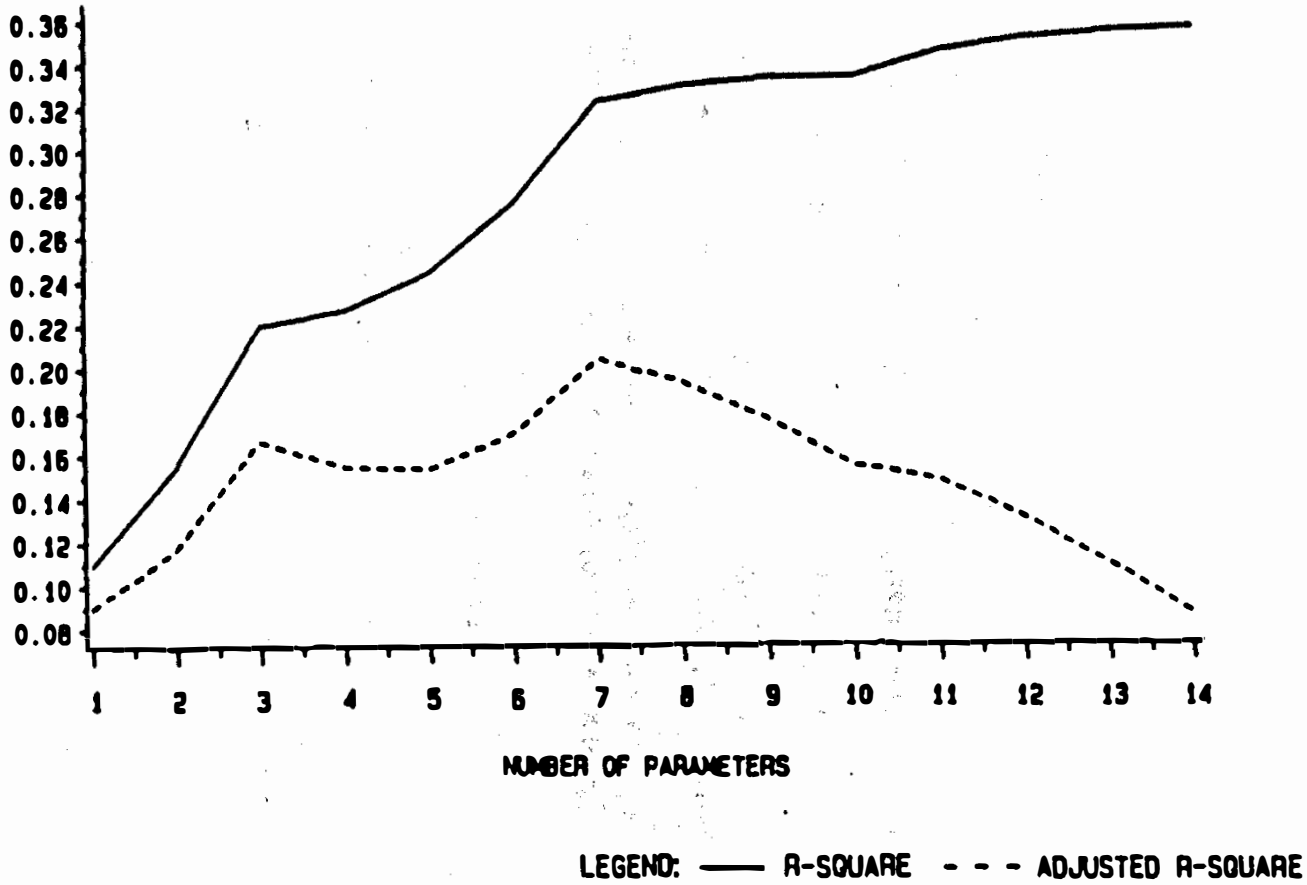


Figure 1. Graphical relationship between $R^2$, adjusted $R^2$, and the number of parameters in the model

## Table 1

### Model Specifications for Prediction of Clinical Skills

| k | $R^2$ (adj.) | Terms in the Equation |
|---|---|---|
| 1 | .090 | (principles x $GPA^a$) |
| | .089 | (Application x $GPA^b$) |
| 2 | .117 | (Questioning), (Questioning X $GPA^b$) |
| | .107 | (Application), ($GPA^b$) |
| 3 | .167* | (Application), ($GPA^a$), (Application X $GPA^a$) |
| | .129 | (Application), ($GPA^b$), (Application X $GPA^b$) |
| 4 | .155 | (Application), ($GPA^a$), (Principles X GPAb), (Application X $GPA^a$) |
| | .152 | (Application), ($GPA^a$), ($GPA^b$), (Application X $GPA^a$) |
| 5 | .154 | (Principles), (Questioning), ($GPA^b$), (Principles X $GPA^b$), (Questioning X $GPA^b$) |
| | .153 | (Principles), (Application), ($GPA^b$), (Principles X $GPA^a$), (Application X $GPA^b$) |
| 6 | .170 | (Principles, (Application), (Principles X $GPA^a$), (Application X $GPA^a$), (Questioning X $GPA^a$), (Questioning X $GPA^b$) |
| | .163 | (Principles, (Application), ($GPA^a$), ($GPA^b$), (Principles X $GPA^b$), (Application X $GPA^a$), (Questioning X $GPA^a$) |
| 7 | .204 | (Principles), (Questioning), ($GPA^a$), (Recall X $GPA^a$), (Recall X $GPA^b$), (Principles X $GPA^b$), (Questioning X $GPA^a$) |
| | .187 | (Principles), (Questioning), ($GPA^b$), (Principles X $GPA^b$), (Application X $GPA^a$), (Application X $GPA^b$), (Questioning X $GPA^a$) |

**Key:**

k Number of parameters in the model

$R^2$ (adj.) Adjusted $R^2$

* Model tested in subsequent analysis

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

# A Case for Interpreting Regression Weights

**Clifford E. Lunneborg**
**University of Washington (U.S.A.)**
**The Open University (U.K.)**

***Abstract***

McNeil (1990) argues against interpreting estimated linear model parameters or weights, largely on the basis of the expected sample-to-sample variability in those estimates. In rebuttal it is noted that not to interpret model parameters is to ignore the strength of regression analysis. Appropriate regard may and should be given parameter uncertainty. But that is only part and parcel of parameter interpretation. Examples of linear model parameter interpretation are given.

***Introduction***

In a recent article in this journal McNeil (1990) writes, "Although most multiple regression texts argue against interpreting regression weights . . . some statistics text authors and researchers still want to place some sort of importance or meaning on the magnitude . . . of regression weights." Count me among them. Let me announce my loyalties even more strongly. I place not just "some sort of importance" on parameter interpretation; I regard interpretability as the central feature of a linear models approach to the analysis of both observational and experimental data. It is what bolds us safe from the sterility of unrelieved null hypothesis testing. The case for interpretation I will base on a series of examples.

***Examples of Linear Model Parameter Interpretation***

**Simple linear regression.** Consider a simple (one explanatory variable) linear model. I'll assume the regression of Annual Income (in thousands of dollars) on Years of Education (in school years satisfactorily completed) is linear in some population of educated and employed individuals. So, we can write:

$$\text{E(Annual Income | Years of Education)} = \beta_0 + \beta_1 \text{ (Years of Education)} .$$

Our response variable (RV) and explanatory variable (EV) both possess metrics. So we have metrics for the regression slope and intercept as well:

$\beta_0$ : The expected Annual Income for an individual with 0 Years of Education. $\beta_0$ is some value in "thousands of dollars."

$\beta_1$ : The increase in expected Annual Income associated with an increase by one year in the number of Years of Education. $\beta_1$ is some value in "thousands of dollars per year."

Our slope or rate of change parameter has a simple and, I believe, very appealing interpretation. It tells us "how much" Education impacts Income. $\beta_1$ might be \$10/year or \$100/year or \$1,000/year or \$10,000/year.

Quite likely a Years of Education score of 0 is outside our range of interest; indeed, the distribution of Annual Income conditional on Years of Education being zero may be without any members. So, any interpretation of the intercept is uninteresting. We might anticipate this and choose to write our linear model in terms of a "Centered" Years of Education. In particular we might reduce Years of Education by a constant of 12 years giving

$$\text{E(Annual Income | Years of Education)} = \beta_0 + \beta_1 \text{ (Years of Education} - 12) .$$

Our slope parameter has its same interpretation. $\beta_0$, though, is now the expected Annual Income (in thousands of dollars) for a high school graduate, a substantively more interesting quantity.

Given an appropriate sample from our population we can estimate these regression parameters. And, granted the satisfactoriness of our sampling assumptions, we can also know how much confidence to place in those estimates. It is my thesis that the point estimate of $\beta_1$ and its standard error are useful because we want to know how big the rate of change is, not because they allow us to "decide" between "rejecting" and "failing to reject" an hypothesis that $\beta_1$ is zero. My support for this borrows heavily from Tukey (1991); I but paraphrase.

Consider the following four possible confidence intervals for $\beta_1$, all, say 95% CIs:

Case A: [−\$10, \$8]

Case B: [−\$3,000, \$4,000]

Case C: [\$4,800, \$5,100]

Case D: [\$10, \$10,000]

Under Cases A or B we "fail to reject" the null hypothesis. But what a difference. Case A ought to tell us that the slope is flat; no question about it, expected Annual Income does not change with Years of Education. On the other hand, Case B ought to tell us that "we haven't the foggiest" whether Annual Income goes up, down or sideways! And, under either Case C or Case D we reject the null hypothesis. Yet Case D is rather like Case B in the lack of precision in our $\hat{\beta}_1$ while Case C allows us to say that an additional Year of Education increases the expected Annual Income by "almost exactly $5,000." How we "decide the null hypothesis" is much less relevant than what we've learned about $\beta_1$.

McNeil (1990) inquires relative to the formula for the circumference of a circle,

Circumference = ($\pi$) (Diameter),

". . . what does $\pi$ mean? $\pi$ is simply the weight which, when multiplied times the diameter, yields the circumference." I have added the emphasis. McNeil dismisses $\pi$ too readily, as if all that were important about it is that it is some constant. But there is more to $\pi$. We think of it as a dimensionless number, but in the context of our Circumference equation it is a rate of change with a metric like in./in. or mm./mm. depending upon how we choose to measure Diameter. $\pi$ is the amount by which the Circumference increases for a one unit increase in Diameter. Increase the Diameter of a circle by 1 inch and you increase its Circumference by (approximately) 3.14 inches. And the value of $\pi$ has practical importance; it is a particular constant and it makes a day-to-day difference that it's value is what it is and not 5 nor 15 nor 1/5. Put another way, it is not sufficient to know that the Circumference of a circle is influenced by its Diameter or, equivalently, that $\pi$ is greater than zero! As with mighty $\pi$, so too with our lowly $\beta$s.

**Multiple Linear Regression.** Now let's extend our Annual Income model by introducing a second EV, Parental Income (also measured in thousands of dollars per year). We write a model additive in the two EVs:

E(Annual Income | Years of Education, Parental Income) =

$$\beta_0 + \beta_1(\text{Years Education}) + \beta_2(\text{Parental Income}).$$

What interpretation do we give the $\beta_1$ of this model? It may be a little easier to see if we rewrite our linear model in the form

$$E(\text{Annual Income} \mid \text{Years of Education, Parental Income}) =$$
$$[\beta_0 + \beta_2(\text{Parental Income})] + \beta_1(\text{Years Education}) .$$

The "slope" parameter, $\beta_1$, is still the expected increase in Annual Income for a one year increase in Years of Education (thousands of dollars per year of education). But, in this model the "intercept" takes different values depending upon Parental Income. So, our $\beta_1$ here has a **conditional** interpretation: The increase in expected Annual Income for a one year increase in Years of Education, for a fixed level of Parental Income.

Often an important question for modelled phenomena like this is whether the $\beta_1$ of our two EV model has essentially the same magnitude as the $\beta_1$ of our one EV model. Is the "influence" of Years of Education on Annual Income the same when we control for Parental Income (our conditional rate of change parameter) as when we ignore Parental Income? Note that the answer to this question has little to do with whether $R^2$ increases signficantly from the one to the two EV model. It has everything to do, of course, with the **substantive importance** of alternative values of $\beta_1$. As we have only estimates of the conditional and marginal rates of change we may seek refuge in the SEs. I emphasize, though, that the comparison is not a statistical but a substantive one.

Kleinbaum, Kupper & Muller (1988) discuss this comparison more fully, albeit under the somewhat pejorative title of "confounding." They take the position that where the two $\beta$s differ, we should prefer the conditional slope. That seems unwarranted. The two answer different questions. "What increase in Annual Income is expected for an additional Year of Education?" is one question. "What increase in Annual Income is expected for an additional Year of Education among those whose parents have identical annual incomes?" is a different question. We may be interested in whether the answers are the same or different, but to prefer one to the other is to pre-suppose the substantive question.

**Moderated Regression Models.** Our two EV model posits additive influences of Years of Education and Parental Income on our RV. More specifically, the "slope" parameter for Years of Education is assumed to be a constant, independent of Parental Income. We might have reason to doubt this assumption. It could be more realistic to assume that the Annual Income contribution of an extra year's education might itself be a function of Parental Income. In more conventional regression lingo we think that (a) Parental Income might **moderate** the influence of Years of Education or (b) we might need **separate slopes** for different Parental Income levels (as well as separate intercepts.)

The usual way of writing a moderated regression model is to allow the intercept and the regression slope of one EV (the moderated EV) each to be linear functions of a second EV (the moderator). In the present context we could write:

$$\text{E(Annual Income} \mid \text{Years of Education, Parental Income)} =$$
$$[\beta_0 + \beta_2(\text{Parental Income})] + [\beta_1 + \beta_3(\text{Annual Income})](\text{Years Education}) =$$
$$\beta_0 + \beta_1(\text{Years Education}) + \beta_2(\text{Parental Income}) + \beta_3(\text{YearsEducation}*\text{ParentalIncome}).$$

The bottom line above describes how we would "input" our regression model, introducing a product variable. It may be a good model to fit but it is quite unsuitable for interpretation. The slope parameter for the product variable, $\beta_3$, has (at least) two strikes against it: (1) Its metric is "thousands of dollars in Annual Income per unit of the product of Years Education and Parental Income." What a "unit" of the latter amounts to is not easy to grasp! (2) Even if we could come to terms with this complicated metric we are warned off interpreting $\beta_3$ because of its conditional nature. In effect, it assesses the contribution of the product variable when the other EVs in the model are held constant. But how can we think about a unit increase in the product of 2 EVs while each is held constant?

Fortunately, the intermediate expression above for our moderated regression model does invite interpretation. The regression slope for Years Education is given as:

$$[\beta_1 + \beta_3(\text{Parental Income})] .$$

This representation is faithful to the moderated regression assumption; the influence on Annual Income of an additional year of education varies with Parental Income. Given estimates of $\beta_1$ and $\beta_3$ the regression slope estimate is easily calculated for a selection of Parental Incomes of interest, say, \$20K, \$40K, \$80K, \$160K, etc. And, if our regression program provides (as it ought) the variance covariance matrix for the $\hat{\beta}$s, it is also easy to calculate SEs for such linear forms of the $\hat{\beta}$s as $(\hat{\beta}_1 + 20{,}000\ \hat{\beta}_3)$. Thus, CIs for the slope estimate at different values of the moderator can be provided.

**Quadratic Regression.** These ideas generalize to quadratic regression and, should the need ever arise to model RVs that double back on themselves in our design space, higher order polynomial regression. Say we thought Annual Income to be influenced quadratically by Years of Education as in this linear model:

$$E(\text{Annual Income} \mid \text{Years Education}) = \beta_0 + \beta_1(\text{Years Education}) + \beta_2(\text{Years Education})^2 .$$

(Years Education)$^2$ is not likely to vary independently of (Years Education) so there is little prospect of interpreting the two separate conditional slope estimates, $\beta_1$ and $\beta_2$.

However, if we rewrite the quadratic model as

$$E(\text{Annual Income} \mid \text{Years Education}) = \beta_0 + [\beta_1 + \beta_2(\text{Years Education})] (\text{Years Education})$$

there is a single slope to be estimated, but one which takes on different values depending upon where in the range of Years Education we want to estimate that slope. Quadratic regression is a special case of moderated regression; moderated and moderator variables are the same variable.

**Interaction Models and Modular Models.** One last example. We make Annual Income now a (probabilistic) function of two *categorical* EVs. We'll assume the population of interest to be college graduates and we are interested in modeling Annual Income (first year post-baccalaureate) as a function of Gender and Degree Major. For simplicity, the later takes only three "levels": Science, Social Science and Humanities. Allowing for the possibility of an interaction between Gender and Major we would likely begin modeling with a six parameter model. If our immediate goal were to test for a (significant) interaction this initial model might look like this:

$$E(\text{Annual Income} \mid \text{Gender, Major}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

where dummy variables have been employed as follows:

$X_1$: 0 for males, 1 for females.
$X_2$: 0 for Science or Humanities, 1 for Social Science
$X_3$: 0 for Science or Social Science, 1 for Humanities
$X_4$: " direct product, $X_1 * X_2$
$X_5$: the direct product, $X_1 * X_3$

The last two EVs can be thought of as "interaction variables" and the hypothesis of no interaction is tested by comparing the overall fits ($R^2$ or SS Residuals) of this model with one in which $\beta_4$ and $\beta_5$ are constrained to be zero (or, equivalently, $X_4$ and $X_5$ are "dropped" from the model.) If the difference in fits is non-significant we declare for the reduced, four parameter, additive model. We detected no interaction. Let's say, though, that the difference in fits was significant; either $\beta_4$ or $\beta_5$ or both are non-zero. Gender and Major do interact in influencing Annual Income. What do we do?

My belief is that we ought to do more than report that the interaction is significant or that the $R^2$ for the six parameter model is significantly larger than the $R^2$ for the four parameter model. We ought to "interpret" the interaction; how do Gender and Major interact? The $\beta$s for our two interaction variables, having as their metrics products of dummy variables, are not the best candidates for yielding up the desired interpretation. What works for me is to re-parameterize the interaction model into one with parameters that are themselves easily defined and give clear insight into the interaction.

First, what does the finding of an interaction mean, substantively? That the relative influence of the several Major levels on Annual Income is different for males than for females. Having learned this, it behooves us to model Major influence for males separately, somehow, from our modeling of Major influence for females. One way of looking at it is to say we want now to examine "simple" rather than "main" Major effects. That is facilitated by the re-parameterization to a modular model. The idea of the modular model is that it is equivalent to the interaction model (in numbers of parameters and fits) but consists of separate "modules" for each level of a categorical EV. (In the case of higher order interactions the modules may be for lower-order interaction "levels".) Modular models have been explicated primarily by writers on the use of weighted least squares in the analysis of categorical response data, e.g., Forthofer and Lehnen (1981) or Freeman (1987). However, they are equally useful in the linear modeling of a continuous RV. Here, we'd like separate modules for males and females.

Our modular model might look like this:

$$E(\text{Annual Income} \mid \text{Gender, Major}) = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_5 + \beta_6 Z_6$$

where the linkage to our earlier dummy variables is as follows:

$Z_1$: $X_1$ (a 0/1 variable coding female)

$Z_2$: 1 - $X_1$ (a 0/1 variable coding male)

$Z_3$: $Z_1$*$X_2$ (a 0/1 variable coding female and social science)

$Z_4$: $Z_2$*$X_2$ (a 0/1 variable coding male and social science)

$Z_5$: $Z_1$*$X_3$ (a 0/1 variable coding female and humanities)

$Z_6$: $Z_2$*$X_3$ (a 0/1 variable coding male and humanities)

$Z_1$, $Z_3$ and $Z_5$, together with their "weights", comprise the female module; $\beta_1$ is the "intercept" for the female module and $\beta_3$ and $\beta_5$ are the female module slopes for the dummy variables coding social science and humanities respectively. Correspondingly, the male module is based on $Z_2$, $Z_4$ and $Z_6$. Given our particular use of dummy variables, the intercepts evaluate to the expected Annual Incomes for (male and female) science graduates and the slopes to the differences between the expected Annual Incomes for either social science or humanities and those expected for science graduates (again, separately for males and females).

In fitting the modular model we obtain SEs for the six parameter estimates. While the presence of an interaction insures that we cannot have $\beta_3 = \beta_4$ and $\beta_5 = \beta_6$ simultaneously, we may be able to simplify the modular mode further, guided substantively by our re-defined parameters. The main point, though, is that the parameters of the modular model are directly interpretable and their estimates can be used to "explain" the interaction.

## *Discussion*

I have tried, by example, to make the case for the directness and substantive importance of parameter interpretation in linear models. Why should it be controversial? I have not addressed that question but I think there are two issues involved. The first has to do with the stages of modeling, from model formulation through fitting and model comparison and on, perhaps, to model adoption. How we view a model and the relevance (or, indeed, acceptability) of parameter interpretation can depend upon the stage of modeling at which we are operating.

The second issue has to do with a contrast between phenomena that are thought to be wholly deterministic and those with an inescapable stochastic element. How we assess our success in modelling will depend on how much determinism we attribute to the phenomenon modeled.

**Model Fit, Comparison and Interpretation.** I ought make it clear that the model parameters whose interpretation concerns me are, for the most part, parameters in "accepted" or final models. I assume that we pursue our modeling with several alternative models in mind. These may all be pre-specified models, well rationalised in advance of any data collection or they may be models whose origin owes something to the "lay of the land" once we have it in sight. In either event, we are interested in identifying one or more of these alternatives as "better" than the others. Better, of course, must take into account the purpose for which we wish to find models.

Whatever our goal, however our alternative models are suggested to us, the path to an accepted model or models involves fitting several alternative models to our data and then comparing those fits. This fitting and comparing are done on statistical turf and parameter interpretation plays no role. Interpretation comes in after final, or, at least, promising, models have been identified. And it is necessary, in my view, if we are to do the best job of communicating our results. Interpretation, or the prospect of it, should also be kept in mind when we parameterise models. Every linear model permits of several alternative parameterizations, all providing the same $R^2$, the same fit to individual observations. We should choose one, our software willing, that will be natural to interpret later on. And, if our goals accommodate any degree of "model snooping", having parameters with simple interpretations makes it that much easier for us to see our way forward in model simplification or modification.

If we keep in mind where in the modeling process we are, we can make parameter interpretation work for us and not against us. I cannot believe that modeling progress is facilitated if the analyst is "blinded" as to the meaning of parameters throughout the entire course of modeling.

**Deterministic and Stochastic Models.** McNeil (1990) writes " . . . when one utilizes MLR one is taking the stance that behavior is complexly determined . . . The goal then is to account for the variation in the criterion by obtaining as high an $R^2$ as possible.." The emphasis is mine. In an appendix to the same article McNeil equates a "correct model" with one yielding an $R^2$ of 1.0. Both remarks suggest that he is modeling deterministic phenomena; given the right set of EVs, all of the response variability can be accounted for. Unquestionably, behavior is complexly influenced (if not wholly determined) and the search for a highest $R^2$ necessarily leads to models with very many EVs. And, indeed, in a model with 100, 200, perhaps more, intercorrelated EVs, parameter interpretation does become, at best, problematic.

Maximizing $R^2$ for deterministic responses is but one goal to be pursued with linear modeling. Let me suggest some alternatives.

(1) Not all behavioral, biological or social phenomena are deterministic. I mean that in two senses. First, there is the possibility of some inherent randomness; in principle we can never account for all of the variability in the free throw accuracy of NBA players. And, there are human limitations; in practice we shall never identify all of the EVs needed to account fully for the variability in the voting behavior of US state legislators. In either event, the "correct model" cannot extend beyond the EVs that are known to be relevant and will have an $R^2$ substantially less than 1.0.

(2) Even if we take the response to be deterministic, but complexly so, we often make scientific headway by considering, at one time, only a few of the many EVs which are known to be relevant. We seek to learn more about how some EV of interest influences a response. Several of my sketchy examples given earlier had a common theme; how does Years of Education influence Annual Income? Many, many factors other than Years of Education impact earnings. But, that's hardly the point if what I'm interested in is learning how Gender or Parental Income or College Major might moderate the influence of Years Education on Annual Income. If I sample randomly I need not worry overmuch about what else I might have put into my model.

(3) McNeil makes the very important point that the magnitude of an influence we detect for some EV in an observational study may be a poor guide to what happens when we attempt to manipulate that EV. That is a caveat to be heeded in the reporting of any observational study. Having said that, we can do worse in our search for potentially effective manipulations than to pay attention to the magnitudes of observational study influences. When I induce a student to remain in college another year I may not have increased her post-educational income by $5,000 per year. Having noticed in an (hypothetical) observational study that, on average, each additional year of education was associated with that amount of additional income, however, suggests it is a manipulation worth trying, and evaluating.

I believe that a very great many, perhaps the substantial majority of, linear models in the biological and behavioral sciences are of these second and third kinds. They involve a limited number of EVs, often fewer than are known to be relevant to the RV. And, they address one or both of these questions: "How great is the influence, if any, of this EV?" and "How is the influence of my EV changed when I take these other things into account?" In neither case is the $R^2$ as relevant as the

interpreted model parameters.

Parameter interpretation, far from being suspect, should be embraced by the multiple linear regression community. For appropriately parameterized models the parameters and their estimates provide natural measures of the magnitude of explanatory influences. Parameter interpretation is essential if we are to understand the meaningfulness (substantive significance) of an influence as well as its "presence" (statistical significance).

## *References*


Forthofer, R. N. and Lehnen, R. G. (1981.) ***Public Program Analysis: A New Categorical Data Approach.*** Belmont, CA: Lifetime Learning Publications.

Freeman, D. H. (1987.) ***Applied Categorical Data Analysis.*** New York: Marcel Dekker.

Kleinbaum, D. G., Kupper, L. L. and Muller, K. E. (1988.) ***Applied Regression Analysis and Other Multivariable Methods.*** (2d. Ed.) Boston: PWS-Kent Publishing.

McNeil, K. (1990.) The case against interpreting regression weights. ***Multiple Linear Regression Viewpoints,*** 17 (2), 40-44.

Tukey, J. W. (1991.) The philosophy of multiple comparisons, ***Statistical Science,*** 6, 100-116.

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

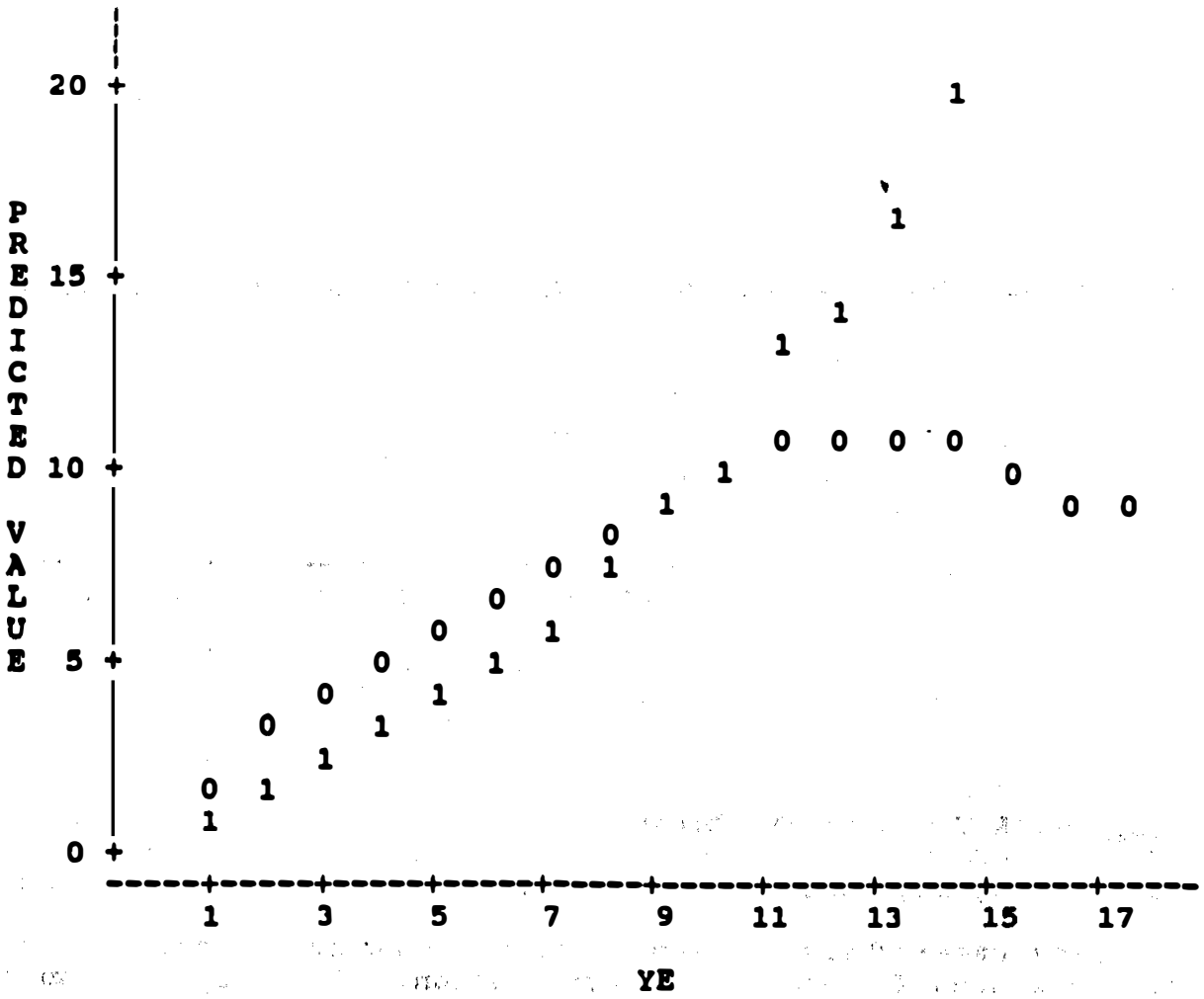


Figure 1. A case wherein linear prediction is high, but misleading for values above 11.

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.

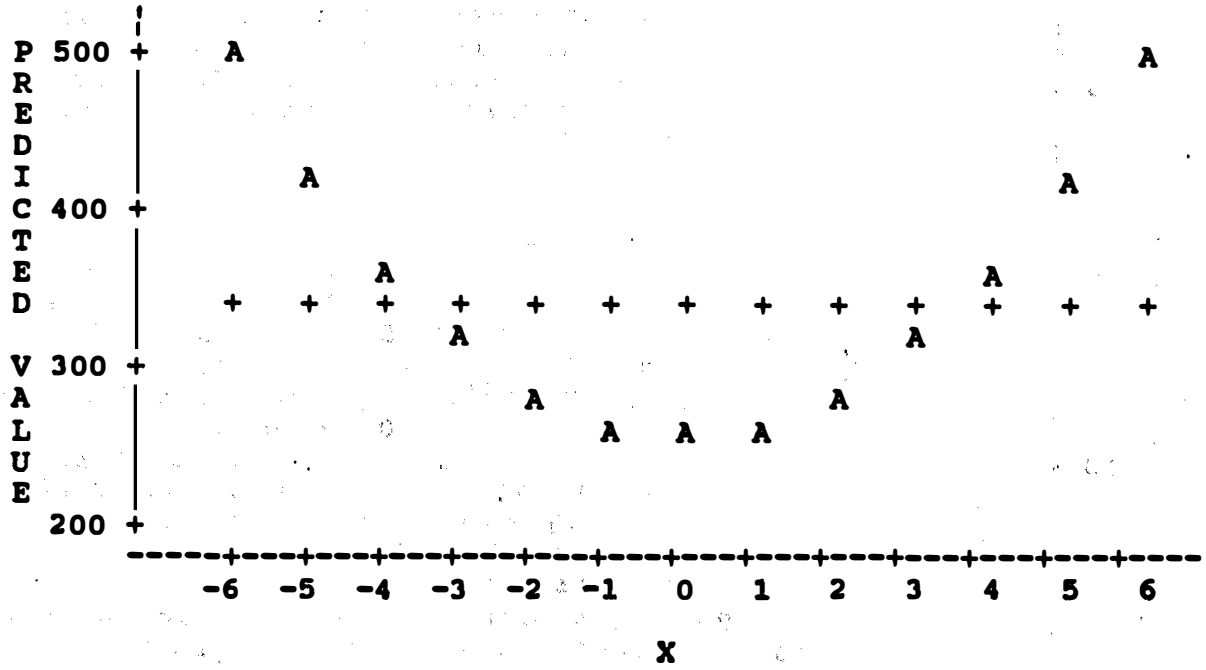


Figure 2. A U-shaped relatinship.

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.

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

If you are submitting a research article other than notes or comments, I would like to suggest that you use the following format if possible:

Title
Author and affiliation
Indented abstract (entire manuscript should be single spaced)
Introduction (purpose—short review of literature, etc.)
Method
Results
Discussion (conclusion)
References

All manuscripts should be sent to the editor at the above address. (All manuscripts should be camera-ready.)

It is the policy of the M.L.R. SIG-multiple linear regression and of *Viewpoints* to consider articles for publication which deal with the theory and the application of multiple linear regression. Manuscripts should be submitted to the editor as original, double-spaced, *camera-ready copy.* Citations, tables, figures and references should conform to the guidelines published in the most recent edition of the *APA Publication Manual* with the exception that figures and tables should be put into the body of the paper. A cost of $1 per page should be sent with the submitted paper. Reprints are available to the authors from the editor. Reprints should be ordered at the time the paper is submitted, and 20 reprints will cost $.50 per page of manuscript. Prices may be adjusted as necessary in the future.

A publication of the Multiple Linear Regression Special Interest Group of the American Educational Research Association, *Viewpoints* is published primarily to facilitate communication, authorship, creativity and exchange of ideas among the members of the group and others in the field. As such, it is not sponsored by the American Educational Research Association nor necessarily bound by the association's regulations.

"Membership in the Multiple Linear Regression Special Interest Group is renewed yearly at the time of the American Educational Research Association convention. Membership dues pay for a subscription to the *Viewpoints* and are either individual at a rate of $5, or institutional (libraries and other agencies) at a rate of $18. Membership dues and subscription requests should be sent to the executive secretary of the M.L.R. SIG."

# TABLE OF CONTENTS



**ISSN 0195-7171**