DOCUMENT RESUME

ED 458 302 TM 033 482

AUTHOR Daniel, Larry G.; Onwuegbuzie, Anthony J.
TITLE Multiple Regression: A Leisurely Primer.
PUB DATE 2001-02-00
NOTE 24p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (24th, New Orleans, LA, February 1-3, 2001).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Analysis; *Regression (Statistics)
IDENTIFIERS *General Linear Model

ABSTRACT

Multiple regression is a useful statistical technique when the researcher is considering situations in which variables of interest are theorized to be multiply caused. It may also be useful in those situations in which the researchers is interested in studies of predictability of phenomena of interest. This paper provides an introduction to regression analysis, focusing on five major questions a novice user might ask. The presentation is set in the framework of the general linear model and builds on correlational theory. More advanced topics are introduced briefly with suggested references for the reader who might wish to pursue the subject. (Contains 11 references.) (Author/SLD)

Running Head: Multiple Regression

Multiple Regression: A Leisurely Primer

Larry G. Daniel
University of North Florida

Anthony J. Onwuegbuzie
Valdosta State University

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA, February 1-3, 2001.

Abstract

Multiple regression is a useful statistical technique when the researcher is considering situations in which variables of interest are theorized to be multiply caused. It may also be useful in those situations in which the researcher is interested in studies of predictability of phenomena of interest. The present paper provides a leisurely introduction to regression analysis, focusing on 5 major questions a novice user might ask. The presentation is set within the framework of the general linear model and builds on correlational theory. More advanced topics are also briefly introduced with suggested references given for the reader who might wish to pursue these further.

Multiple Regression: A Leisurely Primer

Multiple linear regression, the most straightforward form of the general linear model, is one of the more useful of the correlational procedures. It is a powerful tool for testing theories about relationships among observables, and it is also useful for researchers interested in the predictive power of a set of variables. An understanding of linear regression provides a foundational understanding of a number of important statistical concepts, including variable weighting, line of best fit, scatterplots, predictive accuracy, and error of prediction. Finding regression a praiseworthy tool in the scientist's arsenal of statistical weaponry, Fox (1991, p. 3) noted, "regression analysis is the most widely used statistical technique in social research and provides the basis for many other statistical methods."

Despite the obvious usefulness and popularity of regression analysis, many readers of research lack even a casual understanding of regression logistics. Others have a general feel for the concepts of correlation and prediction but get lost in the sea of coefficients and statistical weights that often accompany the reporting of regression results. Still others had training in regression analysis early in their careers but have found their knowledge of regression analysis has become weak due to lack of use of the procedure.

The present paper offers a leisurely, non-technical explanation of linear regression analysis. Emphasis is placed on the use of regression to understand relationships among theoretically important variables. More technical information, such as formulae, where

used, is explained so as to avoid confusion on the part of the less technical reader. In pursuing our purpose, we will focus on five questions that an unfamiliar user of regression might ask: (a) What is the general linear model? (b) What is simple linear regression? (c) What is multiple linear regression? (d) How do I sort through the various correlations, weights, and coefficients available in regression? (e) How do I determine variable contributions to a regression analysis?

What Is the General Linear Model?

The terms "regression analysis" and "general linear model" are often used in the same contexts. The general linear model (GLM) is a broad class of interrelated statistical procedures focusing on linear relationships among variables or variable composites. The term "linear" is used as these techniques can be visually represented by plotting variables against one another on two dimensional charts and utilizing mathematical formulae for determining where to draw one or more lines that will visually represent the relationships among the variables. Regression analysis is the most broad, or general, form of the GLM. What this means, in essence, is that regression analysis forms the basis for many other statistical techniques, or, stated differently, that a number of other common statistical procedures (e.g., analysis of variance, analysis of covariance, *t*-test, Pearson product moment correlation, Spearman rho correlation) are all specially designed versions of regression analysis. Further, regression serves as a general framework for understanding a host of related multivariate statistics, most generally subsumed under canonical correlation analysis (Knapp, 1978).

GLM models include various ways for determining the statistical estimates necessary to producing the desired results of a given procedure. The most commonly used procedure for determining these estimates is the "least squares regression" method. This technique is so named because it computes linear relationships in such a way that the squared differences between actual and predicted values in an analysis are minimized. Though other procedures for estimating GLM models are available, the present discussion is limited to linear least squares methods.

What Is Simple Linear Regression?

Simple linear regression examines the relationship between two variables, one of which is referred to as the predictor variable (i.e., the variable that usually precedes the other), and the other of which is referred to as the criterion variable (i.e., the variable the researcher is interested in explaining, predicting, or better understanding). Because simple regression results give us an understanding of the patterns of relationships between the two variables of interest in a given context, we often use the term "prediction" in describing the relationship. To provide an understanding of prediction, we offer a simple example. When looking out the window and seeing frost on the ground, one would be more likely to put on a coat prior to going outdoors than if one did not observe frost on the ground. In fact, knowing that frost on the ground is usually accompanied by lower temperatures and higher humidity, one might even be able to predict with some reasonable accuracy that individuals who observe the appearance of the ground before going outdoors would be more likely to put on a coat than individuals who do not observe

the appearance of the ground. If we observed enough persons in this type of situation and found a pattern of behavior across these persons, we could then say that there is some reasonable degree of correlation between the amount of frost on the ground and the likelihood of a person wearing a coat. After gathering data on a number of persons, we could predict (with some degree of accuracy) the number of persons that might wear a coat on a given day by observing the amount of frost on the ground.

The type of prediction mentioned in the above example could be addressed using a simple regression analysis involving a measure of the amount of frost on the ground on a given day (the predictor variable) and a measure of the number of persons that day that wore coats (the criterion variable). Simple linear regression allows the researcher to develop a predictive equation to indicate the degree to which one continuous variable (e.g., amount of frost) can accurately predict another continuous variable (e.g., coat wearing behaviors). This procedure is called "simple" because it includes only one predictor variable. Prediction with two or more predictor variables is called "multiple regression analysis."

In prediction studies, the term "prediction" does not necessarily infer that the relationship between the variables of interest is causal. "Prediction" only indicates that the relationship is correlational (i.e., that the predictor variable(s) and criterion variable occur together in a certain pattern). However, some researchers may use predictive studies to *infer* causality based on the theoretical specification of the variables included in a given study.

As in studies employing Pearson product moment correlation (*r*), the linear relationship between the predictor (X) and the criterion (Y) variables can be shown on a two-way scatterplot. A line of best fit drawn through the scatterplot is called the "regression line," and the statistic representing the relationship between the two sets of points is called multiple *R*. One might recall that in Pearson product moment correlation the coefficient *r* is used to show the strength and directionality of a relationship between two variables. A value of *r* close to zero indicates a low or negligible correlation between two variables, while a value closer to $|1|$ indicates a more appreciable amount of correlation between the variables. Negative *r* values depict inverse relationships between variables, while positive *r* values depict direct relationships. Multiple *R* is very similar to the Pearson *r* with the exception that it is always positive in value ($0 \leq R \leq 1$) due to a set of variable weights developed as a part of the analysis. Because there is only one predictor in simple linear regression, *R* will be equivalent to the absolute value of the Pearson correlation coefficient ($|r|$) between the two variables. The square of multiple *R* (R^2) represents the "effect size" for the regression analysis, and can be interpreted as the percentage of similarity in the patterns of the two variables across a given set of observations.

Predictive Equations

Predictive equations are used to determine the degree of accuracy in prediction for any given observation in the regression data set. The predictive equation in simple linear regression applies both additive (a) and multiplicative (b) weights to one variable (the

predictor variable, X) so as to maximally "reproduce" the other variable (the criterion, or dependent variable, Y). The predicted Y score ($\hat{Y}$, or "Y-hat") for each observation in the sample is derived from this predictive equation. The simple regression equation ("regression of Y on X") is:

$$\hat{Y} = a + bX$$

(As you may recall from an algebra class, the above equation looks very much like the mathematical formula you probably learned for a straight line. It is indeed virtually the same formula. Hence, the visual resulting from the formula is called the "regression line.")

Because $\hat{Y}$ is the predicted estimate of Y, this is often represented in the formula as follows:

$$Y \leftarrow \hat{Y} = a + bX$$

The left-pointing arrow ($\leftarrow$) between Y and $\hat{Y}$ may be interpreted "is yielded by," indicating that $\hat{Y}$ is the best estimate of Y given the data in hand. If the order of the equation is reversed:

$$a + bX = \hat{Y} \rightarrow Y,$$

the right pointing arrow ($\rightarrow$) would be interpreted "yields."

Error Scores

It is important to note that the □ values are based on the average goodness of prediction across Y scores in the entire data set. Thus, some $\hat{Y}$ values will be better approximations of their corresponding Y than others. The accuracy of prediction for any

given case in the analysis can be found by subtracting $\hat{Y}$ from Y. The difference is the "error of prediction" (Y_e):

$$Y_e = Y - \hat{Y}$$

The standard deviation (s) of the error (Y_e) scores is called the "standard error of the estimate" and represents the "average" amount of error in the $\hat{Y}$ scores. In a simple regression prediction study, one would want the value of s to be relatively small (close to zero) because a small value would indicate a high degree of accuracy in the prediction equation and a high degree of correlation between the variables of interest to the researcher.

Assumptions in Simple Linear Regression

As with any statistical procedure, simple regression analysis is based on certain assumptions about the data. As noted in various treatises on regression analysis (e.g., Fox, 1995; Hamilton, 1992; Kerlinger & Pedhazur, 1973; von Eye & Schuster, 1998), these assumptions include, but are not limited to, the following:

1. Simple linear regression analysis assumes that all of the variables are normally (or at least quasi-normally distributed). Oddly shaped distributions can yield biased regression results.
2. Regression analysis also assumes that the sample employed is randomly drawn from (or at least representative of) the population of interest.
3. Simple linear regression further assumes that a straight line will be the best way to capture the nature of the relationship between the two variables of interest.

Forcing regression lines on data relationships that are curvilinear will lead to a misunderstanding about the relationship between the predictor and criterion variable.

4. Regression also is based on the assumption of "homoscedasticity" (i.e., that the conditional distribution of the Y_e scores for each value of X is an approximately normal distribution). This is also called the "constancy of error variance assumption." Violation of this assumption may yield results that are atypical in the population.

While it is common for researchers not to take any precaution to assure that these assumptions are met, failing to meet the assumptions can lead to erroneous conclusions about the theoretical concerns the researcher is attempting to test in a given study. Fortunately, there is emerging a greater recognition that assumptions underlying data as well as other common data problems need to be investigated in any study (Wilkinson & Task Force on Statistical Inference, 1999). Fox (1991, 1995) offers a variety of strategies for testing and making data adjustments for violation of regression assumptions and also discusses other data problems in regression analysis.

What Is Multiple Linear Regression?

Multiple linear regression is an extension of simple regression with the difference being the number of predictor variables employed. Multiple regression analyses will include at least two predictor variables (X_1, X_2. . .,X_k). In other words, the analysis

addresses the probability that the criterion variable, Y, is a function of the *set* of predictor variables:

$$p(Y|\ X_1, X_2 . . ., X_k) = f\ (X_1, X_2 . . ., X_k)$$

In other words, the probability p that the dependent variable Y is a function of predictor variables $X_1, X_2 . . ., X_k$ is conditional upon specific values of the predictor variables.

The linear equation for multiple regression is simply an extension of the simple regression equation:

$$Y \square \hat{Y} = a + b_1X_1 + b_2X_2 + . . . + b_kX_k$$

This formula illustrates that the predicted dependent variable score $\hat{Y}$ is a function of the linear weighted combination of the predictor variables $X_1, X_2, . . ., X_k$. Notice that there is a single additive (a) weight (or constant) for the equation and that each of the predictor variables has its own multiplicative (b) weight.

In the social sciences, most events are theorized to be multiply caused, or at least multiply accompanied, by a host of preceding variables. Multiple linear regression allows the researcher to look simultaneously at a host of predictor variables of interest and to examine the collective ability of these variables to predict the criterion variable. The multiple linear regression equation shown above simply combines all of the predictor variables into one composite variable ($\hat{Y}$), and this composite variable then serves as a single (synthetic) predictor variable representing the host of predictors. Once the researcher determines the degree of multiple correlation (R) between the composite of predictor variables and the criterion variable, methods are then typically employed to

further understand the complex set of relationships among the variables. This is accomplished by consulting various coefficients yielded by the analysis.

How Do I Sort Through the Various Correlations, Weights, and Coefficients Available in Regression?

Like most statistical procedures, regression yields a variety of statistical indices that aid in making interpretations of the data. Unfortunately, these various indices can become difficult to understand particularly when they are reported apart from explanatory statements as to their value or meaning. We discuss here three types of indices commonly used in social science literature employing regression analyses, namely correlations, weights, and structure coefficients.

Correlations

There are essentially two types of correlations that one might generate and interpret when conducting a regression analysis, namely simple correlations (Pearson r values) and the multiple correlation (R). Pearson r is used to express a simple linear relationship between two variables of interest, whereas R, as previously indicated, expresses the relationship between the composite of predictor variables and the single criterion variable. In the simple regression case, only one Pearson r can be generated, and the value of this r will be the same as the multiple R (except, possibly, for the sign of the r coefficient) considering that there is only one predictor variable in the analysis. In a two-predictor case, there will be three different r values possible: (a) correlation between X_1 and Y, (b) correlation between X_2 and Y, and (c) correlation between X_1 and X_2. It is

our opinion that simple correlations between each of the predictors and the criterion variable should routinely be reported in regression studies. These values help us to determine at a basic level whether each of the several predictor variables, in and of itself, is appreciably related to the criterion variable. Based on these values, the researcher might determine whether a variable should be discarded prior to conducting the regression analysis (e.g., in the case in which the correlation with the criterion variable is nearly zero). We wish to warn, however, that this should not simply be a "fishing expedition" designed to select out a set of promising variables from an array of variables for which the researcher just so happens to have data. Variables should always be selected for consideration in an analysis due to their theoretical importance, not due to their availability. The practice of examining the correlation of each predictor variable with the criterion variable prior to conducting a regression analysis, when *all of the variables were preselected due to their theoretical importance* may assist in building more parsimonious models for the regression analysis.

Examining the correlations (r's) between each pair of predictor variables can also be useful. These values help us to determine the degree to which the predictors are "collinear" with one another. Collinearity introduces a variety of problems into the regression analysis, most of which are associated with the derivation of the statistical weights for the various predictor variables (Schroeder, Sjoquist, & Stephan, 1986). These problems can be dealt with, as described later in this paper, and it is common to expect that predictor variables collected within a single context will be at least somewhat

correlated with one another; however, it is important to note that when collinearity exists, the statistical weights yielded by the regression analysis will not be uniquely determined. Further, including two predictor variables in the regression analysis that are nearly perfectly collinear (i.e., that are correlated at or near $|1.0|$) is not sensible considering that, once one of the variables is included, the second variable will offer no information that is not already provided by the first. In this case, it would be preferable to drop one of the variables before proceeding with the regression analysis.

Once the Pearson *r*'s have been consulted (and, if necessary, variables have been eliminated from the analysis), the regression analysis is then conducted. The most important correlational statistic yielded by the regression analysis is multiple *R*. The squared value of multiple *R* (R^2) represents the statistical effect size, and, as previously, noted, it can be interpreted as a percent of relationship between the predictor variable set and the criterion variable. If testing for statistical significance of the regression results, R^2 can also be useful as it expresses the portion of the criterion variable's sum of squares that is accounted for by the predictor variables. In absence of other information, R^2 along with the total sum of squares and the sample size provides enough information to calculate an *F* statistic used in testing for statistical significance.

Weights

The a and b weights yielded by the regression analysis each can be useful in interpreting the results. The multiplicative weight (b) used in the equation is also called the "slope" and the "regression coefficient". The additive weight (a) is also called the "Y-

intercept," the "constant," and the "regression constant." The a weight represents the point at which the regression line will cross the Y-axis (the value of Y when X = 0); hence, the a weight estimates the value of the dependent variable when all predictors have a value of zero. The value of the b weight indicates the number of units that the criterion value is predicted to change if the given predictor variable is increased by one unit. If the b weight is negative, the value of the predicted dependent variable will decrease by b units when X is increased by one unit. Conversely, if b is positive, the value of the predicted dependent variable will increase when X is increased.

If all of the variables in the analysis are standardized (i.e., converted to z-scores), the regression equation for k variables takes the following form:

$$z_Y = \beta_1 z_{X1} + \beta_2 z_{X2} + \ldots + \beta_k z_{Xk} + e$$

Note that with standardized data, there is no longer an additive weight (e in the equation represents the error [Y_e] score, not a statistical weight). The beta weights shown in this equation are also known as "standardized regression coefficients" and are often used over the b weights derived using the raw data because it is easy when comparing betas to make direct comparisons as to the relative amount of weight each variable is being assigned. The b weights, on the other hand, are not as easily comparable as they are affected by the metric in which the particular predictor variable is measured. Variables collected in very large units will tend to have relatively small b weights even when the variables are substantially related to the criterion variable.

If weights are omitted from a report of a regression analysis, they can often be

determined from other information available to the researcher. For example, if the correlation (r) between X and Y is known, the b weight can be computed by multiplying r by the quotient of the standard deviation of Y (s_Y) divided by the standard deviation of X (s_X):

$$b = r\,(s_Y / s_X).$$

Once b is known, the constant (a) can be computed using the following formula:

$$a = Y - bX$$

<u>Structure Coefficients</u>

Although structure coefficients have not typically enjoyed the popularity of other statistical indices accompanying regression analyses, they are very useful aids in interpreting regression results. As explained by Thompson and Borrello (1985), a structure coefficient (r_s) is the correlation between each of the predictors and $\hat{Y}$. Structure coefficients, therefore, express the degree of relationship of a predictor with the predicted values of the dependent variable, or, stated differently, express the degree to which a given predictor is "reproduced" in the computation of $\hat{Y}$. In the simple regression case, the structure coefficient for the single predictor is $|1.00|$ considering that the □ is simply a linear transformation of the predictor. In multiple regression, structure coefficients are often large for some predictors and smaller for others.

How Do I Determine Variable Contributions in Regression Analysis?

Variable contributions to a given statistical result are often among the most important outcomes of research studies. Determining that a particular variable has had

an appreciable impact on the regression results often has implications for both theory and practice. Conversely, determining that a theorized variable is not a major contributor to the regression results may lead to modification of a theory to exclude the variable. Obviously, findings regarding variable contributions would have to replicate across a number of studies over time and research conditions in order for the researcher to make confident decisions about theoretical matters; however, within the context of a single study, it is still important to accurately determine what variables are contributing to the regression results and to what extent. At least four means for determining variable contributions have been proposed for use with regression analysis: (a) stepwise variable entry procedures, (b) consultation of beta weights and structure coefficients, (c) follow-up commonality analysis, and (d) all possible subsets regression. Each of these procedures is discussed briefly.

Stepwise Variable Entry Procedures

Stepwise methods have been used with regression and other related statistical procedures for some time. These methods use a variety of procedures for determining what order to enter variables from the predictor set into the regression equation. Some researchers erroneously think that the order of stepwise entry gives information regarding the order of importance of the variables to the analysis. However, this is not the case. Stepwise methods are problematic for many reasons and have often been criticized by reflective methodologists (e.g., Huberty, 1994; Thompson, 1995). For one thing, stepwise methods use miscalculations of degrees of freedom when determining variable entry

order. This can have serious implications on the statistical significance of the results obtained.

Perhaps, more significantly, stepwise procedures can also lead to misunderstandings about order of variable importance. For a set of *k* predictor variables used in an analysis, stepwise procedures seek first to determine which one predictor serves as the best predictor of a given criterion variable. Determination is then made as to the predictor making the second best contribution to explaining the criterion once the variance explained by the first predictor is removed from the analysis, the third best contributor once the variance explained by the first two predictors is removed, and so on. An alternative to this "forward" stepwise procedure is the "backward elimination" method. Backward elimination puts all the predictors into the model and then removes them one at a time based on the variable that makes the least unique contribution to the analysis at any step of the procedure. Frequently, researchers utilize stepwise routines that combine both of these procedures (largely because this is the default for a stepwise routine in many statistical programs), resulting in a stepwise analysis that may alternately include or exclude variables at any step of the analysis based on the degree to which a previously excluded variable makes a uniquely "new" contribution or to which a previously included variable is found to offer a primarily redundant contribution at the particular step of the analysis.

Unfortunately, because at each step of the analysis, the variance previously accounted for is no longer considered as "explainable" variance, stepwise procedures, at

each step, yield information based on the "best" predictor in absence of the full variance being accounted. Hence, a predictor that is highly correlated with the criterion variable but that is also highly collinear with another predictor may be entered into the analysis at a much later step than variables that are only marginally related to the criterion variable. Consequently, we advocate that researchers seriously consider abandoning stepwise procedures altogether. This will not only eliminate the problems mentioned herein, but will also eliminate other deleterious problems related to stepwise methods (Thompson, 1995).

Consultation of Beta Weights and Structure Coefficients

There is a tendency for researchers to utilize beta weights when assessing variable contributions in regression. As noted earlier, these are the statistical weights applied to the standardized predictor variables in a regression. Though it seems logical to conclude that a larger beta weight implies a greater contribution, this is not necessarily the case. This is particularly problematic when predictor variables are collinear. In fact, in cases in which collinearity among predictors is moderate to large, beta weights are extremely subject to distortion and can lead to erroneous conclusions about importance of variables (Thompson & Borrello, 1985). By contrast, regression structure coefficients are not as prone to distortion. Recall that structure coefficients are correlations, not weights. Hence, even in cases in which predictors are collinear, correlations remain relatively constant, and hence structure coefficients are not as acceptable to the collinearity problem. Therefore, in determining which variables are most instrumental in regression,

it is our opinion that structure coefficients provide more reliable evidence than do beta weights.

Follow-up Commonality Analysis

Yet another method for examining variable contributions in regression is the use of follow-up commonality analysis. Once a regression is initially run, commonality analysis uses a series of algebraic equations to determine the degree to which the contribution of any variable, or set of variables, to the analysis is unique and the degree to which the contribution of any variable is common with any other variable or variable set. Though infrequently used, this procedure is a useful means for comparing the behavior of variables in isolation versus their behavior in the company of other variables (Thompson, 1990).

All Possible Subsets Regression

All possible subsets (or best subset) regression is a procedure similar to commonality analysis. This procedure compares results obtained from all possible combinations of the predictor variables included in an analysis and determines which combination yields the best results with the most parsimonious use of researcher and/or data resources. A problem with this procedure is the large number of possible subsets that result from a given set of predictors. For example, von Eye and Schuster (1998) noted that with 10 predictors, a total of 1024 unique combinations of the predictors is possible. Though this problem has been addressed by some developers of computing software, many software packages that include all possible subsets routines do not really

compute all subsets, but rather give "a few of the best subsets for each number of independent variables" (von Eye & Schuster, 1998, p. 244). If all subsets are truly considered, this method, like commonality analysis, provides a meaningful way to consider variables both in isolation and in sets. One needs to take caution however, that some of the erroneous conclusions common to stepwise procedures do not enter in to interpretation of all possible subsets results.

References

Fox, J. (1991). *Regression diagnostics*. Newbury Park, CA: Sage.

Fox, J. (1997). *Applied regression analysis, Linear models, and related methods*. Thousand Oaks, CA: Sage.

Hamilton, L. C.(1992). *Regression with graphics: A second course in applied statistics*. Belmont, CA: Duxbury.

Huberty, C.J. (1989). Problems with stepwise methods--better alternatives. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 1, pp. 43-70). Greenwich, CT: JAI Press.

Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart, and Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin, 85*, 410-416.

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (1986). *Understanding regression analysis: An introductory guide*. Newbury Park, CA: Sage.

Thompson, B . (1990, April). *Variable contribution in multiple and canonical correlation/regression*. Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED 317 615)

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55,* 525-534.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement, 45*, 203-209.

von Eye, A., & Schuster, C. (1998). *Regression analysis for the social sciences*. San Diego, CA: Academic Press.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM033482

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Multiple Regression: A Leisurely Primer	
Author(s): Larry G. Daniel; Anthony J. Onwuegbuzie	
Corporate Source: Univ. of North Fla; Valdosta State University	Publication Date: 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Larry G. Daniel

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

[X]

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

[ ]

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

[ ]

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Larry G. Daniel	Printed Name/Position/Title: Larry G. Daniel, Associate Dean	
Organization/Address: Univ. of North Florida / COEHS 4567 St. Johns Bluff Rd. Jacksonville, FL 32224-2676	Telephone: 904-620-2520	FAX: 904-620-2522
	E-Mail Address: ldaniel@unf.edu	Date: 2/7/01

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)