DOCUMENT RESUME

ED 464 932		TM 033 862
AUTHOR	Thayer, Jerome D.	
TITLE	Stepwise Regression as an Exploratory Data Analysis Procedure.	
PUB DATE	2002-04-02	
NOTE	39p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).	
PUB TYPE	Numerical/Quantitative Data (110) -- Opinion Papers (120) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	*Data Analysis; Predictor Variables; *Regression (Statistics); *Research Methodology	
IDENTIFIERS	*Stepwise Regression	

ABSTRACT

This paper identifies specific problems with stepwise regression, notes criticisms of stepwise methods by statisticians, suggests appropriate ways in which stepwise procedures can be used, and gives examples of how this can be done. Although the stepwise method has been routinely criticized by statisticians, it is still frequently used in the literature. This paper suggests research situations when stepwise regression may have a valuable function. Stepwise methods can be appropriate for variable evaluation. Since the value of a variable as a predictor is highly specific to the other variables in the prediction model, the use of stepwise methods can provide many reduced models in which the characteristics of the variable can be examined. As variables are found to be good predictors in different models, the different prediction characteristics of the variables in the various models can be used to recognize how the variables function as predictors and can be used to develop a theory or models that can be tested with further research. In order for stepwise methods to be used effectively, they should be used in conjunction with a best subsets procedure and zero-order correlations, default criterion values should be modified, models should not be selected by the computer, and, where possible, models should be generated from multiple subsets of the data. (Contains 21 tables and 17 references.) (SLD)

Stepwise Regression as an Exploratory Data Analysis Procedure

Jerome D. Thayer

Andrews University

A paper presented at the annual meeting of the American Educational Research Association

New Orleans, Louisiana

April 2, 2002

Stepwise Regression as an Exploratory Data Analysis Procedure

To some researchers, stepwise regression is the most attractive application of multiple regression. To many others it is the worst. Researchers with a limited statistical background who conduct research in areas where theory is weak or non-existent are attracted to stepwise regression as a wonderful procedure that, with little or no personal intervention, can find the best combination of explanatory or causal variables. On the other hand, many researchers with a strong statistical background view stepwise regression as a method that seldom, if ever, should be used. There has been little consideration in the literature concerning appropriate uses for stepwise procedures. This paper will identify specific problems with stepwise regression, note criticisms of stepwise methods by statisticians, suggest appropriate ways in which stepwise procedures can be used, and give examples of how this can be done.

Problems with Stepwise Regression

While the stepwise method has been routinely criticized by statisticians, it is still frequently used in the literature. When it is used, it is usually used inappropriately. An examination of articles using stepwise procedures found that they routinely report that the "best model" has been found, or that the beta weights or entry order are interpreted as reflecting the importance of the variables (Thayer, 1990).

An examination of textbooks and journal articles dealing with multiple regression methodology (Thayer, 1990) showed that almost all authors criticized the stepwise method. Examples of general criticisms are:

> Someone has characterized the user of stepwise regression as a person who checks his or her brain at the entrance of the computer center (Wittink, 1988, p. 259).
>
> Stepwise regression is probably the most abused computerized statistical technique ever devised. If you think you need stepwise regression to solve a particular problem you have, it is almost certain that you do not. Professional statisticians rarely use automated stepwise regression (Wilkinson, 1984, p. 196).
>
> I think stepwise methods (e.g., stepwise regression, stepwise descriptive discriminant analysis) are bad, evil, rotten, worthless, and wrong. Plus I do not like them (Thompson, 2001).
>
> The principal problem with stepwise methods is that they take the researcher out of the picture. . . . Stepwise methods are inappropriate within the framework of the scientific method. . . This method requires a hypothesis . . . Stepwise procedures do not fit within this framework (Knapp & Sawilowsky, 2001a).

The extent of the criticism is illustrated by the title of articles and chapters such as "The case against using stepwise research methods" (Davidson, 1988), "Problems with stepwise methods–better alternatives" (Huberty, 1989), "Why won't stepwise methods die?" (Thompson, 1989), "Three reasons why stepwise regression models should not be used by researchers" (Snyder, 1991), and "Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial" (Thompson, 1995).

Multiple Regression Procedures

This paper will suggest research situations when stepwise regression may have a valuable function. To form a context, six regression procedures (simple, simultaneous, hierarchical, forward stepwise, backward stepwise, and best subsets) will be described.

Simple regression evaluates how well a single independent variable predicts a dependent variable. *Simultaneous regression* evaluates how well a pre-specified combination of independent variables predict a dependent variable. *Hierarchical regression* evaluates how one or more independent variables predict a dependent variable in addition to one or more other independent variables; the independent variables are ordered in a hierarchy and entered as predictors in a pre-determined sequence. *Forward stepwise regression* forms a prediction model from the bottom up from a set of independent variables. Variables are entered one at a time, in a series of steps, to build a prediction model. At each step, the computer program automatically adds the variable that would increase the explained percentage of the variance of the dependent variable the most in addition to the previously entered variables. Criteria are set to determine a stopping point for the procedure. Criteria can be set to allow variables to be removed from the model if they no longer meet the criteria. *Backward stepwise regression* builds a prediction model from the top down. Initially, all variables are entered into a prediction model. Then variables are removed one at a time, in a series of steps, to build a prediction model. At each step, the computer program automatically removes the variable that would decrease the explained percentage of the variance of the dependent variable the least. Criteria are set to determine a stopping point for the procedure. Criteria can be set to allow variables to be added to the model if they meet the criteria. *Best subsets regression* or all-possible-subsets regression identifies one or more models models of different sizes that maximize a given criterion. This is done either by examining all possible models or using an algorithm that approximates this procedure.

Exploratory and Confirmatory Multiple Regression

Multiple regression is an appropriate procedure to use to provide information to answer research questions based on either strong theory or weak theory. Frequently an exploration phase is needed to gain an understanding of the data prior to beginning to think about how to model it. In this situation an exploratory phase based on weak theory might be followed by a model selection or model validation phase based on strong theory.

Strong theory research questions are most appropriately answered using simultaneous or hierarchical regression. Information needed to answer weak theory research questions can be gained from using all six regression procedures.

Confirmatory multiple regression procedures (simultaneous or hierarchical) can be used to answer strong theory research questions such as: "Can a specific combination of independent variables predict or explain the variance of a dependent variable?", "Is a specific variable in a given set of independent variables necessary to predict or explain the variance of a dependent variable?", and "Can a specific combination of independent variables predict or explain the variance of a dependent variable in addition to other controlled variables".

Exploratory multiple regression procedures (any of the six methods) can be used to answer weak theory research questions such as: "How do variables work in combination to predict a dependent variable?", "What independent variables should be considered to be good predictors of a dependent variable?", and "How do independent variables predict a dependent variable in various circumstances?" This paper will describe how stepwise regression can be used in conjunction with other regression procedures in exploratory research to answer weak theory research questions.

Weak theory research questions where multiple regression can be used generally specify a set of independent variables that have potential value in predicting or explaining variability in a dependent variable. The independent variables to be considered typically includes many variables that have only weak theory to support their consideration or it is known that some of the variables are good predictors but the nature of the intercorrelations confuse which will be the more stable predictors or what the causal relationships might be. The purpose of exploratory regression should not be to find a "best" model or to find out what variables are the "best" predictors, but to provide information that be used to understand the relationship between the variables to allow a specific hypothesis or theory to be constructed which can be confirmed with later research.

In a situation where a set of independent variables are hypothesized to be related to a dependent variable (by weak or strong theory), the relationship between each independent variable and the dependent variable can be examined in many circumstances. The simplest relationships to examine are those with each independent variable as a single predictor (zero-order correlation). The most complex relationships would be to examine each independent variable as a predictor in the presence of all other independent variables being considered (simultaneous regression). It would also be helpful to examine the relationship of each independent variable with the dependent variable in reduced (smaller) models. Stepwise procedures can be useful in providing a variety of reduced models that can be examined. This use of stepwise regression has support in the literature.

> Algorithms such as stepwise regression analysis should be reserved for situations where the research is entirely exploratory, and where the researcher has extreme difficulty justifying any model specification prior to data analysis (Wittink, 1988, p. 259).

> Although some researchers suggest that variable selection procedures are always useful (i.e., the number crunchers) or they are never useful (e.g., the statistical purists), my personal philosophy lies somewhere in between. I believe that the best situation to use these procedures is in exploratory research where prior research and theory are weak or lacking (Lomax, 2001, p. 258-259).

> But, does this mean that stepwise methods are worthless? No; here are two possible roles for such analyses: 1. A great deal of emphasis has recently been placed on 'data mining.' . . . Stepwise methods can be useful as mining tools (Knapp & Sawilowsky, 2001a).

> The desirable exuberance in pointing out that stepwise methods are useless in hierarchical analysis, theory building, or the testing of theory has little to do with data mining or the construction of predictive equations that capitalize on nontheory-based R^2s. Although neither of us supports these practices, we do not extend our disdain of stepwise methods to nonmodel-based exploratory applications. (Knapp & Sawilowsky, 2001b).

Regression can be used for hypothesis testing, model building or variable evaluation. While it appears that stepwise regression could be used for model building (it is frequently called a model-selection procedure in textbooks), it has two major problems. First, it cannot be used to confirm whether a given model is good and second, the model selected by the computer is frequently not the model with the highest R^2. Thayer (1990) gives an example of a data set in which forward stepwise and backward stepwise methods were compared to the best subsets method. The models selected by the computer for the 2, 4, 5, 6, and 7 predictor cases were different for the three methods. Thayer (1986) gives examples of 7 data sets in which the three methods give different models. He concludes that "it is recommended that the stepwise . . . methods NEVER be used alone in selecting a model for any purpose."

However, stepwise methods are appropriate for variable evaluation. Since the value of a variable as a predictor is highly specific to the other variables in the prediction model, the use of stepwise methods can provide many reduced models in which the characteristics of the variable can be examined. As variables are found to be good predictors in different models, the different prediction characteristics of the variables in the various models can be used to recognize how the variables function as predictors and can be used to develop a theory or models that can be tested with further research.

Proposed Use for Stepwise Regression

Ideally, it would be helpful to examine the relationship between each independent variable (IV) and the dependent variable (DV) in every possible combination of predictors. When there are more than just a few predictors in the data set this is not feasible, and obviously there comes a point of diminishing returns as many models are examined. The procedure recommended in this paper is to compare many models of four types of combinations of predictors and to examine how each variable functions differently in the models to understand the value of the variable as a predictor.

Initially it is helpful to examine the relationship between the DV and each IV alone. This usually provides the largest estimate of the value of an IV since the variable can be expected to claim any explanatory value it shares with other predictors. When other variables have a causal effect on both the IV and the DV, the zero-order correlation between the IV of interest and the DV overestimates their true relationship. In cases where suppression exists, the zero-order correlation may actually underestimate the true value of the relationship.

Next it is helpful to examine the relationship between the DV and all IVs together in a simultaneous regression analysis. This allows the researcher to examine the unique contribution of each IV when controlled for all of the other predictors. This contribution would be measured by a squared part (or semi-partial) correlation which would give the proportion of the variance of the DV accounted for by the IV in addition to all the other IVs in the regression model. Normally, the percent of variance contributed in addition to the other variables would be less than the variance contributed when the IV is considered alone. However, when suppression exists, the part correlation may be larger than the zero-order correlation.

Models with smaller numbers of predictors (reduced models) can be generated by using the forward stepwise, backward stepwise, and best subsets (or all-possible-subsets) procedures. The models used can be the different intermediate models at each step in the stepwise procedures or one or more models generated through each of many stepwise regression analyses with multiple subsets of the data.

Thayer (1986) identified different problems for the forward stepwise and backward stepwise methods. The forward stepwise procedure misses many good models, particularly if variables are only good predictors when combined with certain other variables. He presented one set of data in which the backward stepwise and best subsets methods selected a two-predictor model with an $R^2 = .967$ while the forward stepwise method indicated that none of the models met the stepping criteria. The two variables only were valuable as predictors when acting in combination.

The backward stepwise method frequently gives models that are larger than necessary. The best subsets method will always identify good models, but occasionally will miss a good model. It also is more difficult to use for comparing models and evaluating how each variable functions in the model.

The forward stepwise procedure is best run by changing the default entry criterion to a high p value such as p=.25 to allow for models with more predictors to be considered. The backward stepwise procedure is best run by changing the default removal criterion to a low p value such as p=.0001 to allow models with fewer predictors to be considered. Since many of the problems with forward and backward stepwise procedures are unique to that method, using both methods with modified entry/removal criteria helps to minimize the probability of missing important information that might be true if fewer models were considered. Also if a more relaxed criterion is used for stopping the stepping procedure, a better model is frequently found. For example, using stepwise regression with a classic data set with four predictors (Hald, 1952) would result in a 2-predictor model using either PIN =.10 or PIN=.05, but the model found with PIN=.10 is a different and better model (higher R^2). The maximum number of stepwise models could be considered if the p value for entry was set to .999999 for forward stepwise (almost all variables would be added) and the p value for removal was set to .0000001 for backward stepwise (almost all variables would be eliminated.

For example, with relaxed entry criteria, with a set of 20 potentially valuable independent variables you could get information about each IV from at least 20 different models. Since in many situations forward and backward stepwise procedures give different models of the same size, you might get more than 20 different models. Realistically, a smaller number of models would be examined as the stepping criteria would be set so that there were no forward stepwise models with most of the predictors and no backward stepwise models with very few predictors.

In following the approach suggested in this paper, you would compare the way each variable functions alone (zero-order correlation) and in each model (examine one or more of the following statistics: beta, part correlation, and tolerance). More importantly, you could see the change in these statistics for each variable in the steps of the stepwise procedure as other variables are added or deleted from the models. This valuable diagnostic function of the stepwise method is an exploratory data procedure.

If statistics from the most informative reduced models selected from the forward and backward stepwise procedures are compared to statistics from the simultaneous (all IVs) model and the zero-order correlations, rich information about each of the variables can be gained. When unusual statistics or patterns of statistics are found, the researcher should try to determine the reasons for them before deciding on the value of the variables being considered. Unusual statistics or patterns of statistics would include:

multicollinearity statistics for each IV changing as models change
betas or part correlations getting larger with larger models
betas or part correlations having different signs in different models
models of the same size with different IVs
IVs with a high zero-order correlation that are not found in larger models
IVs with a low zero-order correlation that are found in larger models

The purpose of exploratory analysis using stepwise regression along with other regression procedures would be to understand each IV, not to select good IVs or to select a good model.

A suggested sequence of steps would be:

1) Identify appropriate variables
2) Produce different models (combinations of variables)
    alone
    in reduced subsets of good predictors
        selected by different methods
            forward stepwise (with modified default values)
            backward stepwise (with modified default values)
            best subsets
    with all good predictors
    with all predictors
3) Compare the relevant statistics for each predictor in each model
    zero-order correlations, betas, and part correlations
4) Determine which variables are worthy of consideration in future research

<u>Examples</u>

To illustrate this procedure, 3 data sets were studied. The data sets are described in Table A. They varied widely in the number of subjects, the number of predictors, the type of variables, the degree of multicollinearity, and the value of the predictors. Two of the data sets were large and one was small. Data set #3 had high multicollinearity, data set #2 had moderate multicollinearity, and data set #1 had relatively low multicollinearity. Most predictors in data set #2 were good predictors by themselves while most predictors in data set #3 were poor predictors. Data Set #1 had some poor predictors and some good predictors.

Insert Table A about here

The following steps were used with these three data sets to illustrate how stepwise regression could be used for exploratory purposes.

1) Reduced models were identified for each data set using the forward stepwise, backward stepwise, and best subsets methods. SPSS 11.0 was used for the stepwise methods and BMDP9R was used for the best subsets method.

2) The models from the forward stepwise and backward stepwise methods were not selected automatically by the computer program. The procedures used for each method were:

forward stepwise method

PIN = .10 (models were not selected using this criterion – this setting allowed the computer program to present larger models than the default setting of PIN = .05 – one or more models were selected by the researcher using other criteria listed below)

models were chosen by the researcher if they would have been selected automatically by the computer using PIN=.01

models were chosen by the researcher if the next variable to be added by the computer increased the total R^2 by less than .01 (add an additional percent of the explained variance of Y of less than 1%)

backward stepwise method

POUT = .001 (models were not selected using this criterion – this setting allowed the computer program to present additional smaller models than the default setting of POUT=.10 – one or more models were selected by the researcher using other criteria listed below)

models were chosen by the researcher if they would have been selected automatically by the computer using POUT=.01

models were chosen by the researcher if the next variable to be removed by the computer decreased the total R^2 by more than .01 (remove an additional percent of the explained variance of Y by more than 1%)

best subsets method

the model chosen was the computer-selected model using the default criterion of minimizing the C_p value

The criteria values of p=.01 and R^2 change=.01 gave models of approximately equal size with these data sets. If different criteria values had were used, the model sizes would have been different. Since the purpose of these analyses was not to identify models but to examine variables, the criterion values were chosen to give models with enough variables in them to give good diagnostic information about many variables.

The three data sets were treated as population data. Multiple samples of two sizes (N=500 and N=5,000) were used to produce models. Eight randomly selected samples of 500 subjects and eight randomly selected samples of 5,000 subjects were selected both from data set #1 and from data set #2. The 3 model-selection procedures (forward stepwise, backward stepwise, and best subsets) were used with each of the 16 samples for each data set producing 48 different analyses for each data set (24 for each sample size).

These sample sizes were used to approximate or go beyond guidelines commonly recommended for reliable use of stepwise methods to produce a good model. For data set #1 (40 independent variables), the sample sizes of 500 and 5,000 resulted in sample size/number of independent variable ratios of 12.5/1 and 125/1. The ratio of 12.5/1 is at the lower range of recommendations and 125/1 is higher than most recommendations. For data set #2 (19 independent variables), the sample sizes resulted in sample size/number of independent variable ratios of 26.3/1 and 263/1.

Data set #3 was not large enough to be subdivided into smaller samples and since it was so small (N=50), it was artificially increased in size to make the sample size equal to 300. One of the criteria used in this study for the stepwise procedures was to use a select models that contained variables all of which had a significance in the model of .01 or less. In order to use this criterion and have models of comparable size for all three data sets, data set #3 was modified by replicating the data 5 additional times to make the N=300. All statistical information for this data other than the significance level of the predictors was not changed by this modification. The same results could have been accomplished by changing the significance level from .01 to a higher value that would give models of approximately the same size as the other data sets.

Using two criteria of p=.01 and R^2 change=.01 resulted in selecting more than one model for some methods and using three different methods (forward stepwise, backward stepwise, and best subsets) sometimes produced different models for the same data set. Six unique models were identified for data set 3 with one sample of data, 16 unique models were identified for data set #2 with 16 samples, and 22 unique models were identified for data set #3 with 16 samples. Table B describes the number of models identified by these procedures.

Insert Table B about here

Data set #1 had 40 independent variables. Eighteen of these variables appeared in at least one of the 22 unique models. The models ranged in size from 5-7 predictors.

Data set #2 had 19 independent variables. Fourteen of these variables appeared in at least one of the 16 unique models. The models ranged in size from 4-5 predictors.

Data set #3 had 14 independent variables. Twelve of these variables appeared in at least one of the 6 unique models. The models ranged in size from 8-10 predictors.

In the 33 samples of data (16 from data set #1, 16 from data set #2, and 1 from data set #3), the forward stepwise, backward stepwise and best subsets methods identified the same model in only 18 samples. The forward stepwise and backward stepwise methods agreed in 19 samples, the forward stepwise and best subsets methods agreed in 29 samples, and the backward stepwise and best subsets methods agreed in 18 samples.

The three methods were run using the population data for data set #1 and data set #2. The three methods agreed on the model for data set #2 but the model identified by the backward stepwise method in data set #1 differed on one variable from the models identified by the forward stepwise and best subsets methods.

The models identified from the population data were found in 11 of the 24 runs of N=5,000 with data set #1, in 9 of the 24 runs of N=5,000 with data set #2, and in none of the 48 runs of N=500 with data sets #1 and #2. A description of the models are found in Tables C-G.

Insert Tables C-G about here

---

3) Statistics were computed from population data for each predictor found in one of the selected models. While there is no standard statistic that determines the value of a predictor in a regression equation, Thayer (1991) suggests three alternatives: standardized regression coefficients (betas), part correlations (semi-partial correlations), and the product of beta and the zero-order correlation. Each statistic provides different information. Each beta shows how much the dependent variable would change for a one-standard deviation change in the independent variable. Each part correlation, when squared, indicates the unique contribution of the independent variable to the R^2 of the model. The product of each beta and the corresponding zero-order correlation gives the contribution of the predictor to the R^2 of the total model (R^2 = the sum of the products of each beta and the corresponding zero-order correlation).

When there is no suppression, the beta and part correlations are highly correlated (Thayer, 1991). However, this is not true when there is suppression. Table H shows statistics from two small data sets to illustrate how different and unreliable these statistics can be in a suppression situation. The two data sets have 3 predictors and 6 subjects, with two of the predictors being highly correlated. By changing one data point the relative values of the betas changed markedly while the relative values of the part correlations changed very little. In some circumstances the betas might provide better information, whereas in other circumstances the part correlation or the product of beta and the zero-order correlation might be superior.

Insert Table H about here

For this research, part correlations were used to understand the predictive value of each predictor in the models with many predictors. Table I is a matrix showing the relationship between these three statistics for the models that included the best variables from the three data sets examined in this paper. The correlation between the betas and the part correlations ranged from .991 to .997 for the three data sets.

Insert Table I about here

Once the models had been selected, statistics were compiled for the variables in each model. Each unique model was re-run using population data to get more comparable statistics. Similar results would have been found if statistics based on the samples had been used. Statistics were reported from 3 types of models: simple models (each predictor alone), many reduced models selected from one of the three selection methods (forward stepwise, backward stepwise, and best subsets) and two simultaneous models, one model composed of all predictors that appeared in any of the models identified by the three methods and one model composed of all predictors. The statistics for these models are reported in Tables J-L.

Insert Tables J-L about here

----

4) Each of the predictors was classified subjectively as "good," "fair," "questionable," or "poor" based on the number of models in which the variable appeared and the statistics associated with the variable in each of the models. Good variables appeared in most of the models with good part correlations while poor variables appeared in few models with poor part correlations. Fair and questionable variables were in some of the models with varying quality of part correlations. A description of the classification of the predictors is found Tables M-O.

Insert Tables M-O about here

----

5) The highest rated predictors were combined into a model of approximately the same size as the models identified by the 3 methods. The model composed of these "best" predictors was found in 10 of the 16 samples of N=5,000 for data sets #1 and #2 and none of the 16 samples of N=500. The two models containing the "best" predictors were identified by both the forward stepwise and best subsets method using the population data for data sets #1 and #2 but by the backward stepwise method only for data set #1. The model of "best" predictors was not found by any of the methods in data set #3. The models are reported in Table P.

Insert Table P about here
----

6) The predictors were evaluated without using stepwise or best subsets methods to see if the "best" variables could be identified. The top 10 variables in terms of zero-order correlation and part correlation in the simultaneous model with all predictors were identified. Variables that were one of the top 10 predictors both alone and together were identified. Five of the "best" variables in the 3 data sets were not identified using just the zero-order and part correlations in the simultaneous model, and 9 variables identified by the zero-order and part correlations in the simultaneous model were not on the "best" predictor list. These evaluations are reported in Tables Q-S.

Insert Tables Q-S about here
----

7) A principle components factor analysis with varimax rotation was conducted with data sets #1 and #2 to try to explain which types of variables were being selected in the reduced models. For data set #1, 11 factors were identified with eigenvalues > 1.00. There were no predictors in any of the models from 4 of the factors, 1 predictor in each identified model from 5 factors, and multiple predictors in each model from 2 factors. For data set #2, 2 factors were identified (oblimin rotation was also used with the same results). There were many predictors in each model from 1 factor and 0-1 predictor in each model from the other factor. An examination of the factor structure of the data set did not help to explain why different combinations of predictors appeared in the models selected. This information is in Tables T-U.

Insert Tables T-U about here

<u>Conclusions</u>

Stepwise methods are useful in identifying variables that are good predictors in reduced models. In order for stepwise methods to be used effectively, they should used in conjunction with a best subsets procedure and zero-order correlations, default criterion values should be modified, models should not be selected by the computer, and, where possible, models should be generated from multiple subsets of the data.

Independent variables that are likely to be good variables in predictive or explanatory models can be identified by comparing betas and/or part correlations from multiple models including single-predictor models, reduced models from the stepwise and best subsets methods, and simultaneous models using all good predictors and/or all predictors. These "good" variables should be combined to form predictive or explanatory models based on information provided with this analysis and theoretical considerations. Models formed with these variables would need to be cross-validation with other data or subjected to confirmatory analysis.

References

Davidson, B. M. (1988). *The case against using stepwise research methods.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED303507)

Gunst, R. F. & Mason, R. L. (1980). *Regression analysis and its application.* New York: Marcel Dekker.

Hald, A. (1952). *Statistical Theory with Engineering Applications.* New York, NY: Wiley.

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.

Knapp, T. R. & Sawilowsky, S. S. (2001a). Constructive criticisms of methodological and editorial practices. *Journal of Experimental Education, 70(1),* 65-79.

Knapp, T. R. & Sawilowsky, S. S. (2001b). Strong arguments: Rejoinder to Thompson. *Journal of Experimental Education, 70(1),* 94-95.

Lomax, R. G. (2001). *An introduction to statistical concepts for education and behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Snyder, P. (1991). Three reasons why stepwise regression models should not be used by researchers. In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments,* (Vol 1, pp. 99-105). Greenwich, CT: JAI Press.

Thayer, J. D. (1986). Testing different model building procedures using multiple regression. *Multiple Linear Regression Viewpoints, 15(1),* 37-52.

Thayer, J. D. (1990, April). *Implementing variable selection techniques in regression.* Paper presented at the annual meeting of the American Education Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED319805)

Thayer, J. D. (1991, April). *Interpretation of standardized regression coefficients in multiple regression.* Paper presented at the annual meeting of the American Education Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED334208)

Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development, 21,* 146-148.

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. (1999). Five methodology errors in educational research: A pantheon of statistical significance and other faux pas. In B. Thompson (Ed.), *Advances in social science methodology,* (Vol 5, pp. 23-86). Greenwich, CT: JAI Press.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: strong arguments move the field. *Journal of Experimental Education, 70(1),* 80-93.

Wilkinson, L (1984). *SYSTAT.* Evanston, IL: SYSTAT.

Wittink, D. R. (1988). *The application of regression analysis.* Needham Heights, MA: Allyn and Bacon.

Table A

Description of Data Sets

					Tolerance		Zero-order correlations (absolute values)	
Data Set	Source of Data	Subjects	Dependent variable	Independent variables	Range	Median	Range	Median
1	Nationwide values study	13,000+ elementary & secondary students	Vertical faith maturity	40 values and home, school, and church characteristics	.747	.458 - .913	.236	.014 - .618
2	Student ratings from a university	65,000+ university students	Overall rating of instructor's teaching effectiveness	19 course and instructor characteristics	.428	.372 - .597	.592	.477 - .718
3	Sample data set in a statistics textbook[a]	50 police department applicants	Reaction time	14 anthropometric and physical fitness measurements	.301	.059 - .715	.138	.032 - .222

[a]A6 data set from Gunst and Mason (1980)

Table B

Number of Models Produced

Data Set	Number of samples	Number of subjects	Number of runs[a]	Number of different models[b]	Number of unique models[c]
1	8	500	24	16	16
	8	5,000	24	13	6
2	8	500	24	10	10
	8	5,000	24	10	6
3	1	300	3	6	6

[a] Models were selected from each sample using three methods:
forward stepwise
backward stepwise
best subsets

[b] Different models were those with different statistical information because of being different predictors from the same sample or the same predictors from different samples

[c] Unique models were those with different predictors disregarding the sample from which they came

Table C

Models from Data Set #1 (Sample N's = 500)

Dependent Variable = 12
24 runs with 8 samples (8 forward stepwise, 8 backward stepwise, 8 best subsets) – 16 models produced (16 unique models)
16 models sorted by R^2

R^2	Method[a]	NIV[b]	1	3	6	7	8	11	14	16	17	19	21	28	30	31	32	35	36	39
			Variables																	
.668	F^7 B^7 S^7	7		*				*					*		*	*	*	*		
.653	B^{4a}	7	*	*			*	*			*				*		*			
.645	F^4 S^4	6		*			*	*						*	*		*			
.640	F^2 S^2	5		*				*		*					*					*
.637	B^2	5		*			*	*							*	*				
.633	B^{4b}	5		*			*	*							*		*			
.616	S^5	5		*			*	*									*			*
.616	F^5 B^5	5		*				*							*		*			*
.616	B^1	8	*	*		*	*	*	*	*										*
.605	F^1	7	*	*		*	*	*							*					*
.604	F^8 B^8 S^8	6	*	*			*	*		*						*				
.604	S^1	7	*	*			*	*							*		*			*
.589	F^3 B^3 S^3	7	*	*			*	*						*	*	*				
.566	B^6	7		*	*		*	*			*	*							*	
.557	S^6	6		*	*		*	*			*								*	
.546	F^6	5		*			*	*						*					*	
N			6	16	2	2	13	16	1	3	2	1	1	3	10	4	7	1	3	6

[a] Superscript indicates sample
F = Forward Stepwise, B = Backward Stepwise, S = Best Subsets
[b] NIV = Number of Independent Variables

16 unique models sorted by variables included

R^2	Method			NIV	1	3	6	7	8	11	14	16	17	19	21	28	30	31	32	35	36	39
					Variables																	
.616		B		8	*	*		*	*	*	*	*										*
.605	F			7	*	*		*	*	*							*					*
.653		B		7	*	*			*	*			*				*		*			
.589	F	B		7	*	*			*	*						*	*	*				
.604			S	7	*	*			*	*							*		*			*
.566		B		7		*	*		*	*			*	*							*	
.668	F	B		7		*				*					*		*	*	*	*		
.604	F	B		6	*	*			*	*		*						*				
.645	F			6		*			*	*						*	*		*			
.557			S	6		*	*		*	*			*								*	
.546	F			5		*			*	*						*					*	
.637		B		5		*			*	*							*	*				
.633		B		5		*			*	*							*		*			
.616			S	5		*			*	*									*			*
.640	F			5		*				*		*					*					*
.616	F	B		5		*				*							*		*			*

Model based on population data:	R^2	Variables					
Forward stepwise	.603	1	3	8	11	31	32
Backward stepwise	.603	1	3	8	11	28	31
Best subsets	.603	1	3	8	11	31	32

Table D

Models from Data Set #1 (Sample N's = 5,000)

Dependent Variable = 12
24 runs with 8 samples (8 forward stepwise, 8 backward stepwise, 8 best subsets) – 13 models produced (6 unique models)
13 models sorted by R^2

R^2	Method[a]			NIV[b]	Variables 1	3	6	7	8	11	14	16	17	19	21	28	30	31	32	35	36	39
.612[p]	F^7		S^7	6	*	*			*	*								*	*			
.612		B^7		6	*	*			*	*								*			*	
.611[p]	F^5		S^5	6	*	*			*	*								*	*			
.609		B^5		6	*	*			*	*						*		*				
.607[p]	F^4		S^4	6	*	*			*	*								*	*			
.606		B^4		6	*	*			*	*						*		*				
.605[p]	F^1	B^1	S^1	6	*	*			*	*								*	*			
.601	F^3		S^3	6	*	*			*	*								*			*	
.599[p]	F^2		S^2	6	*	*			*	*								*	*			
.599	F^8	B^8	S^8	6	*	*	*		*	*								*				
.596	F^6	B^6	S^6	6		*	*		*	*		*						*				
.591		B^3		5	*	*			*	*								*				
.589		B^2		5	*	*			*	*								*				
N					12	13	2	0	13	13	0	1	0	0		2	0	13	5	0	2	0

[a] Superscript indicates sample
F = Forward Stepwise, B = Backward Stepwise, S = Best Subsets
[b] NIV = Number of Independent Variables
[p] Model found with population data

6 unique models (a-f) sorted by variables included

R^2	Method			NIV	Variables 1	3	6	7	8	11	14	16	17	19	21	28	30	31	32	35	36	39
.599[b]	F	B	S	6	*	*	*		*	*								*				
.609[c]		B		6	*	*			*	*						*		*				
.606[c]		B		6	*	*			*	*						*		*				
.612[d]	F		S	6	*	*			*	*								*	*			
.611[d]	F		S	6	*	*			*	*								*	*			
.607[d]	F		S	6	*	*			*	*								*	*			
.605[d]	F	B	S	6	*	*			*	*								*	*			
.599[d]	F		S	6	*	*			*	*								*	*			
.612[e]		B		6	*	*			*	*								*			*	
.601[e]	F		S	6	*	*			*	*								*			*	
.596[f]	F	B	S	6		*	*		*	*		*						*				
.591[a]		B		5	*	*			*	*								*				
.589[a]		B		5	*	*			*	*								*				

Model based on population data:

	R^2	Variables					
Forward stepwise	.603	1	3	8	11	31	32
Backward stepwise	.603	1	3	8	11	28	31
Best subsets	.603	1	3	8	11	31	32

Table E

Models from Data Set #2 (Sample N's = 500)

Dependent variable = 1
24 runs with 8 samples (8 forward stepwise, 8 backward stepwise, and 8 best subsets) – 10 models produced (10 unique)
10 models sorted by R^2

R^2	Method[a]	NIV[b]	4	5	6	7	8	9	10	14	15	16	17	18	19	20
.677	F^7 S^7	5		*				*			*	*				*
.677	B^7	5		*					*		*	*				*
.676	F^4 B^4 S^4	5	*	*					*						*	*
.671	F^8 B^8 S^8	5								*	*			*	*	*
.670	F^2 B^2 S^2	5			*		*			*	*					*
.656	F^1 B^1 S^1	4		*									*		*	*
.647	F^5 S^5	5						*	*	*					*	*
.646	B^5	5		*					*	*	*					*
.625	F^3 B^3 S^3	4		*										*	*	*
.620	F^6 B^6 S^6	5		*		*								*	*	*
N			1	7	1	1	1	2	4	4	5	2	1	3	6	10

[a] Superscript indicates sample
F = Forward Stepwise, B = Backward Stepwise, S = Best Subsets
[b] NIV = Number of Independent Variables

10 unique models sorted by variables included

R^2	Method	NIV	4	5	6	7	8	9	10	14	15	16	17	18	19	20
.676	F B S	5	*	*					*						*	*
.620	F B S	5		*		*								*	*	*
.677	F S	5		*				*			*	*				*
.646	B	5		*					*	*	*					*
.677	B	5		*					*		*	*				*
.670	F B S	5			*		*			*	*					*
.647	F S	5						*	*	*					*	*
.671	F B S	5								*	*			*	*	*
.656	F B S	4		*									*		*	*
.625	F B S	4		*										*	*	*

Model based on population data:	R^2	Variables				
Forward stepwise	.649	5	10	15	19	20
Backward stepwise	.649	5	10	15	19	20
Best subsets	.649	5	10	15	19	20

Table F

Models from Data Set #2 (Sample N's = 5,000)

Dependent variable = 1
24 runs with 8 samples (8 forward stepwise, 8 backward stepwise, and 8 best subsets) – 10 models produced (6 unique)
10 models sorted by R^2

R^2	Method[a]	NIV[b]	4	5	6	7	8	9	10	14	15	16	17	18	19	20
$.664^p$	F^1 B^1 S^1	5		*					*		*				*	*
.664	F^4 S^4	5				*			*		*				*	*
$.664^p$	B^4	5		*					*		*				*	*
.651	F^7 B^7 S^7	4							*		*				*	*
$.651^{p**}$	B^6	5		*					*		*				*	*
.650	F^6 S^6	5						*	*		*				*	*
.641	F^8 B^8 S^8	4		*					*						*	*
.640	F^2 B^2 S^2	5		*					*	*					*	*
.632	F^3 B^3 S^3	4		*					*						*	*
$.627^p$	F^5 B^5 S^5	5		*					*		*				*	*
N			0	7	0	1	0	1	10	1	7	0	0	0	10	10

* = Superscript indicates sample
** = not one of the top 10 best subsets with 5 predictors
p Model found with population data

6 unique models (a-f) sorted by variables included

R^2	Method	NIV	4	5	6	7	8	9	10	14	15	16	17	18	19	20
$.640^a$	F B S	5		*					*	*					*	*
$.664^b$	F B S	5		*					*		*				*	*
$.664^b$	B	5		*					*		*				*	*
$.651^b$	B	5		*					*		*				*	*
$.627^b$	F B S	5		*					*		*				*	*
$.664^c$	F S	5				*			*		*				*	*
$.650^d$	F S	5						*	*		*				*	*
$.641^e$	F B S	4		*					*						*	*
$.632^e$	F B S	4		*					*						*	*
$.651^f$	F B S	4							*		*				*	*

Model based on population data:	R^2	Variables				
Forward stepwise	.649	5	10	15	19	20
Backward stepwise	.649	5	10	15	19	20
Best subsets	.649	5	10	15	19	20

Table G

Models from Data Set #3 (N = 300)

Dependent Variable = 1
1 run with 1 sample (8 forward stepwise, 8 backward stepwise, 8 best subsets) – 6 models produced (6 unique models)
Sorted by R^2

R^2	Method[a]			NIV[b]	2	4	5	6	7	8	9	10	11	12	13	15
.388		B	S	10	*		*	*	*		*	*	*	*	*	*
.378	F		S	9	*		*	*	*	*	*			*	*	*
.370		B	S	8	*		*	*	*		*			*	*	*
.363	F			9	*	*	*	*		*	*			*	*	*
.332	F			8	*	*	*			*	*			*	*	*
.318	F			8	*	*	*			*	*	*		*	*	
N					6	3	6	4	3	4	6	2	1	6	6	5

[a] F = Forward Stepwise, B = Backward Stepwise, S = Best Subsets
[b] NIV = Number of Independent Variables

6 unique models sorted by variables included

R^2	Method			NIV	2	4	5	6	7	8	9	10	11	12	13	15
.388		B	S	10	*		*	*	*		*	*	*	*	*	*
.363	F			9	*	*	*	*		*	*			*	*	*
.378	F		S	9	*		*	*	*	*	*			*	*	*
.318	F			8	*	*	*			*	*	*		*	*	
.332	F			8	*	*	*			*	*			*	*	*
.370		B	S	8	*		*	*	*		*			*	*	*

Table H

Comparison of Beta and Part Correlations

	One-predictor model statistics					Three-predictor model statistics						
Data	r_{Y1}[a]	r_{Y2}	r_{Y3}	r_{12}[b]	r_{23}	R_{Y123}[c]	β_1[d]	β_2	β_3	$r_{Y(1.23)}$[e]	$r_{Y(2.13)}$	$r_{Y(3.12)}$
Y X1 X2 X3 1 6 60 2 2 2 19 3 3 3 29 3 4 10 98 2 5 12 **90** 4 6 10 100 5	.740	.722	.777	.962	.367	.926	.103	.432	.603	.027	.117	.555
Y X1 X2 X3 1 6 60 2 2 2 19 3 3 3 29 3 4 10 98 2 5 12 **110** 4 6 10 100 5	.740	.740	.777	.997	.367	.920	-.298	.825	.589	-.024	.066	.547

*Numbers in bold are the only numbers different in the two data sets.

[a] zero-order correlation between Y and X_1
[b] intercorrelation between X_1 and X_2
[c] multiple correlation between Y and X_1, X_2, and X_3
[d] part correlation between Y and X_1, controlled for X_2 and X_3

Table I

Comparison of Standardized Coefficients (betas) and Part Correlations

Data Set #1 (R^2 = .603)					Data Set #2 (R^2 = .649)					Data Set #3 (R^2 = .360)				
IV	r	β	Part	rβ	IV	r	β	Part	rβ	IV	r	β	Part	rβ
V1	.482	.133	.114	.064	V5	.618	.146	.107	.090	V2	.222	.564	.407	.125
V3	.618	.311	.252	.192	V10	.664	.176	.120	.117	V5	-.056	-.430	-.288	.024
V8	.468	.189	.164	.088	V15	.654	.154	.100	.101	V6	-.032	-.367	-.227	.012
V11	.547	.311	.278	.170	V19	.673	.178	.113	.120	V8	.163	.147	.115	.024
V31	.302	.120	.115	.036	V20	.718	.308	.205	.221	V9	.147	.224	.193	.033
V32	.438	.119	.104	.052						V12	-.076	-.420	-.320	.032
										V13	-.149	-.317	-.250	.047
										V15	.165	.383	.254	.063

Correlations:

Data Set	β - Part	β - rβ	Part - rβ
1	.993	.988	.964
2	.991	.989	.986
3	.997	.716	.687

Table J

Zero-order & Part Correlations from Data Set #1
24 Models (22 Forward Stepwise/Backward Stepwise/Best Subsets Models and 2 Simultaneous Models)
Statistics Based on Population Data (N = 13,103)

		Number of Predictors																								
IV	No of Models	1	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	8	18	40
1	11	.482	.119							.108	.121	.114	.109		.115			.118	.125	.116			.115	.119	.105	.117
3	22	.618	.272	.302	.306	.293	.333	.348	.307	.261	.266	.252	.254	.279	.250	.289	.290	.261	.238	.252	.287	.320	.245	.255	.212	.202
6	4	.423								.096				.112			.104				.104				.048	.051
7	2	.273																.062						.059	-.006	-.001
8	19	.468	.202	.169	.184	.156			.164	.177	.179	.164	.184	.119	.146	.140	.149	.161	.138	.162	.148		.150	.131	.092	.079
11	22	.547	.287	.321	.326	.327	.338	.327	.336	.269	.284	.278	.271	.304	.285	.325	.304	.288	.281	.278	.304	.311	.283	.286	.252	.211
14	1	.109																						-.056	-.031	-.032
16	4	.442					.142							.094	.085									.100	.062	.053
17	3	-.232															-.074		-.073		-.074				-.043	-.030
19	1	.236																			-.010				-.025	-.019
21	1	.148																				.033			.005	.005
28	4	.301		.114							.101					.101				.095					.046	.046
30	10	.350			.098	.088	.102	.110								.083		.092	.076	.085		.098	.082		.058	.057
31	10	.302	.131		.134					.121	.108	.115	.121	.121	.126					.105		.118			.066	.065
32	8	.438				.117		.147	.114			.104				.100			.104			.121	.096		.030	.023
35	1	.335																				.089			.040	.035
36	4	.424		.106									.093				.093				.093				.049	.048
39	6	.195					.117	.108	.100									.106					.098	.101	.034	.033
R^2			.593	.586	.588	.584	.570	.571	.586	.602	.603	.603	.601	.599	.600	.594	.589	.601	.602	.610	.589	.584	.607	.603	.634	.643

Table K

Zero-order & Part Correlations from Data Set #2
18 Models (16 Forward Stepwise/Backward Stepwise/Best Subsets Models and 2 Simultaneous Models)
Statistics Based on Population Data (N = 65,535)

IV	No of Models	Number of Predictors																		
		1	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	14	19
4	1	.503					.035												.008	.005
5	10	.618	.119	.150	.145		.101	.113	.098	.112	.109	.110	.107						.056	.053
6	1	.599												.093					.013	.010
7	2	.592						.084							.078				.022	.022
8	1	.545												.073					.030	.029
9	3	.602							.067							.090	.092		.018	.014
10	10	.664	.138			.155	.132			.116	.117	.127	.120		.123	.120	.121		.072	.070
14	5	.644								.085	.091			.086		.090		.094	.049	.044
15	9	.654				.113			.173	.147		.149	.100	.168	.104		.102	.107	.071	.071
16	2	.535							.068			.067							.018	.012
17	1	.526		.086															.006	.006
18	3	.559			.113			.102										.102	.052	.051
19	12	.673	.181	.186	.185	.126	.178	.170			.159		.113		.117	.165	.116	.118	.084	.083
20	16	.718	.221	.267	.254	.233	.216	.230	.270	.197	.179	.228	.205	.248	.214	.186	.210	.224	.158	.147
r^2			.639	.628	.633	.638	.641	.640	.630	.644	.648	.641	.649	.628	.644	.644	.646	.637	.665	.666

Table L

Zero-order & Part Correlations from Data Set #3
8 Models (6 Forward Stepwise/Backward Stepwise/Best Subsets Models and 2 Simultaneous Models)
Statistics Based on Population Data (N = 300)

IV	No of Models	Number of Predictors								
		1	8	8	8	9	9	10	12	14
2	6	.222	.410	.412	.431	.399	.412	.353	.304	.308
4	3	-.094	-.160	-.151		-.053			.031	.043
5	6	-.056	-.211	-.258	-.327	-.254	-.298	-.347	-.291	-.273
6	4	-.032			-.287	-.177	-.262	-.293	-.239	-.150
7	3	.132			-.150		-.131	-.120	-.116	-.113
8	4	.163	.157	.150		.119	.090		.058	.057
9	6	.147	.188	.188	.219	.185	.193	.218	.202	.203
10	2	-.158	-.096					.107	.090	.073
11	1	.160						.110	.112	.100
12	6	-.076	-.244	-.260	-.264	-.291	-.278	-.294	-.293	-.295
13	6	-.149	-.146	-.174	-.262	-.220	-.253	-.274	-.242	-.242
15	5	.165		.153	.256	.232	.238	.271	.236	.236
R^2			.318	.332	.370	.363	.378	.388	.394	.399

Table M

Evaluation of Variables from Data Set #1

	Variable	Number of Models	Zero-order correlation	Median Part correlation in reduced models	Part correlation in complete model	Tolerance
Good predictors – 1	3	22	.618	.276	.202	.548
	11	22	.547	.296	.211	.537
Good predictors – 2	8	19	.468	.164	.079	.543
Fair predictors – 1	1	11	.482	.116	.117	.630
	30	10	.350	.090	.057	.800
	31	10	.302	.121	.065	.756
	32	8	.438	.109	.023	.458
Fair predictors – 2	6	4	.423	.104	.051	.479
	16	4	.442	.097	.053	.532
	28	4	.301	.101	.046	.642
	36	4	.424	.093	.048	.706
Questionable predictors	7	2	.273	.061	-.001	.552
	17	3	-.232	-.074	-.030	.769
	19	1	.236	-.010	-.019	.795
	35	1	.335	.089	.035	.782
	39	6	.195	.104	.033	.709
Poor predictors – 1	14	1	.109	-.056	-.032	.691
	21	1	.148	.033	.005	.788
Poor predictors – 2	2	0	.210		.000	.733
	5	0	.259		-.057	.666
	10	0	-.288		-.025	.737
	13	0	-.247		-.026	.769
	15	0	.446		.031	.602
	18	0	.227		.007	.559
	23	0	.236		-.001	.710
	25	0	.331		.005	.765
	29	0	.327		-.005	.541
	33	0	.200		-.011	.781
Poor predictors – 3	4	0	.000		-.010	.723
	9	0	.095		-.009	.862
	20	0	.186		.000	.668
	22	0	.137		-.002	.824
	24	0	-.054		.028	.798
	26	0	.157		-.015	.823
	27	0	.179		.009	.788
	34	0	-.066		-.006	.796
	37	0	-.022		.010	.854
	38	0	.014		-.001	.845
	40	0	.105		-.011	.913
	41	0	-.092		-.018	.887

Subjective criteria for classification

	Number of Models	Zero-order correlation	Part correlations
Good	Most	Good	Good
Fair	Few-Many	Fair	Fair-Good
Questionable	Few	Poor	Fair-Good
Poor	Few	Low-Fair	Poor

Table N

Evaluation of Variables from Data Set #2

	Variable	Number of Models	Zero-order correlation	Median Part correlation in reduced models	Part correlation in complete model	Tolerance
Good predictors – 1	19	12	.673	.162	.083	.372
	20	16	.718	.223	.147	.372
Fair predictors – 1	5	10	.618	.111	.053	.381
	10	10	.664	.122	.070	.386
	15	9	.654	.113	.071	.392
Questionable predictors	14	5	.644	.090	.044	.404
	18	3	.559	.102	.051	.380
Poor predictors – 1	6	1	.599	.093	.010	.399
	7	2	.592	.081	.022	.428
	8	1	.545	.073	.029	.498
	9	3	.602	.090	.014	.387
	16	2	.535	.068	.012	.516
	17	1	.526	.086	.006	.461
Poor predictors – 2	2	0	.523		.020	.555
	3	0	.477		.002	.597
	4	1	.503	.035	.005	.534
	11	0	.554		.004	.506
	12	0	.581		.023	.475
	13	0	.482		-.014	.480

Subjective criteria for classification

	Number of Models	Zero-order correlation	Part correlations
Good	Most	-	Good
Fair	Many	-	Fair-Good
Questionable	Few	-	Fair-Good
Poor	Few	-	Poor-Fair

Table O

Evaluation of Variables from Data Set #3

	Variable	Number of Models	Zero-order correlation	Median Part correlation in reduced models	Part correlation in complete model	Tolerance
Good predictors	2	6	.222	.411	.308	.299
	5	6	-.056	-.278	-.273	.302
	9	6	.147	.191	.203	.693
	12	6	-.076	-.271	-.295	.434
	13	6	-.149	-.237	-.242	.413
	15	5	.165	.238	.236	.068
Fair predictors	4	3	-.094	-.151	.043	.253
	6	4	-.032	-.275	-.150	.111
	7	3	.132	-.131	-.113	.158
	8	4	.163	.135	.057	.517
Questionable predictors	11	1	.160	.110	.100	.486
Poor predictors	3	0	.056		.061	.059
	10	2	-.158	.006	.073	.288
	14	0	-.053		-.036	.715

Subjective criteria for classification

	Number of Models	Zero-order correlation	Part correlations
Good	Most	-	Good
Fair	Many	-	Fair-Good
Questionable	Few	-	Fair-Good
Poor	Few	-	Poor

Table P

Models with "Good" Predictors

Data Set #1:
6 variables: 1, 3, 8, 11, 31, 32

Found in 0 of the 8 samples with N = 500
Found in 5 of the 8 samples with N = 5,000
Found 5 times by the forward stepwise method
Found 2 times by the backward stepwise method
Found 5 times by the best subsets method

Models found with population data:

Forward stepwise	1, 3, 8, 11, 31, 32
Backward stepwise	1, 3, 8, 11, 28, 31
Best subsets	1, 3, 8, 11, 31, 32

Data Set #2:
5 variables: 5, 10, 15, 19, 20

Found in 0 of the 8 samples with N = 500
Found in 5 of the 8 samples with N = 5,000
Found 2 times by the forward stepwise method
Found 4 times by the backward stepwise method
Found 2 times by the best subsets method

Models found with population data:

Forward stepwise	5, 10, 15, 19, 20
Backward stepwise	5, 10, 15, 19, 20
Best subsets	5, 10, 15, 19, 20

Data Set #3:
8 variables: 2, 5, 6, 8, 9, 12, 13, 15 (8 IV)

Not found by the forward stepwise, backward stepwise, or best subsets method

Table Q

Identifying Good Predictors from Data Set #1 without Stepwise Procedures Population Data

Variable	Zero-order Correlation	Good alone[a]	Good together[a]	Part correlation in simultaneous model
1	.482	x	x	.117
2	.210			.000
3	.618	x	x	.202
4	.000			-.010
5	.259		x	-.057
6	.423	x	x	.051
7	.273			-.001
8	.468	x	x	.079
9	.095			-.009
10	-.288			-.025
11	.547	x	x	.211
13	-.247			-.026
14	.109			-.032
15	.446	x		.031
16	.442	x	x	.053
17	-.232			-.030
18	.227			.007
19	.236			-.019
20	.186			.000
21	.148			.005
22	.137			-.002
23	.236			-.001
24	-.054			.028
25	.331			.005
26	.157			-.015
27	.179			.009
28	.301			.046
29	.327			-.005
30	.350	x	x	.057
31	.302		x	.065
32	.438	x		.023
33	.200			-.011
34	-.066			-.006
35	.335			.035
36	.424	x	x	.048
37	-.022			.010
38	.014			-.001
39	.195			.033
40	.105			-.011
41	-.092			-.018

[a]Top 10 variables

8 variables are in the top 10 alone and together:	1, 3, 6, 8, 11, 16, 30, 31
6 variables are identified as "best" predictors:	1, 3, 8, 11, 31, 32
2 "best" predictors are not identified here:	31, 32
3 predictors identified here are not "best" predictors:	6, 16, 30

Table R

Identifying Good Predictors from Data Set #2 without Stepwise Procedures Population Data

Variable	Zero-order Correlation	Good alone[a]	Good together[a]	Part correlation in simultaneous model
2	.523			.020
3	.477			.002
4	.503			.005
5	.618	x	x	.053
6	.599	x		.010
7	.592	x	x	.022
8	.545		x	.029
9	.602	x		.014
10	.664	x	x	.070
11	.554			.004
12	.581	x	x	.023
13	.482			-.014
14	.644	x	x	.044
15	.654	x	x	.071
16	.535			.012
17	.526			.006
18	.559		x	.051
19	.673	x	x	.083
20	.718	x	x	.147

[a]Top 10 variables
8 variables are in the top 10 alone and together: 5, 7, 10, 12, 14, 15, 19, 20
5 variables are identified as "best" predictors: 5, 10, 15, 19, 20
0 "best" predictors are not identified here:
3 predictors identified here are not "best" predictors: 7, 12, 14

Table S

Identifying Good Predictors from Data Set #3 without Stepwise Procedures Population Data

Variable	Zero-order Correlation	Good alone[a]	Good together[a]	Part correlation in simultaneous model
2	.222	x	x	.308
3	.056			-.061
4	-.094	x		.043
5	-.056		x	-.273
6	-.032		x	-.150
7	.132	x	x	-.113
8	.163	x		.057
9	.147	x	x	.203
10	-.158	x	x	.073
11	.160	x	x	.100
12	-.076	x	x	-.295
13	-.149	x	x	-.242
14	-.053			-.036
15	.165	x	x	.236

[a]Top 10 variables

8 variables are in the top 10 alone and together: 2, 7, 9, 10, 11, 12, 13, 15
8 variables are identified as "best" predictors: 2, 5, 6, 8, 9, 12, 13, 15
3 "best" predictors are not identified here: 5, 6, 8
3 predictors identified here are not "best" predictors: 7, 10, 11

Table T

Predictors for Data Set #1 Categorized by Factors

IV	Factor	Number of Predictors																					
		1	5	5	5	5	5	5	6	6	6	6	6	6	6	7	7	7	7	7	8	18	40
6	1	.423							.096				.112						.104			.048	.051
7	1	.273														.062					.059	-.006	-.001
32	1	.438				.117		.147			.104				.100		.104			.121		.030	.023
36	1-2	.424		.106								.093							.093			.049	.048
1	2	.482	.119						.108	.121	.114	.109		.115		.118	.125	.116			.119	.105	.117
11	2	.547	.287	.321	.326	.327	.338	.327	.269	.284	.278	.271	.304	.285	.325	.288	.281	.278	.304	.311	.286	.252	.211
3	2-3	.618	.272	.302	.306	.293	.333	.348	.261	.266	.252	.254	.279	.250	.289	.261	.238	.252	.287	.320	.255	.212	.202
8	3	.468	.202	.169	.184	.156			.177	.179	.164	.184	.119	.146	.140	.161	.138	.162	.148		.131	.092	.079
14	3	.109																			-.056	-.031	-.032
16	3	.442					.142						.094	.085							.100	.062	.053
19	3	.236																	-.010			-.025	-.019
17	5	-.232															-.073		-.074			-.043	-.030
21	6	.148																		.033		.005	.005
28	7	.301		.114						.101					.101			.095				.046	.046
35	7	.335																		.089		.040	.035
31	8	.302	.131		.134				.121	.108	.115	.121	.121	.126				.105		.118		.066	.065
39	8	.195					.117	.108								.106					.101	.034	.033
30	x	.350			.098	.088	.102	.110							.083	.092	.076	.085		.098		.058	.057
R^2			.593	.586	.588	.584	.570	.571	.602	.603	.603	.601	.599	.600	.594	.601	.602	.610	.589	.584	.603	.634	.643

11 factors with eigenvalues > 1.00 using Principle Components Analysis with Varimax rotation
No predictors from factors 4, 9, 10, or 11
Only 1 predictor per model from factors 1, 5, 6, 7, 8
Multiple predictors from factors 2 & 3

Table U

Predictors For Data Set #2 Categorized by Factors

IV	Factor	Number of Predictors																		
		1	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	14	19
4	1	.503					.035												.008	.005
5	1	.618	.119	.150	.145		.101	.113	.098	.112	.109	.110	.107						.056	.053
6	1	.599												.093					.013	.010
7	1	.592						.084							.078				.022	.022
8	1	.545												.073					.030	.029
9	1	.602							.067							.090	.092		.018	.014
10	1	.664	.138			.155	.132			.116	.117	.127	.120		.123	.120	.121		.072	.070
14	1	.644								.085	.091			.086		.090		.094	.049	.044
15	1	.654				.113			.173	.147		.149	.100	.168	.104		.102	.107	.071	.071
19	1	.673	.181	.186	.185	.126	.178	.170			.159		.113		.117	.165	.116	.118	.084	.083
20	1	.718	.221	.267	.254	.233	.216	.230	.270	.197	.179	.228	.205	.248	.214	.186	.210	.224	.158	.147
16	2	.535							.068			.067							.018	.012
17	2	.526		.086															.006	.006
18	2	.559			.113			.102										.102	.052	.051
R^2			.639	.628	.633	.638	.641	.640	.630	.644	.648	.641	.649	.628	.644	.644	.646	.637	.665	.666

2 factors with eigenvalues > 1.00 using Principle Components Analysis with Varimax or Oblimin rotation
3-5 predictors from factor 1
0 -1 predictors from factor 2

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

ERIC®

TM033862

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: STEPWISE REGRESSION AS AN EXPLORATORY DATA ANALYSIS PROCEDURE

Author(s): JEROME D. THAYER

Corporate Source:

Publication Date:

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	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	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	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	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:	Printed Name/Position/Title: JEROME THAYER	
Organization/Address: ANDREWS UNIVERSITY BERRIEN SPRINGS, MI 49104	Telephone: 616-471-6214	FAX:
	E-Mail Address: THAYERJ@ANDREWS.EDU	Date: 3-27-02

(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:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

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