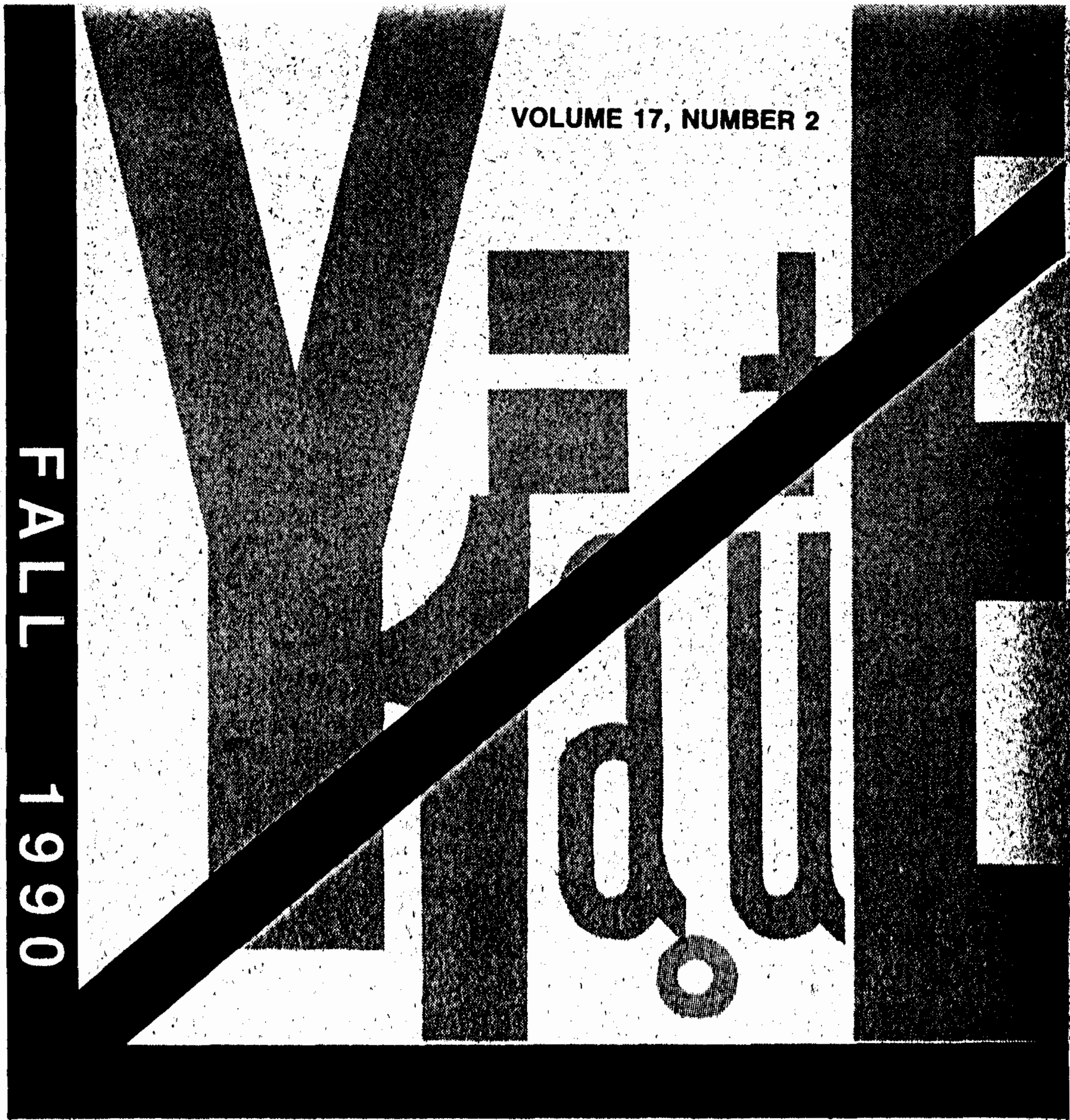
VOLUME 17, NUMBER 2

FALL 1990



MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group on Multiple Linear Regression

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

ISSN 0195-7171

MULTIPLE LINEAR REGRESSION VIEWPOINTS

TABLE OF CONTENTS

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

Teaching ANCOVA: The Importance of Random Assignment

Ralph O. Mueller
University of Toledo

Abstract

The purpose of this paper is to present a simple approach to teaching the fundamental concepts underlying the Analysis of Covariance (ANCOVA) with particular attention to the assumption of random assignment. The main advantage of using ANCOVA in experimental research is the gain in statistical power due to a reduction in error variance. As a by-product, ANCOVA provides statistically modified group means that compensate for non-systematic group differences on the covariate. A continuing misconception, however, is that ANCOVA "equates" previously unequal groups with respect to a covariate even if these preexisting differences are systematic ones. Teachers of research methodologies are urged to clarify and expand on the sometimes insufficient presentations of ANCOVA to prevent further misapplications.

Introduction

The Analysis of Covariance (ANCOVA) has long been used in the behavioral sciences as an important data analysis tool. In many contemporary texts on research methodologies an entire chapter is devoted to ANCOVA, as, for example, in Cohen and Cohen (1983), Hinkle, Wiersma, and Jurs (1988), Howell (1987), Keppel (1982), Kirk (1982), Marascuilo and Serlin (1988), or Pedhazur (1982), to name just a few. Usually described as an integration of Analysis of Variance (ANOVA) and Multiple Linear Regression (MLR), the ANCOVA model can be represented as a special case of the

An earlier version of this paper was presented at the 1989 meeting of the Mid-Western Educational Research Association in Chicago, Illinois. The author thanks the editor and reviewers for their helpful suggestions and Mr. Tito Mendoza, Research Assistant, for helping with preparing the final document.

General Linear Model (GLM), including model components of both, ANOVA and MLR. The ANCOVA's general goal can be viewed as being very similar to that of ANOVA: the technique helps answering the question of whether observed group differences on some dependent variable are attributable to sampling fluctuations alone or to true population differences between the groups (in fact, in a true experiment both procedures test the same null-hypothesis, $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$, as explained below).

In experimental research settings ANCOVA has the main advantage of error variance reduction so that true group differences are easier to detect; that is, compared to ANOVA, the Analysis of Covariance provides an increase in statistical power provided certain assumptions are met (Keppel, 1982, p. 483). The error reduction is achieved by adding one or more continuous explanatory variables to the model, called the *covariate(s)*, that (a) are related to the dependent variable as much as possible, but (b) are unrelated to each other and to the independent variable(s) that indicate group membership. A by-product of the application of ANCOVA is the calculation and subsequent interpretation of the *adjusted means* which are group means on the dependent variable that have been statistically adjusted for preexisting *non-systematic* group differences on the covariate(s) (Keppel, 1982, p. 483). Generally, adjusted means can be interpreted as predicted mean scores that would be expected if all group covariate means were exactly equal (to the grand covariate mean) rather than different due to random sampling fluctuations.

Recognizing the advantages of ANCOVA, researchers soon began to apply the technique to data obtained from quasi- and non-experimental research settings as well, and soon, first warnings against the use of the Analysis of Covariance began to appear in the literature; see, for example, Cook and Campbell (1979), Elashoff (1969), Lord (1967, 1969), or, more recently, Huitema (1980). One of the focal points of the discussion continues to be the potential misinterpretation of adjusted means. Some authors argue that "the analysis of covariance, which is also used in experimental studies, is a statistical method that can be used to equate groups on one or more variables" (Gay, 1987, p.254). But statements similar to the one above overstate and misinterpret the real advantage of ANCOVA especially when used in quasi- or non-experimental research. Group differences on the covariate are likely to be systematic when dealing with in-tact groups; ANCOVA, however, is not intended to adjust for systematic differences, just for non-systematic ones (Keppel, 1982, pp. 481-492). For an excellent and comprehensive discussion on interpretation problems associated with ANCOVA, consult Huitema (1980, chap. 7) who warned that "in general, ANCOVA is not an appropriate procedure for the analysis of nonequivalent group studies" (p. 154).

Today, ANCOVA's advantages are well known and its disadvantages and limitations are recognized and understood by most. Some introductory texts in research methodology, however, still mislead the research neophyte somewhat by stating in very general terms that the use of ANCOVA will statistically "equate" previously unequal groups on the covariate (e.g., Borg & Gall, 1989, p. 556; Gay, 1987, p. 254; Huck, Cormier, & Bounds, 1974, pp. 134-136). Others make conflicting remarks regarding the interpretation of adjusted means and the appropriateness of ANCOVA in quasi- or non-experimental research (e.g., Marascuilo & Serlin, 1988, p. 608 and p. 611; Wiersma, 1986, p. 354). The intent of this paper is not to criticize specific textbook authors; rather, it serves to present a simple approach to teaching the fundamental concepts of ANCOVA in a beginning research methodology or applied statistics course. The emphasis here is on the importance of the assumption of random assignment and the potential misapplications of ANCOVA in quasi and non-experimental research. Especially students of research methods that do not specialize in the field need to be aware of common misuses of this widely used technique.

The Statistical Model and Adjusted Means

A suggested approach to teaching the underlying concepts is to begin with a presentation of the General Linear Model (GLM) expression of ANCOVA. Under certain statistical assumptions (see Cook & Campbell, 1979, Elashoff, 1969, or Huitema, 1980, chap. 6), the model for a one-way linear ANCOVA can be expressed as

(1) $Y_{ik} = \mu + \alpha_k + \beta_w(X_{ik} - \mu_X) + e_{ik}$

where Y_{ik} denotes the i^{th} score on the dependent variable in the k^{th} group, μ is the grand mean of the dependent variable, $\alpha_k = (\mu_k - \mu)$ is the k^{th} group effect, β_w denotes the regression coefficient representing the linear relationship between the dependent variable and the covariate. X_{ik} is a score on the covariate. μ_X is the grand mean of the covariate, and e_{ik} denotes random error associated with each subject's score.

Without loss of generality, assume that two groups are being compared ($k=2$). Mean group differences on the dependent variable can then be expressed as

$$(2) \quad \mu_{Y_1} - \mu_{Y_2} = [\mu + \alpha_1 + \beta_w(\mu_{X_1} - \mu_X)] - [\mu + \alpha_2 + \beta_w(\mu_{X_2} - \mu_X)]$$
$$= (\alpha_1 - \alpha_2) + \beta_w(\mu_{X_1} - \mu_{X_2})$$

The last expression in Equation 2 shows that observed sample differences cannot be uniquely attributed to group effects but could also be due to mean differences on the covariate. Rewriting Equation 1 as

$$(3) \quad Y_{ik(adj)} = Y_{ik} - \beta_w(X_{ik} - \mu_X) = \mu + \alpha_k + e_{ik}$$

where $Y_{ik(adj)}$ denotes an *adjusted score*, and defining the *adjusted mean* in the k^{th} group as

$$(4) \quad \mu_{Y_k(adj)} = \mu_{Y_k} - \beta_w(\mu_{X_k} - \mu_X) = \mu + \alpha_k$$

proves to be helpful since now differences between adjusted means can be attributed to group effects alone:

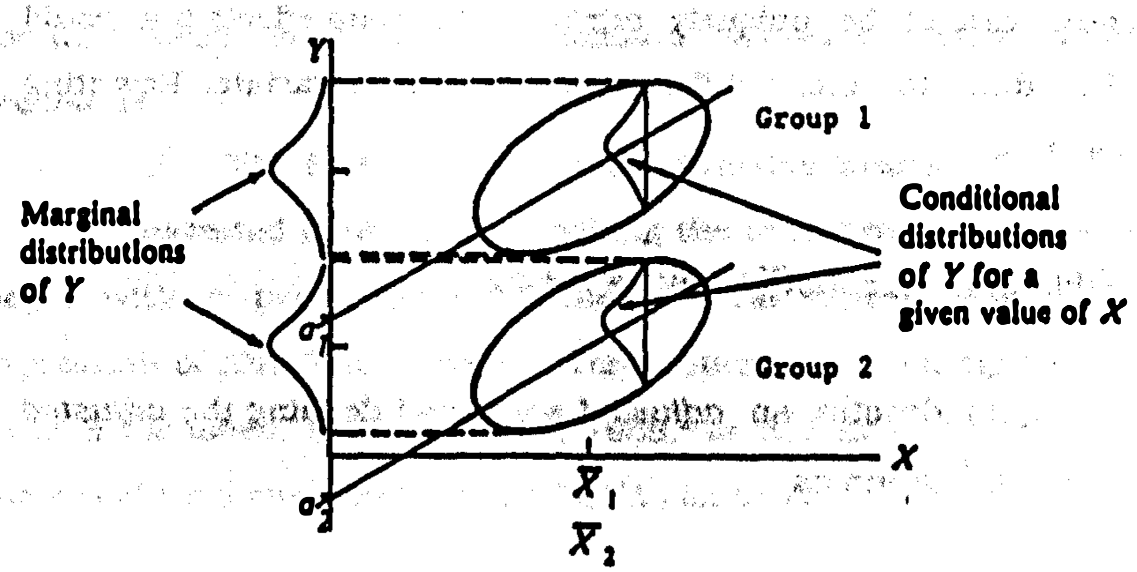
$$(5) \quad \mu_{Y_1(adj)} - \mu_{Y_2(adj)} = (\mu + \alpha_1) - (\mu + \alpha_2) = \alpha_1 - \alpha_2$$

Additionally, using Equation 4, it can be seen that

$$(6) \quad \mu_{Y_1(adj)} - \mu_{Y_2(adj)} = (\mu_{Y_1} - \beta_w \mu_{X_1}) - (\mu_{Y_2} - \beta_w \mu_{X_2}) = a_1 - a_2$$

where a_k is the intercept term for the regression of the dependent variable on the covariate in Group k. Thus, differences between adjusted means can also be interpreted as differences between regression intercepts in the separate regressions of the dependent variable on the covariate (see Figure 1).

FIGURE 1
Error Reduction in ANCOVA

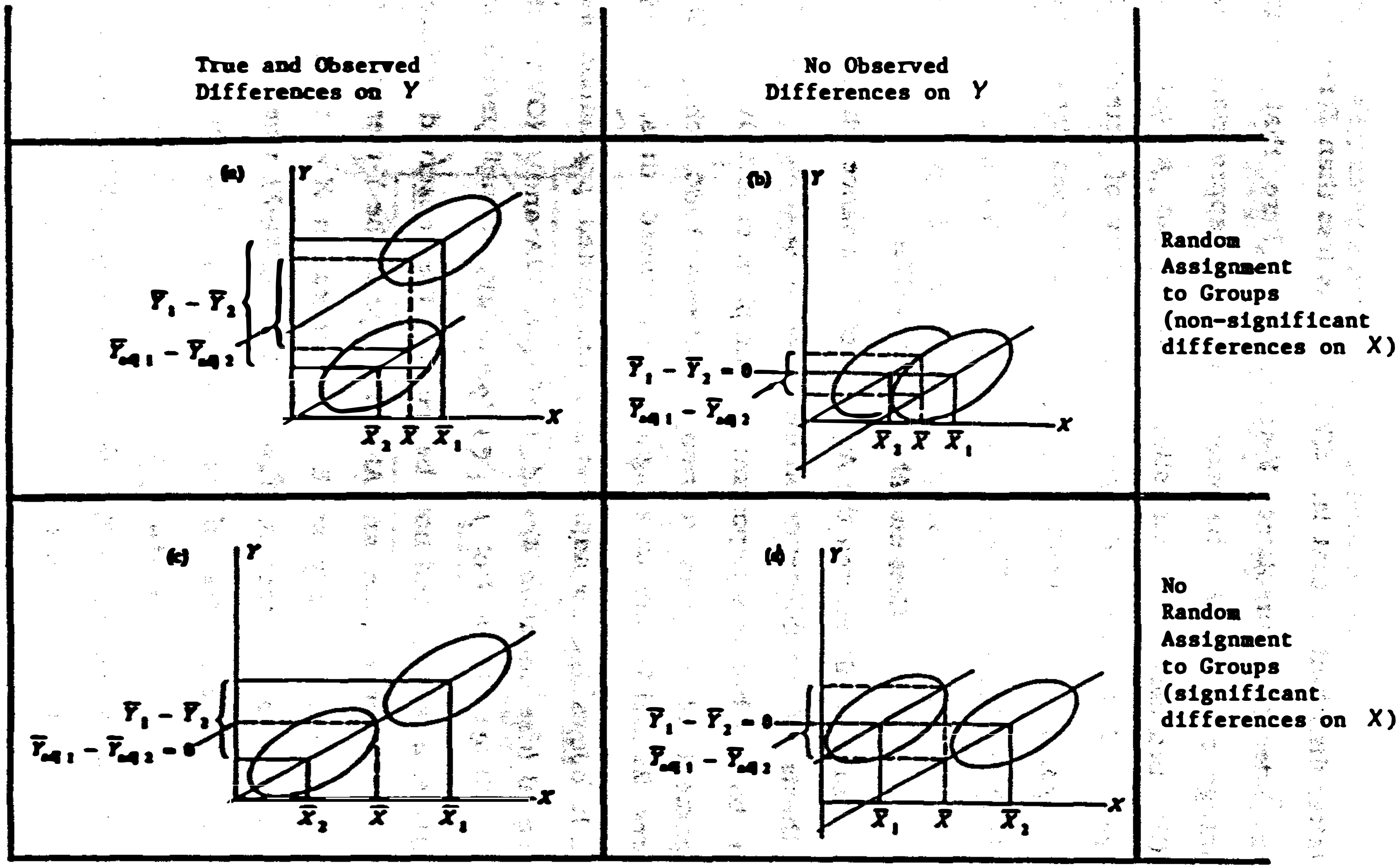


Assuming that $\beta_w \neq 0$, note the reduction in error variance as shown in Figure 1. When a covariate is included in the model, error variance is determined by the *conditional*, rather than *marginal* distribution of Y. The former has a smaller variance estimate than the latter distribution, that is, $\Sigma(Y-\hat{Y})^2/df$ is less than $\Sigma(Y-\bar{Y}_k)^2/df$, where $\hat{Y}$ denotes a predicted Y-score. It is here that the main advantage of ANCOVA becomes apparent: appropriately used, ANCOVA provides more statistical power than a conventional Analysis of Variance design; the probability of detecting true differences on the dependent variable is increased by a decrease in estimated error variance.

Uses and Abuses of ANCOVA

Figure 1 illustrated the Analysis of Covariance when the null-hypothesis of no difference on the covariate is true, a consequence of a basic - but very important - assumption of ANCOVA: random assignment of subjects to groups (Huitema, 1980, chap. 6). In Figures 2a and 2b random assignment is assumed; thus, $\mu X_k = \mu X$ for all k. It follows that the adjusted and unadjusted *population* means are equal (use Equation 4) and that ANCOVA and ANOVA test the same null-hypothesis, H_0: $\mu Y_1 = \mu Y_2 = ... = \mu Y_k$. *Sample* means on the covariate, however, need not be equal. The observed differences are due to chance alone and ANCOVA adjusts the Y-means for these non-systematic - usually small - differences on X via the definition of adjusted sample means, $\bar{Y}_{k(adj)} = \bar{Y}_k - b_W(\bar{X}_k - \bar{X})$, where the terms are sample estimates of the corresponding terms in Equation 4. Figures 2a and 2b illustrate that the analysis will lead to correct conclusions regarding group differences on the dependent variable provided ANCOVA is used in conjunction with random assignment of subjects to groups. In such a case, the difference between adjusted sample means is an unbiased estimate of what the

FIGURE 2
Uses and Abuses of ANCOVA



difference between group means on the dependent variable would have been if each group had equal covariate means. Only in this sense can one claim that groups were "equated" with respect to the covariate (recall that population covariate means were *assumed* to be equal).

When ANCOVA is used in quasi- or non-experimental research settings, it is often the case that the groups under study systematically differ on the covariate and possibly on other relevant variables, that is, the randomization assumption was violated. What effect will this have on research results based on an ANCOVA? Huitema (1980, chap. 6 and chap. 7) provided a comprehensive and detailed discussion on the consequences of assumption violation and there is no need to repeat his arguments here. However, consider a less technical treatment of the potential misinterpretation of an ANCOVA when indeed the groups differ on the covariate.

At the beginning of this paper one possible way of expressing the general goal of ANCOVA was stated: to detect whether groups significantly differ on some dependent variable. When large group differences on the covariate exist, ANCOVA might mislead the researcher regarding this general question. Consider Figures 2c and 2d. Misinterpretations are possible in two situations. First, although the two groups are *different* with respect to the dependent variable, ANCOVA leads to a conclusion of *equality* in adjusted means (indicated by equal regression intercepts in Figure 2c). This is often interpreted by stating that the covariate "explains" true differences, especially after a significant ANOVA analysis. The fact remains, however, that in situations like this

ANCOVA will not indicate differences between the two groups eventhough the groups differ on the dependent variable. Second, if the groups are *equal* with respect to the dependent variable, ANCOVA can lead to the conclusion that they *differ* after covariate adjustment (indicated by unequal intercepts in Figure 2d). Situations like these are sometimes referred to as cases of "Lord's paradox" (Lord, 1967). In a very illuminating and critical paper Bock (1969) claimed that the "paradox" is merely a misunderstanding: ANOVA and ANCOVA answer different questions since the former technique is based on the marginal *Y*-distribution, while the latter deals with the distribution of *Y*-scores conditional on the covariate. Note, however, that ANCOVA is not likely to provide unbiased adjusted means when used in nonequivalent group designs (Huitema, 1980, p. 142). The difference between adjusted sample means might be a *biased* estimate of what the difference between group means on the dependent variable would have been if each group had equal covariate means.

The brief discussion above - in addition to other potential interpretation problems (Huitema, 1980) - indicates that it might be of advantage to test for differences on the covariate as a preliminary step in the data analysis. If the hypothesis of no difference is rejected, ANCOVA might motivate false (or at least misleading) conclusions regarding group differences on the dependent variable; if the hypothesis is retained, ANCOVA might be appropriate and lead to a more powerful analysis. But what are the consequences of a Type I or Type II error in such a preliminary test? In the first situation one would erroneously conclude that covariate differences exist and, given the previous discussion, might not use ANCOVA for the data analysis eventhough it would have been appropriate; the loss of statistical power is the consequence. Under the second situation - that is, falsely concluding that no systematic covariate differences are present - ANCOVA might be used inappropriately and lead to false conclusions. The latter is the reason for testing group differences on the covariate at a more liberal level of significance, say .10 or .20; protection against a Type II error seems more important than protection against a Type I error.

Conclusion

The Analysis of Covariance model can be represented as a special case of the General Linear Model; it includes both, Analysis of Variance and Multiple Linear Regression components. The main advantage of using ANCOVA is a reduction in error variance achieved through the inclusion of additional explanatory variables (covariates) when assessing mean group differences on some dependent variable. As such, ANCOVA provides a statistically powerful way of detecting true group differences but can also lead to false conclusions regarding these group differences when the assumption of random assignment is violated and groups significantly differ on the covariate. Teachers are urged to discuss potential misapplications and discourage the use of ANCOVA when the random assignment assumption is not met. One indication of possible misuse can be provided by rejecting the hypothesis of no difference between covariate group means at a liberal level of significance to guard against a possible Type II error. The best protection against potentially serious misinterpretations of ANCOVA results, however, is to restrict its use to true - or nearly true - experimental designs. In accordance with others (Elashoff, 1969; Huitema, 1980; Keppel, 1982), the Analysis of Covariance is not recommended in nonequivalent group studies.

ANCOVA still is an important and powerful data analysis tool in a variety of applied research situations. Nearly every comprehensive textbook on research methodologies includes a discussion on ANCOVA and the technique is presented in most university courses on applied statistics or research design. However, the technique is also frequently misunderstood; misconceptions like "ANCOVA can equate previously nonequivalent groups on the covariate(s)" still circulate through some uninitiated minds. Teachers of research methods and authors of textbooks are in the position to start the initiation process - or should there be an alumni initiation first?

References

Bock, R. D. (1969). *Conditional and unconditional inference in the analysis of repeated measurements.* Paper presented at the Symposium on the Application of Statistical Techniques to Psychological Research, Canadian Psychological Association, York University.

Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed.). New York, NY: Longman.

Cohen, J., & Cohen P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston, MA: Houghton Mifflin.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal, 6* (3), 383-401.

Gay, L. R. (1987). *Educational research: Competencies for analysis and application* (3rd ed.). Columbus, OH: Merrill.

Hinkle, D. E., Wiersma, W., & Jurs, S. (1988). *Applied statistics for the behavioral sciences* (2nd ed.) Boston, MA: Houghton Mifflin.

Howell, D. C. (1987). *Statistical methods for psychology* (2nd ed.). Boston, MA: Duxbury.

Huck, S. W., Cormier, W. H., & Bounds, W. G. (1974). *Reading statistics and research.* New York, NY: Harper & Row.

Huitema, B. E. (1980). *The analysis of covariance and alternatives.* New York, NY: Wiley.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Monterey, CA: Brooks/Cole.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304-305.

Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin, 72*, 336-337.

Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical methods for the social and behavioral sciences.* New York, NY: Freeman.

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

Wiersma, W. (1986). *Research methods in education* (4th ed.). Boston, MA: Allyn and Bacon.

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

Corporate Manager's Leadership Style and Existence of Employee Health Promotion Programs

Elizabeth Kinion
The University of Akron

ABSTRACT

The establishment and the quality of health promotion programs depend on supportive corporate management. However, there is a paucity of research investigating the area of leadership in corporations as it relates to health promotion programs. In general, the research on health promotion consists primarily of types of programs, cost effectiveness, and physiological responses to specific health behaviors.

The purpose of this study was to examine the relationship of corporate managers' leadership style, determined by Likert's Profile of Organizational Characteristics, and the existence of employee health promotion programs. One hundred eighty-seven corporate officers in Northeastern Ohio completed the questionnaire entitled Corporate Leadership Styles and the Existence of Employee Health Promotion Programs which included questions from Likert's Profile of Organizational Characteristics, general information, demographic data, and questions about the effects of health promotion. Multiple linear regression procedures were used to analyze the variance in predicting one variable to another. The F test was applied to determine statistical significance at the .05 level.

The results of hypothesis testing for the sample indicated leadership style, as measured by Likert's Profile of Organizational Characteristics, does not aid in predicting the existence of an employee health promotion program. Leadership styles of the respondents in this study clustered around System 2 and System 3. System 2, the benevolent-authoritative system, and System 3, the consultive system, are intermediate systems. These systems resemble the extreme from which they deviate. However, data from a subset of the sample (managers from corporations with health promotion programs) indicated knowledge of leadership style may be used to predict corporate officers' perception that health promotion programs increase employee morale. In addition, data from this subset indicated corporate officer participation in the decision to establish a health promotion program leads to a predictive relationship that health promotion programs are cost effective, increase employee productivity, and decrease absenteeism.

INTRODUCTION

Historically, the practice of medicine and, therefore, health care was disease and acute care oriented. In the years between 1875 and 1924 medical advances were based on environmental factors such as improved sanitation and antiseptic surgery. From around 1925 to 1950, discoveries of sulfa and penicillin decreased the mortality rate by providing a "cure" for infectious diseases. Americans viewed the physician as a person who could cure their ills. Medicine has continued to respond with cures, such as open heart surgery, organ transplants, and pharmaceutical break throughs such as synthesis of hormones and genetic engineering of DNA. Until recently, this curative approach to health care has continued without scrutiny, in spite of the fact that 7 of the 10 leading causes of death in the United States during the 1980's are related, directly or indirectly, through risk factors, to behavior or lifestyle (Brady, 1983).

As a nation, we have expended large amounts of money for health care. In the years from 1960 to 1978, annual health care expenditures increased over 700%. Hookman (1984) notes that although the national inflation rate declined in 1983 and 1984, hospital room costs increased in 1981, 1982, and 1983. Three hundred and thirty-two billion dollars, or 10.5%, of the Gross National Product was spent on health care in 1982. This exceeds federal outlays for defense by nearly $150 billion and averages out to $1,365 per person, or $140 more than in 1981. To a large extent, the increased expenditures focused on disability and disease, not prevention (Department of Health, Education, and Welfare, 1979). Fielding (1984) reports that health promotion programs do not work if not strongly supported by top management.

A nationwide survey (Health Maintenance, 1973) of barriers toward better health and ways of overcoming them conducted among representative samples of the American public, business, and labor leaders indicates that:

> in the real world, the actual level of participation (in employer sponsored preventive health programs) would depend on the quality and availability of the programs, as well as the quality of the campaign within the company used to sell the employees (p. 82).

It would appear that Corporate America is striving to improve its competitiveness and productivity in the world market. A healthy workforce is essential since high absenteeism and poor performance due to physical or mental problems diminish productivity. Corporate management is one of the keys to the success of health promotion programs.

PROCEDURES

The population for this study included corporate managers of manufacturing companies within Northeastern Ohio that have 500 or more employees and were identified in the Ohio Directory of Manufacturers (1986). All 310 companies which met the previously stated criteria were surveyed. The total design method (TDM) was utilized in conducting this survey. Dillman (1978) notes that:

> in order to maximize the quantity and quality of survey responses, attention must be given to every detail that might affect response behavior. The TDM relies on a theoretically based view of why people do and do not respond to questionnaires and a well confirmed belief that attention to administrative details is essential to conducting surveys (p. viii).

Of the 310 companies surveyed, one hundred eighty-seven questionnaires were returned, representing 60% response rate.

The research design that was used was ex post facto. This ex post facto study was guided by hypotheses. Alternative or rival hypotheses are hypotheses that propose explanations for the effect other than the stated ones. Internal validity of the design can be increased when more of the rival hypotheses can be eliminated. However, Newman and Newman (1977) state:

> one must still keep in mind that by its very nature ex post facto research can never have total internal validity. Therefore, causation can never be inferred (p. 125).

The instrument used to identify the leadership style of corporate officials was the Profile of Organizational Characteristics (POC). This instrument which measures managerial styles was developed by Rensis Likert, and has been used extensively in previous research (Likert, 1978). Likert's Profile of characteristics identifies four leadership styles: (a) System 1, exploitive-authorative; (b) System 2, benevolent-authorative; (c) System 3, consultive; and (d) System 4, participative-group.

Likert Associates (personal communication, March 13, 1986) report the 18-item Form S usually yields split-half reliabilities in the .90 to .96 range when applying the Spearman-Brown formula for estimating reliability from the r between two halves of the form. Validity of the POC, found the rank order correlation (rho) between POC scores and performance data for a West Coast manufacturing firm was +.61. Data from 10 pairs of plants in Yugoslavia and two firms in Japan show consistent differences in profiles between high and low performing plants or departments in the expected direction.

Since this investigator was interested in the relationship of leadership style, personal characteristics, and demographic variables to the existence of health promotion programs the POC was only one component of the questionnaire. The POC was reproduced in booklet form. Transitional statements were used to facilitate transition from the POC questions to demographic data and, finally, questions about health promotion. The questionnaire booklet was entitled Corporate Leadership Styles and the Existence of Employee Health Promotion Programs.

STATISTICAL ANALYSIS

Specific research hypotheses were derived from the following research questions.

1. Are there differences in leadership styles (predictor variable) as identified on Likert's Profile of organizational characteristics instrument, of managers in corporations whith health promotion programs and those in corporations without such programs (criterion variable)?

2. Are there differences in leadership styles (predictor variable) of managers who favor health promotion programs (criterion variable) and those who do not?

3. Are there differences in leadership styles (predictor variable) of managers who have always advocated the establishment of health promotion programs and those who were not initially favorable but support such programs (criterion variable) after seeing them in operation?

4. Do age, sex, education, tenure in position, tenure with the corporation, or previous area of specialization within the corporation (predictor variables) relate to perceptions of health promotion programs (criterion variable)?

5. Is there a relationship between the managers leadership style (predictor variable) and the managers perception of health promotion programs (criterion variable)?

6. Does the origin of the idea for the health promotion program (predictor variable) or the manager's participation in the decision to provide a health promotion program relate to the manager's perception (criterion variable) of the program?

The F test was used to test the statistical significance of the proposed relationships in the research hypotheses. The F test was chosen because it is very robust. The assumptions of random selection of subjects and normal distribution of the variables can be violated without doing serious harm to the procedure.

Multiple linear regression was used in analyzing the variance in predicting from one variable to another and in covarying some of the variables to test the alternative hypotheses. Multiple linear regression was chosen because it is more flexible than traditional analysis of variance. With multiple linear regression, one can write the models that reflect the specific research question being asked. In addition, Newman (1976) points out that with multiple linear regression one can test relationships between categorical variables, between categorical and continuous variables, or between continuous variables.

Two tailed tests of significance were used to test the relationship of those variables where the direction of the correlation was uncertain. The .05 level of significance was used since it was the opinion of the investigator that the consequences of rejecting a true null hypothesis were not so serious as to warrant a more stringent confidence level.

Since four leadership styles were being tested, a correction for multiple comparisons was made if the overall *F* was significant. Newman and Newman (1977) report:

> When an overall *F* is significant and there are more than two groups, the question of where the difference is, always arises. To find out where the difference is, one generally runs multiple comparisons between the groups. That is, Group 1 is compared to Group 2, Group 1 is compared to Group 3, Group 2 is compared to Group 3, etc. As the number of comparisons (tests of significance), which are not independent of each other, increases, the more likely one will find significance (p. 221).

A variety of corrections may be used to control for alpha error buildup when making multiple comparisons. This researcher used $\frac{\alpha}{n-1}$.

Power analyses were performed to determine the probability of making a Type II error. Effect size ($\underline{f}^2$) was subjectively set at .15 which is defined as medium effect. The following formula noted by Newman and Newman (1983) was used to calculate power:

$$L = \underline{f}^2 v$$

where: N = number of replications
$v = df_2 \ (\underline{N} - m_1)$
~~$u = \underline{df_1}$~~ $(m_1 - m_2)$
m_1 = number of linearly independent vectors in full model
m_2 = number of linearly independent vectors in restricted model

Power was calculated for the most stringent model case, that is, the case in which power would be the lowest; therefore, the power estimates that follow for this study will be at least this high or higher. Three power estimates were given for small .02, medium .15, and large .35 effect sizes. For this study, therefore, power for effect size would be .15 if effect size was truly small for this population. Medium effect size would be .85 and large effect would be .92. Therefore, we can be fairly certain that if a medium or large effect does exist in the population, this study would be capable of detecting it. This study has low power and could detect a small effect size in a population 15 times out of 100. However, since the researcher is most interested in at least medium size effects, the researcher feels the power is sufficient for this study.

RESULTS

A vast majority of corporate officers are male, between ages 30 and 59, and have approximately 4 years experience as a corporate officer. This majority of corporate officers have at least a bachelor's degree. Of the companies returning completed questionnaires, 88 offer health promotion programs and 99 do not offer health promotion programs.

Of the respondents, 163 (88%) favor health promotion at the worksite. The majority of the respondents in this study were clustered in two systems. Systems 2, benevolent-authorative, and System 3, consultive. This study addressed six research questions. Responses from the entire sample (N = 187) were used to answer Research Questions 1 and 2 and related hypothesis. A subset (n = 88) of the sample responses from corporations with health promotion programs were used to answer Research Questions 3 through 6 and related hypotheses.

Hypotheses 1 through 10 relate to Research Question 1. These hypotheses and results are stated in Table 1. An examination of table one reveals that there is not a significant difference among leadership styles in predicting whether a corporation has a health promotion program. Leadership styles are not significantly different over and above corporate officer title, age, or gender, tenure in current position, tenure with corporation, education, and area of specialization in corporation prior to current position.

Hypothesis 11 relates to Research Question 2 and Hypotheses 12 and 13 relate to Research Question 3. These hypotheses and results are detailed on Table 2. An examination of Table 2 reveals the there is not a significant difference among leadership styles of managers who favor health promotion programs at the worksite and those who do not. Nor is there a significant difference among leadership styles in managers who always advocated the establishment of health promotion programs and those who initially were not in favor of the program, but now support such a program.

Table 1

Results of Hypotheses 1-10

Hypothesis	R^2	df	Alpha	F	P	s/ns
1. There is a significant difference among leadership styles in predicting whether a corporation has a health promotion program.						
Full Model	.071	17/162	.05	.729	.770	NS
Restricted Model	.0					
2. Leadership style, as measured by Likert's Profile of Organizational Characteristics, will be significantly difference between corporations that have a health promotion program and ones that do not over and above corporate officer title, age, and sex.						
Full Model	.099	18/151	.05	.648	>.05	NS
Restricted Model	.029					
3. Leadership style, as measured by Likert's Profile of Organizational Characteristics, will be significantly different between corporations that have health promotion programs and ones that do not over and above years in current position and years with corporation.						
Full Model	.106	18/154	.05	.734	>.05	NS
Restricted Model	.029					
4. Leadership style, as measured by Likert's Profile of Organizational Characteristics, will be significantly different between corporations that have health promotion programs and ones that do not over and above highest degree held and area of specialization prior to current positions.						
Full Model	.192	18/147	.05	.909	>.05	NS
Restricted Model	.102					
5. Age of the corporate officer accounts for a significant amount of variance in predicting the presence of a health promotion program.						
Full Model	.014	4/182	.05	.653	.625	NS
Restricted Model	.0					
6. Sex of the corporate officer accounts for a significant amount of variance in predicting the presence of a health promotion program.						
Full Model	.002	1/185	.05	.532	.466	NS
Restricted Model	.0					
7. Education of the corporate officer accounts for a significant amount of variance in predicting the presence of a health promotion program.						
Full Model	.055	6/180	.05	1.768	.108	NS
Restricted Model	.0					
8. Tenure as a corporate officer accounts for a significant amount of variance in predicting the presence of a health promotion program.						
Full Model	.0200	3/183	.05	1.244	.2948	NS
Restricted Model	.0					
9. Tenure with the corporation accounts for a significant amount of variance in predicting the presence of a health promotion program.						
Full Model	.014	3/183	.05	.906	.441	NS
Restricted Model	.0					
10. Previous area of specialization in the corporation prior to becoming a corporate officer accounts for a significant amount of variance in predicting the presence of a health promotion program.						
Full Model	.045	6/177	.05	1.409	.213	NS
Restricted Model	.0					

Table 2

Results of Hypotheses 11, 12, and 13

Hypothesis	R^2	df	Alpha	F	P	s/ns
11. There is a significant difference in leadership style in managers who favor better health promotion programs at the worksite and those who do not favor health promotion programs.						
Full Model	.090	17/160	.05	.937	.531	NS
Restricted Model	.0					
12. There is a significant difference in leadership style in managers who have always advocated the establishment of health promotion programs and those who were not initially in favor of the program, but now support such programs.						
Full Model	.122	15/65	.05	.606	.860	NS
Restricted Model	.0					
13. There is a significant difference between corporate officers who initially favored health promotion programs and those who later favored health promotion programs based on whether or not they participated in the decision to establish the program.						
Full Model	.017	1/84	.05	1.533	.219	NS
Restricted Model	.0					

Hypotheses 14 through 37 relate to Research Question 4. These hypotheses and results are detailed in Tables 3-5. An examination of Tables 3-5 reveals there is not a significant difference in demographics as they relate to the officers' perception of cost effectiveness of health promotion programs.

Table 3

Hypothesis 14 - 23

Hypothesis	R^2	df	Alpha	F	P	s/ns
14. Age of corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as cost effective.						
Full Model	.015	4/81	.05	.311	.870	NS
Restricted Model	.0					
15. Sex of corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion as cost effective.						
Full Model	.0004	1/84	.05	.033	.856	NS
Restricted Model	.0					
16. Education of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as cost effective.						
Full Model	.071	6/79	.05	1.017	.420	NS
Restricted Model	.0					
17. Tenure in the current position accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion as cost effective.						
Full Model	.042	3/82	.05	1.208	.312	NS
Restricted Model	.0					
18. Tenure with the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as cost effective.						
Full Model	.0419	3/82	.05	1.196	.3164	NS
Restricted Model	.0					
19. Previous area of specialization within the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as cost effective.						
Full Model	.046	6/78	.05	.635	.702	NS
Restricted Model	.0					
20. Age of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee morale.						
Full Model	.017	4/81	.05	.352	.841	NS
Restricted Model	.0					
21. Sex of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee morale.						
Full Model	.010	1/84	.05	.356	.873	NS
Restricted Model	.0					
22. Education of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee morale.						
Full Model	.065	6/79	.05	.929	.479	NS
Restricted Model	.0					
23. Tenure in current position accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee morale.						
Full Model	.022	3/82	.05	.617	.610	NS
Restricted Model	.0					

Table 4

Hypothesis 24 - 33

Hypothesis	R^2	df	Alpha	F	P	s/ns
24. Tenure with the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee morale.						
Full Model	.029	3/82	.05	.834	.481	NS
Restricted Model	.0					
25. Previous area of specialization within the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee morale.						
Full Model	.067	6/78	.05	.943	.469	NS
Restricted Model	.0					
26. Age of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee productivity.						
Full Model	.048	4/81	.05	1.032	.395	NS
Restricted Model	.0					
27. Sex of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee productivity.						
Full Model	.007	1.84	.05	.616	.434	NS
Restricted Model	.0					
28. Education of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee productivity.						
Full Model	.088	6/79	.05	1.270	.280	NS
Restricted Model	.0					
29. Tenure in position accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee productivity.						
Full Model	.084	3/82	.05	2.509	.063	NS
Restricted Model	.0					
30. Tenure with the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee productivity.						
Full Model	.032	3/82	.05	.914	.439	NS
Restricted Model	.0					
31. Previous area of specialization with the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as increasing employee productivity.						
Full Model	.070	6/78	.05	.986	.440	NS
Restricted Model	.0					
32. Age of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as decreasing absenteeism.						
Full Model	.020	4/81	.05	.424	.791	NS
Restricted Model	.0					
33. Sex of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program decreases absenteeism.						
Full Model	.003	1/84	.05	.262	.609	NS
Restricted Model	.0					

Table 5

Hypothesis 34 - 37

Hypothesis	R^2	df	Alpha	F	P	s/ns
34. Education of the corporate officer accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as decreasing absenteeism.						
Full Model	.043	6/79	.05	.597	.731	NS
Restricted Model	.0					
35. Tenure in current position accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as decreasing absenteeism.						
Full Model	.009	3/82	.05	.267	.849	NS
Restricted Model	.0					
36. Tenure in the corporation accounts for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as decreasing absenteeism.						
Full Model	.021	3/82	.05	.591	.626	NS
Restricted Model	.0					
37. Previous areas of specialization, within the corporation, account for a significant amount of variance in predicting whether the corporate officer perceives the health promotion program as decreasing absenteeism.						
Full Model	.017	6/78	.05	.231	.965	NS
Restricted Model	.0					

Hypotheses 38 through 42 relate to Research Question 5. These hypotheses and results are detailed in Table 6. An examination of Table 6 reveals that there is not a significant difference among leadership styles in predicting the managers' perception of health promotion programs as cost effective, increasing employee productivity, or decreasing absenteeism. There is not a significant difference in leadership styles of corporate officers who participated in the decision to establish a health promotion program and those who did not. However, there is a significant difference among leadership styles in predicting the managers' perception that health promotion programs increase employee morale.

Table 6

Hypothesis 38-42

Hypothesis	R^2	df	Alpha	F	P	s/ns
38. There is a significant difference among leadership styles in predicting the corporate official's perception that the health promotion program is cost effective.						
Full Model	.131	15/66	.05	.687	.788	NS
Restricted Model	.0					
39. There is a significant difference among leadership styles in predicting the corporate official's perception that the health promotion program increases employee morale.						
Full Model	.401	15/66	.05	2.950	.0013	S
Restricted Model	.0					
40. There is a significant difference among leadership styles in predicting the corporate official's perception that the health promotion program increases employee productivity.						
Full Model	.103	15/66	.05	1.490	.134	NS
Restricted Model	.0					
41. There is a significant difference among leadership styles in predicting the corporate official's perception that the health promotion program decreases employee absenteeism.						
Full Model	.145	15/66	.05	.752	.724	NS
Restricted Model	.0					
42. There is a significant difference among leadership styles in corporate officers who participated in the decision to establish a health promotion program, and those who did not participate in the decision.						
Full Model	.049	15/67	.05	.591	.872	NS
Restricted Model	.0					

Hypotheses 43–50 relate to Research Question 6. These hypotheses and results are detailed on Table 7. There is not a significant difference in where the idea for a health promotion program originates and the managers' perception of whether the program is cost effective, increases employee morale, increases employee productivity, and decreases absenteeism.

Table 7

Hypothesis 43-50

Hypothesis	R^2	df	Alpha	F	P	s/ns
43. There is a significant difference in where the idea for the health promotion program originated in predicting the manager's perception of whether the program is cost effective.						
Full Model	.007	4/81	.05	.153	.961	NS
Restricted Model	.0					
44. There is a significant difference in where the idea for the health promotion program originated in predicting the manager's perception of whether the program increases employee morale.						
Full Model	.068	4/81	.05	1.479	.216	NS
Restricted Model	.0					
45. There is a significant difference in where the idea for the health promotion program originated in predicting the manager's perception of whether the program increases employee productivity.						
Full Model	.037	4/81	.05	.777	.543	NS
Restricted Model	.0					
46. There is a significant difference in where the idea for the health promotion program originated in predicting the manager's perception of whether the program decreases employee absenteeism.						
Full Model	.056	4/81	.05	1.218	.309	NS
Restricted Model	.0					
47. There is a significant difference if the manager participates in the decision to establish a health promotion program in predicting the manager's perception of whether the program is cost effective.						
Full Model	.133	1/84	.05	12.923	.0005	NS
Restricted Model	.0					
48. There is a significant difference if the manager participates in the decision to establish a health promotion program in predicting the manager's perception of whether the program increases employee morale.						
Full Model	.018	1/84	.05	1.622	.206	NS
Restricted Model	.0					
49. There is a significant difference if the manager participates in the decision to establish a health promotion program in predicting the manager's perception of whether the program increases employee productivity.						
Full Model	.065	1/84	.05	5.885	.017	S
Restricted Model	.0					
50. There is a significant difference if the manager participates in the decision to establish a health promotion program in predicting the manager's perception of whether the program decreases employee absenteeism.						
Full Model	.058	1/84	.05	5.19	.025	S
Restricted Model	.0					

There is not a significant difference in where the idea for the health promotion program originated and the manager's perception of the program. Nor is there a significant difference in the managers' perception that the program increases employee morale, when the manager participates in the decision to establish the program. However, there is a significant difference in the managers' perception that the program was cost effective, increased employee productivity and decreased employee absenteeism when the manager participated in the decision to establish the program.

REFERENCES

Brady, J. P. (1983). Behavior, the environment and the health of the individual. Preventive Medicine, 12, 600-609.

Department of Health, Education, and Welfare. (1979). Healthy people: The Surgeon General's report on health promotion and disease prevention. Washington, D.C.: Government Printing Office.

Dillman, D. (1978). Mail and telephone surveys: the total design method. New York: John Wiley & sons.

Fielding, J. E. (1984). Health promotion and disease prevention at the worksite. Annual Review: Public Health, 5, 237-265.

Health maintenance. (1973). Newport Beach, CA: Pacific Mutual Insurance.

Likert, R. (1978). Profile of organizational characteristics. Ann Arbor, MI: R. Likert Associates.

Newman, I. (1976). Brief note on the justification for using multiple linear regression. Multiple Linear Regression Viewpoints, 6(4), 50-53.

Newman, I. & Newman, C. (1977). Conceptual statistics for beginners. (3rd ed.). Lanham, MD: University Press of America.

Ohio directory of manufacturers. (1986). Midland Park, NJ: Commerce Register.

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

Two Stage Smoothing of Scatterplots

Timothy H. Lee, Ph.D.
California State University

Donald T. Searls, Ph.D.
University of Northern Colorado

Abstract

Scatterplot smoothing is a simple but a very useful tool for data analysis. A smooth curve superimposed on the scatterplot greatly enhances the visual information, especially, the bivariate association between the prediction variable and the response variable. In this article some smoothers are reviewed with respect to consistency and sensitivity to discontinuities on the underlying functions. Robust centered span smoothers produce smooth and consistent curves but they tend to smooth over or blur the discontinuities. Non-centered span smoothers are sensitive to the discontinuities but they tend to be rough and lack consistency. Two stage smoothing is proposed as a technique that provides consistency as well as sensitivity to discontinuities.

Key words: smoother, underlying function, discontinuity, consistency, centered span, non-centered span

1. Introduction.

Scatterplots are a very useful tool for analyzing a bivariate relationship between two variables, say X and Y.

The observed bivariate data points,

$(x_1,y_1), (x_2,y_2), ..., (x_n,y_n),$

constitute acatterplots. They visually explain the relationship. It was pointed out by Cleveland (1979) that the extreme points in the point cloud of acatterplots diatract the eyes and they tend to miss the structure of the bulk of the data. As a remedy,scatterplots are smoothed, then the visual information is enhanced and the association between the two variables is clarified. Unfortunately, if discontinuities are present the smooth curve may tend to conceal this fact. If the smoothers are sensitive to discontinuities they tend to be somewhat rough. Two stage smoothing is proposed as a technique that tends to provide smooth fits with detection of discontinuities.

Scatterplot smoothing is a procedure that operates over the bivariate data points to decompose the observed y_i values into two parts, System (or Smooth) and Noise (or Rough). That is, the i-th observed value of Y can be written as

$$y_i = s(x_i) + r_i,$$

where s is a system or a smoothing function and r_i is a residual (or rough). Here, we assume that y_i is generated from an underlying function and noise with a certain distribution. That is,

$y_i = f(x_i) + e_i$.

The underlying function $f(x_i)$ is estimated by $s(x_i)$ in the smoothing procedure. The requirement of a good smoother is that it should not be affected by occasional outliers and the output results should be smooth regardless of the input data. In this regard, Cleveland (1979) proposed Locally Weighted Regression Scatterplot Smoothing ("LOWESS") which meets the robustness condition of good smoothers. Friedman (1984) proposed a variable span smoother in which local cross validation is used to estimate the optimal span as a function of the abscissa value. McDonald and Owen (1984) proposed a split linear fit smoothing algorithm that can produce discontinuous output. It can be used for smoothing with edge detection. One feature of the split linear fit method that distinguishes it from most of the other smoothers is that it uses non-centered spans.

One of the problems encountered in smoothing scatterplots is how to estimate, as closely as possible, the $f(x)$ by $s(x)$ using the given scatterplots. Therefore, a good smoother should be robust and consistent. When the underlying function, $f(x)$, is smooth (continuous) most of the centered span smoothers perform well. However, if $f(x)$ is discontinuous or kinked, the centered span smoothers usually blur the discontinuous points and produce a smooth curve; while the non-centered span smoothers are quite sensitive to discontinuities.

In this study, the smoothers sensitive to the discontinuities, namely, the non-centered span smoother, running medians of three, and Tukey's 3RSSH, are compared for consistency. Also, an exploration was made of a two-stage smoother that is more consistent but at the same time can produce a discontinuous curve.

For computational economy, the updating formula of the sample variance proposed by Chan, T. et al (1980) were used to update the regression parameter estimations.

Next, we discuss smoothers with two different types of spans and consider detection of the discontinuities of $f(x)$.

2. Centered Span Smoother.

The centered span smoother is the most commonly used smoother. To estimate $f(x_i)$ take a number of observations around x_i so that x_i is a center of the observations. These observations constitute a span for x_i. Cleveland's LOWESS, Running Median, Moving Average, and 3RSSH are examples of the centered span smoother. Here, as a centered span smoother, we use a robust fixed span smoother which is similar to LOWESS. The basic procedure is:

(a) Find initial fitted value $\hat{y}_i$ for x_i by using local linear regression.

Fit a simple local straight line to the data in the span for x_i, $i = 1,\ldots,n$.
Then, find the initial smooth value $\hat{y}_i$, $i = 1,\ldots,n$.(Updating formula can be used with unit weight.)

(b) Depending on the residual ($r_i = y_i - \hat{y}_i$) for each x_i, assign a weight.

A weight for each x_i is based on each r_i.

Let $m = \text{Median}\{|r_i|, i = 1,\ldots,n\}$, and let $d_i = r_i/(6*m)$.

Then, the weight for the k-th observation in the span for x_i will be

$$w_k(x_i) = \begin{cases} (1 - d_i^2)^2 & \text{for } |d_i| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Based on the new weight, fit a locally weighted straight regression line.

(d) Repeat steps (b) and (c) until the convergence criterion, $|\hat{y}_{old} - \hat{y}_{new}|/|\hat{y}_{old}| < \sigma$ is satisfied.
In this study, $\sigma = 10^{-5}$ is used.

This procedure is applied for three different sizes of spans in order to give points on the boundaries of the span less weight than the points in the center. So, three values (i.e.,$\hat{y}1_i$, $\hat{y}2_i$, $\hat{y}3_i$) for x_i are computed. The weight for each estimate is given depending on the span size. Let w1, w2, and w3 be weights for each of 3 spans. Then, the final smooth value for x_i will be obtained by,

$$\hat{y}_i = w1\hat{y}1_i + w2\hat{y}2_i + w3\hat{y}3_i,$$

where $w1 + w2 + w3 = 1$,

and

$w1 > w2 > w3$,

if the relationships among the spans are

span 1 < span 2 < span 3 .

In this study, the three spans used are 18, 20, and 22, respectively.

The advantages of this procedure are:

(a) It is computationally effective in terms of number of operations.

(b) It is more robust than a simple local straight line fit.

(c) Using a straight line reduces computational cost and makes the updating easier.

As seen in Figure 1, this smoother blurs the discontinuous points and produces an overall smooth curve. Running medians of three (referred to as "3R") and 3RSSH are also simple centered span smoothers. They are quite sensitive to discontinuities but produce rough (or bumpy) fits to the data.

3. Non-centered Span Smoother.

Unlike most of the smoothers, spans for x_i are not set up such that x_i is the center of a span. For example, McDonald and Owen's (1984) split linear fit smoother is such a smoother. They pointed out the weakness of the centered span smoothers and proposed a smoother that can be used for smoothing with edge detection. The idea is to make several linear fits for x_i; some of them are left-sided fits, some are central fits, and some are right-sided fits. In practice, three linear fits (one for each type of fit) are enough. Then, the three estimated values from the three types of fits are assessed depending on the basis of the mean squared residual about the line fitted over all of the data except x_i (referred to as "PMSE"). Any fitted value with PMSE greater than the average PMSE for x_i is ignored. Weights for the remaining fitted values are based on the squared differences between each PMSE and the average PMSE. Using these remaining fitted values and their respective weights, a weighted average is computed as a fitted value for x_i.

This smoother is very sensitive to discontinuities but there is a tendency for this smoother to produce a curve with a somewhat jagged appearance. This problem can be solved to some extent by applying the above algorithm repetitively to its own output. In this study, it is repeated once to avoid possible digression of the fitted curve from the underlying function f(x). See Figure 2. In this study, the span size for this smoother is 20.

4. Measurement of Consistencies.

To compare the consistencies of smoothers it is necessary to quantify them. A possible candidate to measure consistency is the average of the sample variances of the B fitted values for each x_i. Efron (1990) presented an example for a bootstrap estimate for the variance of regression coefficients. A similar idea is applied in this study as follows.

First, assuming that the underlying function is not known, apply a smoother on a generated data set and find

$s(x_i)$ and $r_i = y_i - s(x_i)$, $i = 1,\dots,n$.

Then,

(a) Construct $\hat{F}$ by assigning $1/n$ as the weight for the residual, r_i.

(b) Draw a bootstrap data set

$y_i^* = s(x_i) + r_i^*$, $i = 1,\dots,n$,

where r_i^*'s are i.i.d. from $\hat{F}$.

Then,

$s^*(x_i)$, $i = 1,\dots,n$

are computed on

y_i^*, $i = 1,\dots,n$.

(c) Independently repeat step (b) B times, obtaining bootstrap replications,

$s^{*1}(x_i), s^{*2}(x_i), \dots, s^{*B}(x_i)$, $i = 1,\dots,n$.

Then, compute

$$CM1 = \frac{1}{Bn}\sum_{b=1}^{B}\sum_{i=1}^{n}[s^{*b}(x_i) - s^{*}(x_i)]^2,$$

where

$$s^{*}(x_i) = \frac{1}{B}\sum_{b=1}^{B}[s^{*b}(x_i)].$$

And

$$CM2 = \frac{1}{Bn}\sum_{b=1}^{B}\sum_{i=1}^{n}[s^{*b}(x_i) - f(x_i)]^2$$

where f is the underlying function.

CM1 measures the consistencies (variation) of the smooth curve around the mean smooth curve and CM2 measures the consistencies around the underlying function. CM2 is measurable only when the underlying function is known. If the underlying function is known, it is more reasonable to use the e_i's rather than r_i's and $f(x_i)$ rather than $s(x_i)$ for step (c) in the above procedure to compare consistency. The reason is that the values of the r_i's depend on the sensitivity of smoothers to discontinuities. In Tables 1 - 4, such measures are computed for comparison of the consistency of smoothers.

5.Smoothing with Detection of the Discontinuities and Improved Consistency

We have seen that the non-centered span smoother is sensitive to the discontinuities, while the centered span smoothers blur them. By using this fact we can detect discontinuities simply by plotting the differences of the two smooth values estimated by the non-centered span smoother and by the centered span smoother. Figure 3 presents the two smooth curves for the purpose of visual comparison. The underlying function in Figure 3 is a sawtooth function. Figure 4 presents the difference plot. A discontinuity is suspected at the local maxima or minima. In the figure, a discontinuity is suspected around x = 50. Also, the difference plot shows the overall pattern of the discontinuity.

We are interested in consistency and, at the same time, in the detection of discontinuities. If a smoother has both properties, the computed values of CM1 and CM2 for that smoother will be lower than those of other smoothers. From Tables 1 - 4, we see that the robust centered span smoother has better consistency than the non-centered span smoother, but the latter has more sensitivity to discontinuities. The problem is how to combine the two desirable properties. One solution is to use two-stage smoothing. In the first step, discontinuities are located and the original data set is split such that each discontinuity serves as a splitting point. In the second step, the robust centered span smoother is applied to each of the split data sets. The consistency measurements of this smoother are shown in Tables 2 and 3 and the smooth curves produced by this method is shown in Figure 5.

6. Discussion.

In this study, the consistency measures of various smoothers are compared. The results show that:

(1) The non-centered span smoother is sensitive to discontinuities and less consistent than the robust centered span smoother,

(2) The robust centered span smoother lacks sensitivity to discontinuities but it is very consistent;

(3) Other sensitive smoothers, such as running medians of three or 3RSSH, produce quite rough curves and lack consistency; and

(4) The two-stage smoother is consistent and produces smooth curves with edge detection.

The detection and the location of the discontinuities on the x-axis are dependent upon the span size of the smoother. The determination of the span size is very important. If the span size is large, then the robust centered span smoother will blur the discontinuities. If x_i is close to a discontinuity, then the difference between the values estimated by the non-centered span smoother and the robust centered span smoother will be large. If the non-centered span smoother has a wide span it tends to ignore the discontinuities, while a narrow span will make it unnecessarily sensitive and may result in false detection of discontinuities. If there are more than one discontinuity on the underlying function the distance between any two discontinuities must be larger than the span size in order to be detected.

Sometimes outliers make the detection of discontinuity very difficult. Outliers near the discontinuities may cause confusion and lead to poor decisions. One possible remedy is to apply the running medians as a filter before the two-stage smoother is applied. The two-stage smoother works well when the discontinuities are separated enough and the functional form of the underlying function is not complicated. It works best when the underlying function is smooth but broken by discontinuities, for example, a saw tooth function. When no discontinuities are detected the two-stage smoother is the same as the robust centered span smoother. The two stage smoother has the advantages of being able to detect discontinuities as well as being very consistent.

Acknowledgements

The authors wish to acknowledge Dr.Sam Houston of University of Northern Colorado for his careful review of the manuscript. The authors also appreciate Arline Nakanishi and John Lichtenstein for their help in the preparation of the document.

References

Chan, T. F., Golub, G. H., & Leveque, R. J. (1980),"Updating formulae and pairwise algorithm for computing sample variances," DAAG29-78-G-0179.

Cleveland, W. S. (1979),"Robust locally weighted regression and smoothing scatterplots," Journal of the American Statistical Association, 74, 829-836.

Efron, B. (1990), "More Efficient Bootstrap Computations", Journal of the American Statistical Association, 85, 79-89.

Friedman, J. H. (1984), "A variable span smoother," LCS Technical Report No. 5.

Hoaglin, D. C., & Velleman, P. F. (1981). Applications, Basics, and computing of exploratory data analysis. Boston: Duxbury Press.

McDonald, J. A., & Owen, A. B. (1984). "Smoothing with split linear fits." Technical Report, Stanford University, Dept. of Statistics.

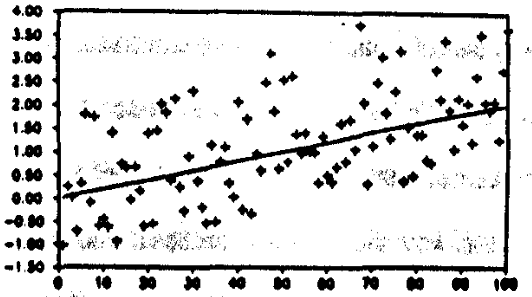
Mosteller, F., & Tukey, J. W. (1977). Data Analysis and Regression. Menlo, CA: Addison Wesley, Inc.

Appendix

A.Tables

Table-1.

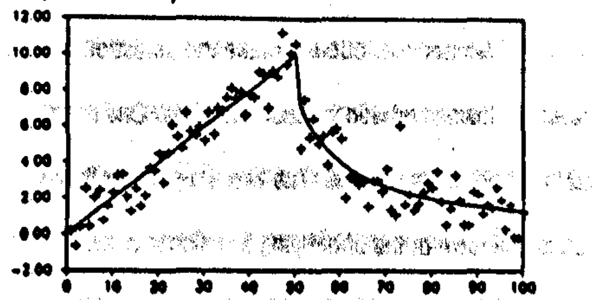
(f(x) =0.02x, monotone increasing function)



Smoother	CM1	CM2
~~Robust Centered Span~~	0.05721	0.07972
Non-centered Span	0.13593	0.17323
3RSSH	0.47333	0.64696
3R	0.49667	0.71916

Table-3.

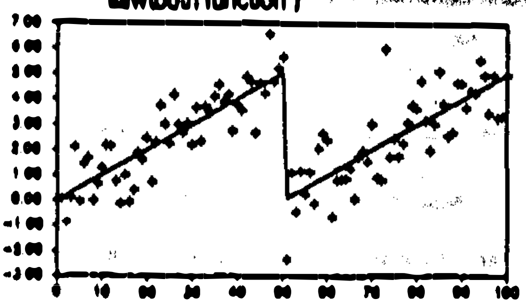
(f(x)=0.2x, for x≤50, f(x)=80/(x-40), for x>50, fin-shaped function)



Smoother	CM1	CM2
Robust Centered Span	0.06246	0.47491
Non-centered Span	0.20740	0.23504
3RSSH	0.37346	0.49477
3R	0.35296	0.51365
Two-stage	0.07766	0.10650

Table-2.

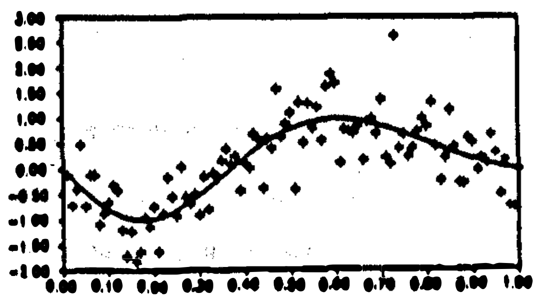
(f(x)=0.1x, for x≤50, f(x)=0.1x(x-50), for x>50, sawtooth function)



Smoother	CM1	CM2
~~Robust Centered Span~~	0.06462	0.76905
Non-centered Span	0.20362	0.39049
3RSSH	1.63010	2.14755
3R	1.66922	2.34655
Two-stage	0.05671	0.31377

Table-4.

(f(x)=sin[2π(1-x)^1.5], expandig cyclical function)



Smoother	CM1	CM2
~~Robust Centered Span~~	0.06246	0.47491
Non-centered Span	0.20740	0.23504
3RSSH	0.37346	0.49477
3R	0.35296	0.51365

B. Figures

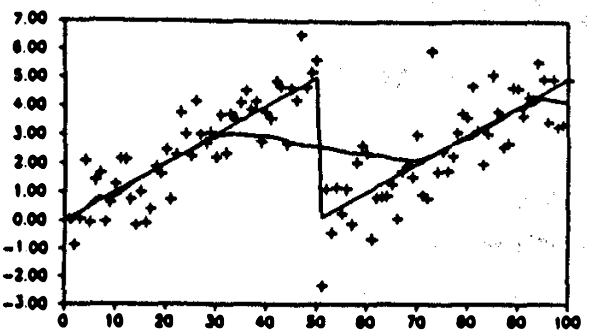


Figure 1. Smooth by Robust Centered Span Smoother.

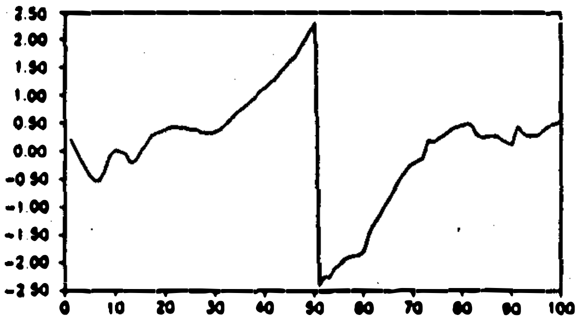


Figure 4. Differences of two smooth curves by Robust Centered Span smoother and Non-centered Span smoother

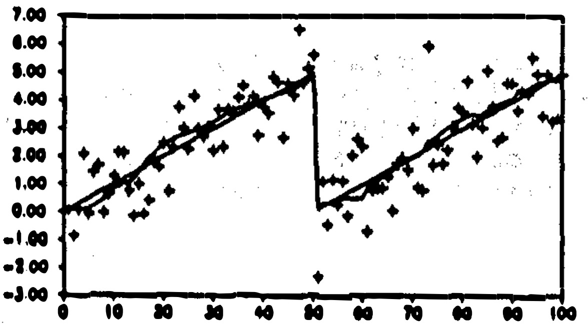


Figure 2. Smooth by Non-centered Span smoother.

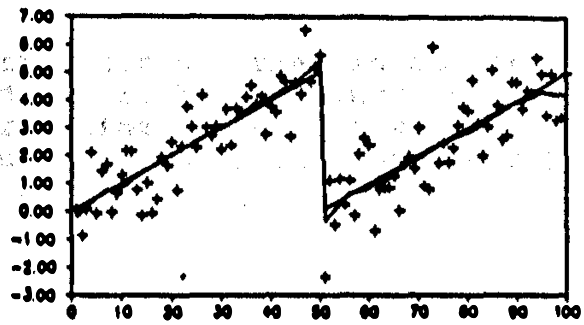


Figure 5. Smooth by Two-stage smoother

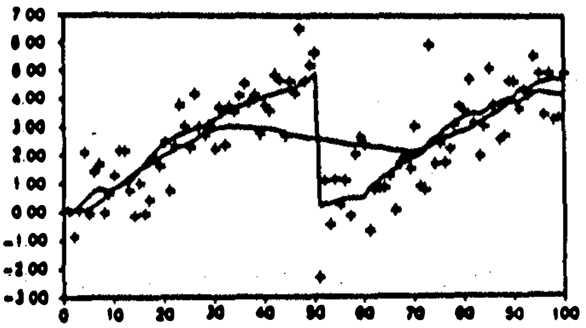


Figure 3. Comparison of two smooth curves by Robust Centered span smoother and Non-centered span smoother.

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

The Case Against Interpreting Regression Weights

Keith McNeil
University of New Mexico

Abstract

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

Introduction

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

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

Orthogonal Predictors

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

Non-Orthogonal Predictors

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

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

An extreme case of perfectly correlated predictors.

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

Control, or Upsetting the Prediction

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

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

Interpretation of Statistical Tests

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

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

Purpose of Using Multiple Predictors

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

The Inverted U Example

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

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

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

Appendix A

1) Draper and Smith (1981) p 117

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

2) Cooley and Lohnes (1962) p 40

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

3) Williams (1959) p 31-32.

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

4) Ward and Jennings (1973) pg 271.

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

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

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

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

5) Kerlinger and Pedhazur (1973) pg 63.

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

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

Bibliography

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

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

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

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

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

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

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

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

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

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

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

Testing Assumptions in Multiple Regression: Comparison of Procedures Available in SAS and SPSSX

Paula L. Woehlke

Patricia B. Elmore

Southern Illinois University Carbondale

Debra L. Spearing

State University of New York at Albany

That the use of multiple linear regression requires satisfying several assumptions has seldom been disputed. However, assessing whether one has met important assumptions is not always easy, and given the limited time available to instructors in a typical multiple regression course, the techniques available for checking assumptions are often not taught, or mentioned only briefly. The purpose of this paper is to compare the most easily available techniques for checking assumptions from two of the most popular statistics packages in use today, SAS (SAS Institute, 1985) and SPSS^X (SPSS, Inc., 1985). It is hoped that the attached examples will make the multiple regression course instructor's job easier by providing concrete examples of computer input and output that illustrate the testing of assumptions.

A condition that should be met for the use of multiple regression, but which is not, strictly speaking, an assumption, is that there be an absence of multicollinearity. Multicollinearity is defined as the existence of substantial correlation among a set of independent variables, and its presence creates three distinct problems:

- the substantive interpretation of partial regression coefficients,
- the sampling stability of these coefficients

and

- computational accuracy of the regression analysis.

Thus, although absence of multicollinearity is not a regression assumption, failure to assure that predictor variables are not multicollinear can result in faulty interpretations of analyses, regression equations that cannot be used for prediction, or both.

In terms of actual theoretical assumptions for using multiple regression analyses, errors of the prediction or residuals from estimated values of the regression provide the basis for assessing the adequacy of the model (Cohen & Cohen, 1983). Specifically, it is assumed that errors

(1) are normally distributed

(2) are independent of one another (that is, errors associated with one observation Y_i are not correlated with errors associated with any other observation Y_j)

(3) are identically distributed (that is, are sampled from the same distribution and have constant variances, also known as the assumption of homoscedasticity)

(4) have a mean of zero

and

(5) are uncorrelated with the independent variables (X's).

In addition to these assumptions about errors, it is further assumed that

(6) the independent variables, (X's) are fixed and measured without error

(7) the regression of Y on X is linear

and

(8) Y is a random variable composed of two components: a fixed component, a + bX, and a random error e_i.

Two conditions under which these assumptions about residuals fail to be met occur when

- the regression of Y on X (or X's) is curvilinear (so that condition 7 above is not met)

and

- there are one or more extreme residual values, known as "outliers, which not only make relatively large contributions to error or residual variance (thus reducing R^2) but also exert a disproportionately strong pull on the regression.

To illustrate the use of SAS and SPSS^X to test these assumptions, we used the (in)famous Longley data set. This data set has multicollinearity and some cases of univariate outliers through which to illustrate the diagnostic procedures available in both SAS and

SPSS^X. The following pages provide annotated output from these two packages, which we will describe in the next section.

Description of Output

The first assumption about errors is that the residuals are normally distributed. This assumption can be assessed by examining the residual scatterplot in Figure 4.SAS and the normal probability plot and statistical analyses shown in Figure 6.SAS; similar plots and statistics are produced by SPSS^X, as shown in Figure 4.SPSS^X, Figure 5.SPSS^X, and Figure 6.SPSS^X. If residuals are normally distributed, the plus signs (+'s) and the asterisks (*'s) will coincide in the SAS normal probability plot (or the asterisks [*'s] and dots [•'s] in the SPSS^X normal probability plot). Also, a statistical test for normality is provided in SAS in Figure 6.SAS; in this case, W:NORMAL = 0.948682, p=.471. It should be noted that SPSS^X's CONDESCRIPTIVE procedure routinely does not provide a comparable statistical test. All of these plots and tests from both SAS and SPSS^X indicate that the assumption about normally distributed residuals has been met.

That residuals are independent of one another or errors associated with one observation are not correlated with errors associated with any other observation is the second assumption to be tested. The Durbin-Watson D statistic shown in Figure 3.SAS and Figure 3.SPSS^X tests for nonindependence of errors when the order of cases is meaningful. For this data set, the Durbin-Watson D statistic is irrelevant. The residual scatterplots in Figure 4.SAS and Figure 5.SPSS^X show that the residuals are independent.

The third assumption is that residuals are identically distributed. This means that the errors are sampled from the same distribution and have constant variance, also known as homoscedasticity. Examination of the residual scatterplots in Figure 4.SAS and Figure 5.SPSS^X indicates that the assumption of homoscedasticity has been met.

Assumption 4, that the residuals have a mean of zero, can be determined by examining Figure 6.SAS or Figure 6.SPSS^X. For this data set, the mean is -7.421E-10 (Figure 6.SAS), which is considered zero for our purposes, or .000 (Figure 6.SPSS^X).

The correlation matrix showing the correlations between all of the independent variables and the residual should be used to assess assumption 5, that the residuals are uncorrelated with the independent variables. Examination of the correlation matrix for this data set, as found in Figure 7.SAS or Figure 7.SPSS^X indicates a correlation between each of the six independent variables and the residual equal to zero.

That the regression of Y on X is linear, assumption 7, can be determined by creating bivariate scatterplots for all predictors with the criterion. One example is shown in Figure 5.SAS and another in Figure 1.SPSS^X; both show the relation between Y and X_1. All six predictors in this data set are linearly related to the criterion Y.

Figure 1.SAS and Figure 2.SPSS^X show a check for multicollinearity. Low tolerance value and high condition number with large variance proportion for two or more variables may indicate multicollinearity. Variables X_5 and X_6 in this data set may be multicollinear with previous terms in the model.

Figure 2.SAS has two indices to check for outliers. A studentized residual value in excess of ±3.00 may indicate a univariate outlier (Tabachnick & Fidell, 1989, p. 67). Also, a data point with a Cook's Distance value greater than 1.00 is suspected of being an outlier. Cook's Distance is discussed in depth in Tabachnick & Fidell (1989), p. 130, and Kleinbaum, Kupper & Muller (1988), p. 201. Also note that in Figure 3.SPSS^X a similar casewise plot appears, as well as a listing and a histogram of standardized residuals.

Discussion

Although the output of the regression modules and related descriptive statistics procedures for SAS and SPSS^X are quite similar, there are a few differences worth noting. First, SPSS^X includes a histogram of standardized residuals to make the spotting of outliers somewhat easier; the program also has a normal probability plot that is a little easier to read than that provided in SAS. SPSS^X also provides standard errors for the skewness and kurtosis values for the variables analyzed in the CONDESCRIPTIVE module; these values are not printed in the SAS output. On the other hand, SAS provides a statistical test of normality when requested through PROC UNIVARIATE, as well as stem and leaf diagrams and boxplots of distributions, through the same PROC. It is also easy to obtain Cook's D values through SAS's PROC REG; it is somewhat more difficult to get similar statistics from SPSS^X, requiring the use of a RESIDUALS subcommand. In most other respects, output is comparable for the data and regression analyses shown here. For more advanced regression applications, it is somewhat easier to obtain leverage (partial regression residual) plots for general linear hypotheses, used in assessing degree of fit, nonfitting points, and multicollinearity (Sall, 1990) from SAS (via an option in PROC REG) than from SPSS^X, which produces "partial regression plots" through a PARTIALPLOT subcommand. It should be noted, however, that some anomalies recently have been detected in SAS's regression and GLM procedures for models using different types of intercept terms (see Uyar & Erdem, 1990). Finally, although it is somewhat more difficult to obtain several diagnostic statistics from SPSS^X, the package supplements its regression module with an extensive and flexible MANOVA procedure that allows one to easily build advanced regression models. With these advantages and disadvantages in mind, it should be possible for the reader to choose which computer package is most appropriate for a particular regression analysis.

References

Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

Draper, N. R. & Smith, H. (1981). *Applied Regression Analysis*, 2nd ed. New York: Wiley.

Kleinbaum, D. G., Kupper, L. L. & Muller, K. E. (1988). *Applied Regression Analysis and Other Multivariable Methods*, 2nd ed. Boston: PWS—Kent.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research*, 2nd ed. New York: Holt, Rinehart & Winston.

SAS Institute, Inc. (1985). *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: Sas Institute, Inc.

Sall, J. (1990). Leverage plots for general linear hypotheses. *The American Statistician*, *44*(4), 308-315.

SPSS, Inc. (1986). *SPSS X User's Guide, Edition* 2. New York: McGraw-Hill.

Tabachnick, B. G. & Fidell, L. S. (1989). *Using Multivariate Statistics*, 2nd ed. New York: Harper & Row.

Uyar, B. & Erdem, O. (1990). Regression procedures in SAS: Problems? *The American Statistician*, *44*(4), 296-301.

```
      SAS(R) LOG  CMS SAS 5.18        VM/CMS CMS USER GA0157

NOTE: COPYRIGHT (C) 1984,1986 SAS INSTITUTE INC., CARY, N.C.  27512, U.S.A.
NOTE: CMS SAS RELEASE 5.18 AT SOUTHERN ILLINOIS UNIVERSITY (01150002).

NOTE: CPUID   VERSION = FF  SERIAL = 023435  MODEL = 3081 .

 * you have questions, call the Computing Affairs Help Desk at 453-4361

   1 DATA AERA; INPUT Y X1 X2 X3 X4 X5 X6;
   2 OPTIONS NODATE LS=74;
   3 CARDS;

NOTE: DATA SET WORK.AERA HAS 16 OBSERVATIONS AND 7 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.06 SECONDS AND 24K.

  20 PROC REG;
  21      MODEL Y=X1 X2 X3 X4 X5 X6/STB P R CLI DW TOL VIF COLLIN;
  22      OUTPUT OUT=NEW
  23             P=PRED
  24             R=RES;
NOTE: THE DATA SET WORK.NEW HAS 16 OBSERVATIONS AND 9 VARIABLES.
NOTE: THE PROCEDURE REG USED 0.18 SECONDS AND 40K
      AND PRINTED PAGES 1 TO 2.

  25 PROC PLOT;
  26      PLOT PRED=RES Y*X1 Y*X2 Y*X3 Y*X4 Y*X5 Y*X6;
NOTE: THE PROCEDURE PLOT USED 0.23 SECONDS AND 52K
      AND PRINTED PAGES 3 TO 9.

  27 PROC UNIVARIATE NORMAL PLOT;
NOTE: THE PROCEDURE UNIVARIATE USED 0.23 SECONDS AND 64K
      AND PRINTED PAGES 10 TO 18.

  28 PROC CORR;
NOTE: THE PROCEDURE CORR USED 0.11 SECONDS AND 76K
      AND PRINTED PAGES 19 TO 21.
NOTE: SAS USED 76K MEMORY.
```

```
SAS

DEP VARIABLE: Y
ANALYSIS OF VARIANCE

                        SUM OF          MEAN
SOURCE      DF         SQUARES        SQUARE      F VALUE       PROB>F

MODEL        6       184172402   30695400.32      330.285       0.0001
ERROR        9       836424.06   92936.00616
C TOTAL     15       185008826

    ROOT MSE          304.8541     R-SQUARE        0.9955
    DEP MEAN             65317     ADJ R-SQ        0.9925
    C.V.             0.4667301

PARAMETER ESTIMATES

                     PARAMETER      STANDARD     T FOR H0:
VARIABLE    DF        ESTIMATE         ERROR   PARAMETER=0      PROB > |T|

INTERCEP     1    -3482258.63     890420.38        -3.911          0.0036
X1           1    15.06187236   84.91492578         0.177          0.8631
X2           1    -0.03581918    0.03349101        -1.070          0.3127
X3           1    -2.02022980    0.48839960        -4.136          0.0025
X4           1    -1.03322687    0.21427416        -4.822          0.0009
X5           1    -0.05110411    0.22607320        -0.226          0.8262
X6           1     1829.15146     455.47850         4.016          0.0030

                STANDARDIZED                        VARIANCE
VARIABLE    DF      ESTIMATE       TOLERANCE       INFLATION

INTERCEP     1             0               .               0
X1           1    0.04628202     0.007378307       135.53244
X2           1   -1.01374635     0.000559124      1788.51348
X3           1   -0.53754258      0.02974518     33.61889060
X4           1   -0.20474069      0.27863456      3.58893019
X5           1   -0.10122111     0.002505317       399.15102
X6           1    2.47966438     0.001317557       758.98060

COLLINEARITY DIAGNOSTICS

                          CONDITION   VAR PROP  VAR PROP  VAR PROP  VAR PROP
NUMBER   EIGENVALUE          NUMBER   INTERCEP  X1        X2        X3

     1     6.861393        1.000000     0.0000    0.0000    0.0000    0.0000
     2     0.082103        9.141721     0.0000    0.0000    0.0000    0.0143
     3     0.045681       12.255735     0.0000    0.0000    0.0003    0.0008
     4     0.010688       25.336607     0.0000    0.0003    0.0011    0.0646
     5    0.0001292         230.424     0.0000    0.4568    0.0157    0.0056
     6    6.246E-06        1048.080     0.0001    0.5046    0.3284    0.2253
     7    3.664E-09       43275.045     0.9999    0.0383    0.6546    0.6893
```

FIGURE 1. SAS MULTICOLLINEARITY
Low tolerance value and high condition number with large variance proportion for two or more variables may indicate multicollinearity. Variables X_5 and X_6 may be multicollinear with previous terms in the model.

NUMBER	VAR PROP X4	VAR PROP X5	VAR PROP X6
1	0.0004	0.0000	0.0000
2	0.0919	0.0000	0.0000
3	0.0634	0.0000	0.0000
4	0.4267	0.0000	0.0000
5	0.1154	0.0097	0.0000
6	0.0000	0.8306	0.0002
7	0.3020	0.1597	0.9998

OBS	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% PREDICT	UPPER95% PREDICT	RESIDUAL
1	60323.0	60055.7	198.6	59232.6	60878.8	267.3
2	61122.0	61216.0	229.1	60353.3	62078.7	-94.0139
3	60171.0	60124.7	183.4	59319.9	60929.6	46.2872
4	61187.0	61597.1	186.0	60789.3	62405.0	-410.1
5	63221.0	62911.3	239.2	62034.7	63787.8	309.7
6	63639.0	63888.3	185.3	63081.2	64695.4	-249.3
7	64989.0	65153.0	213.7	64310.8	65995.3	-164.0
8	63761.0	63774.2	216.6	62928.2	64620.1	-13.1804
9	66019.0	66004.7	206.1	65172.2	66837.2	14.3048
10	67857.0	67401.6	175.3	66606.1	68197.1	455.4
11	68169.0	68186.3	182.9	67382.1	68990.5	-17.2609
12	66513.0	66552.1	211.9	65712.2	67391.9	-39.0550
13	68655.0	68810.5	186.5	68002.1	69619.0	-155.5
14	69564.0	69649.7	145.7	68885.3	70414.0	-85.6713
15	69331.0	68989.1	186.2	68181.0	69797.1	341.9
16	70551.0	70757.8	253.0	69861.6	71653.9	-206.8

OBS	STD ERR RESIDUAL	STUDENT RESIDUAL	-2 -1 0 1 2	COOK'S D
1	231.3	1.1560	| |** |	0.141
2	201.1	-0.4676	| | |	0.041
3	243.5	0.1901	| | |	0.003
4	241.5	-1.6979	| ***| |	0.244
5	189.0	1.6304	| |*** |	0.614
6	242.1	-1.0300	| **| |	0.009
7	217.4	-0.7547	| *| |	0.079
8	214.6	-0.0614	| | |	0.001
9	224.6	0.0637	| | |	0.000
10	249.4	1.8258	| |*** |	0.235
11	243.9	-0.0708	| | |	0.000
12	219.2	-0.1782	| | |	0.004
13	241.1	-0.6451	| *| |	0.036
14	267.8	-0.3199	| | |	0.004
15	241.4	1.4163	| |** |	0.170
16	170.1	-1.2154	| **| |	0.467

FIGURE 2.SAS OUTLIERS

Check STUDENTIZED RESIDUAL for values in excess of ±3.00 for univariate outliers. See Tabachnick & Fidell (1989), p. 67.

COOK'S DISTANCE values greater than 1.00 are suspected of being outliers. See Tabachnick & Fidell (1989), p. 130 and Kleinbaum, Kupper, & Muller (1988), p. 201.

```
UM OF RESIDUALS                 -1.18744E-08
UM OF SQUARED RESIDUALS             836424.1
REDICTED RESID SS (PRESS)            2886893

URBIN-WATSON D                  2.559
FOR NUMBER OF OBS.)                16
ST ORDER AUTOCORRELATION -0.348
```

FIGURE 3.SAS INDEPENDENCE OF RESIDUALS

The Durbin-Watson D statistic tests for nonindependence of errors when the order of cases is meaningful. Tables are found in Draper & Smith (1981), pp. 164-166. It is irrelevant for this data set.

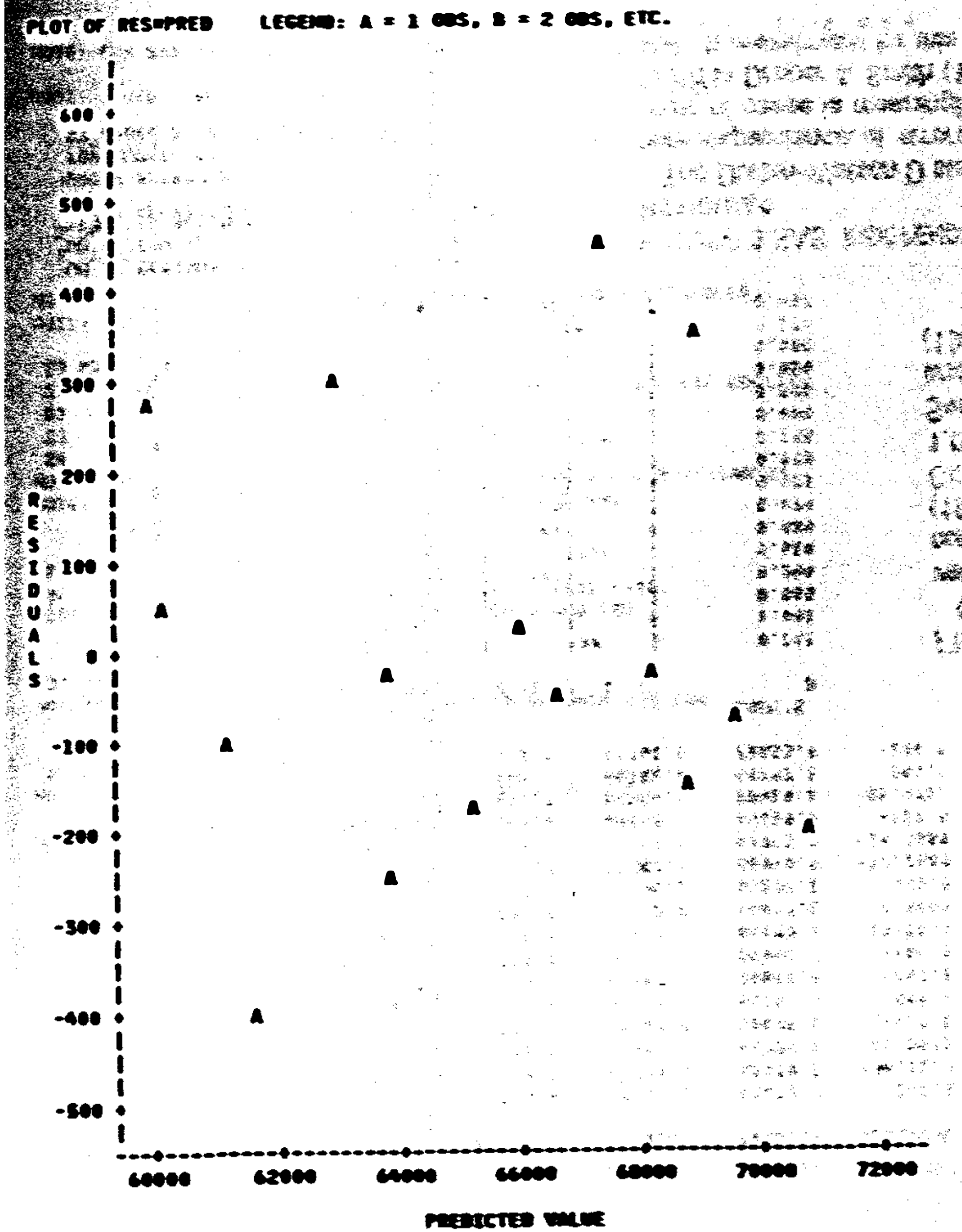


FIGURE 4. SAS NORMALITY, LINEARITY, HOMOSCEDASTICITY, AND INDEPENDENCE OF RESIDUALS

Examine the residual scatterplot to assess all four assumptions. All four assumptions are met in this data set.

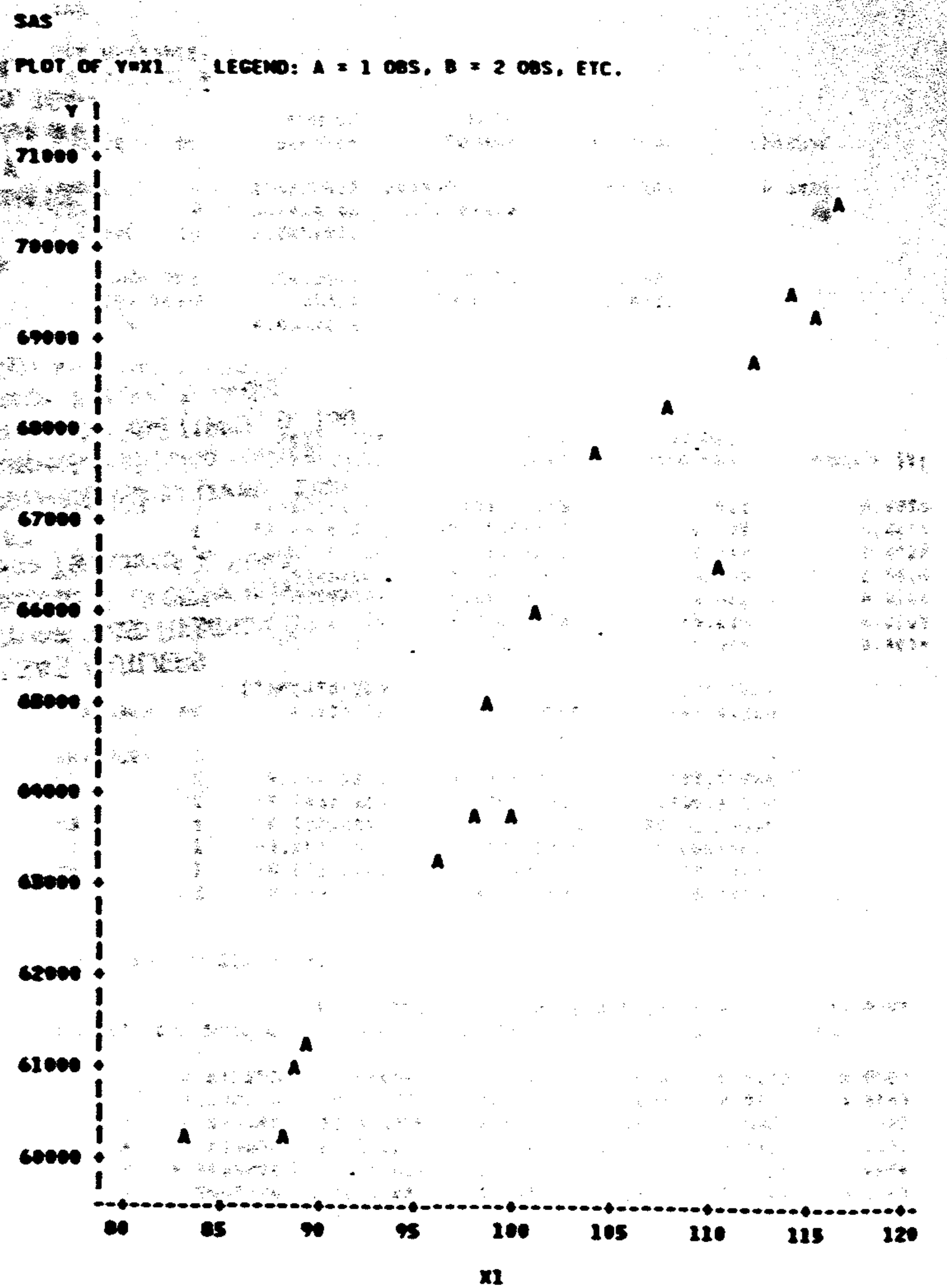


FIGURE 5. SAS LINEARITY

Use bivariate scatterplots to assess linearity of predictor - criterion association.

```
UNIVARIATE

VARIABLE=RES                RESIDUALS

                  MOMENTS

N                16  SUM WGTS         16
MEAN     -7.421E-10  SUM       -1.187E-08
STD DEV     236.139  VARIANCE     55761.6
SKEWNESS   0.464739  KURTOSIS   -0.298894
USS          836424  CSS           836424
CV           -99999  STD MEAN     59.0347
T:MEAN=0 -1.257E-11  PROB>|T|           1
SGN RANK         -6  PROB>|S|    0.776105
NUM ^= 0         16
W:NORMAL   0.948682  PROB<W         0.471


            QUANTILES(DEF=4)                        EXTREMES

100% MAX    455.394     99%      455.394       LOWEST    HIGHEST
 75% Q3     212.077     95%      455.394     -410.115    46.2872
 50% MED    -28.162     90%       375.97     -249.311     267.34
 25% Q1    -161.924     10%     -297.552     -206.758    309.715
  0% MIN   -410.115      5%     -410.115     -160.049    341.932
                         1%     -410.115      -155.55    455.394
 RANGE      865.509
 Q3-Q1      374.001
 MODE      -410.115

STEM LEAF                          #  BOXPLOT
   0 5                             1     |
   0 333                           3  +--+--+
  -0 422221100000                 12  *-----*
     ----+----+----+----+
 MULTIPLY STEM.LEAF BY 10**+03

                        NORMAL PROBABILITY PLOT
  750+                                                      ++++
     |                       ++++++*+*+*+*+*+*+*+*+++++++
 -250+   ***+*+*+*+*+*+*+*+*+*+** **
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

FIGURE 6. SAS NORMALITY OF RESIDUALS

If residuals are normally distributed, the plus signs (+'s) and asterisks (*'s) should coincide in the normal probability plot. A statistical test for normality is also provided. In this case, W:NORMAL = 0.948682, p = .471.

SKEWNESS can be detected by observing the stem and leaf and the boxplot as well as the skewness index of .464739 indicating the residuals are positively skewed. The kurtosis value of -.298894 indicates the residuals are platykurtic. Statistical tests of skewness and kurtosis are discussed in Tabachnick & Fidell (1989), pp. 72-73.

For these data,

$$z_{SKEWNESS} = \frac{S - 0}{\sqrt{\frac{6}{N}}} = \frac{.464739}{\sqrt{\frac{6}{16}}} = .76$$

$z_{SKEWNESS} < 1.96 \therefore$ not skewed.

$$z_{KURTOSIS} = \frac{K - 0}{s_K} = \frac{K - 0}{\sqrt{\frac{24}{N}}} = \frac{-.298894}{\sqrt{\frac{24}{16}}} = -.24$$

$z_{KURTOSIS} > -1.96 \therefore$ no kurtosis

SAS

VARIABLE	N	MEAN	STD DEV	SUM	MINIMUM	MAXIMUM
Y	16	65317.0	3511.97	1045072	60171.0	70551.0
X1	16	101.7	10.79	1627	83.0	116.9
X2	16	387698.4	99394.94	6203175	234289.0	554894.0
X3	16	3193.3	934.46	51093	1870.0	4896.0
X4	16	2606.7	695.92	41707	1456.0	3594.0
X5	16	117424.0	6956.10	1878784	107600.0	130081.0
X6	16	1954.5	4.76	31272	1947.0	1962.0
PRED	16	65317.0	3504.02	1045072	60055.7	70757.8
RES	16	-7.421E-10	236.14	-1.187E-08	-410.1	455.4

SAS

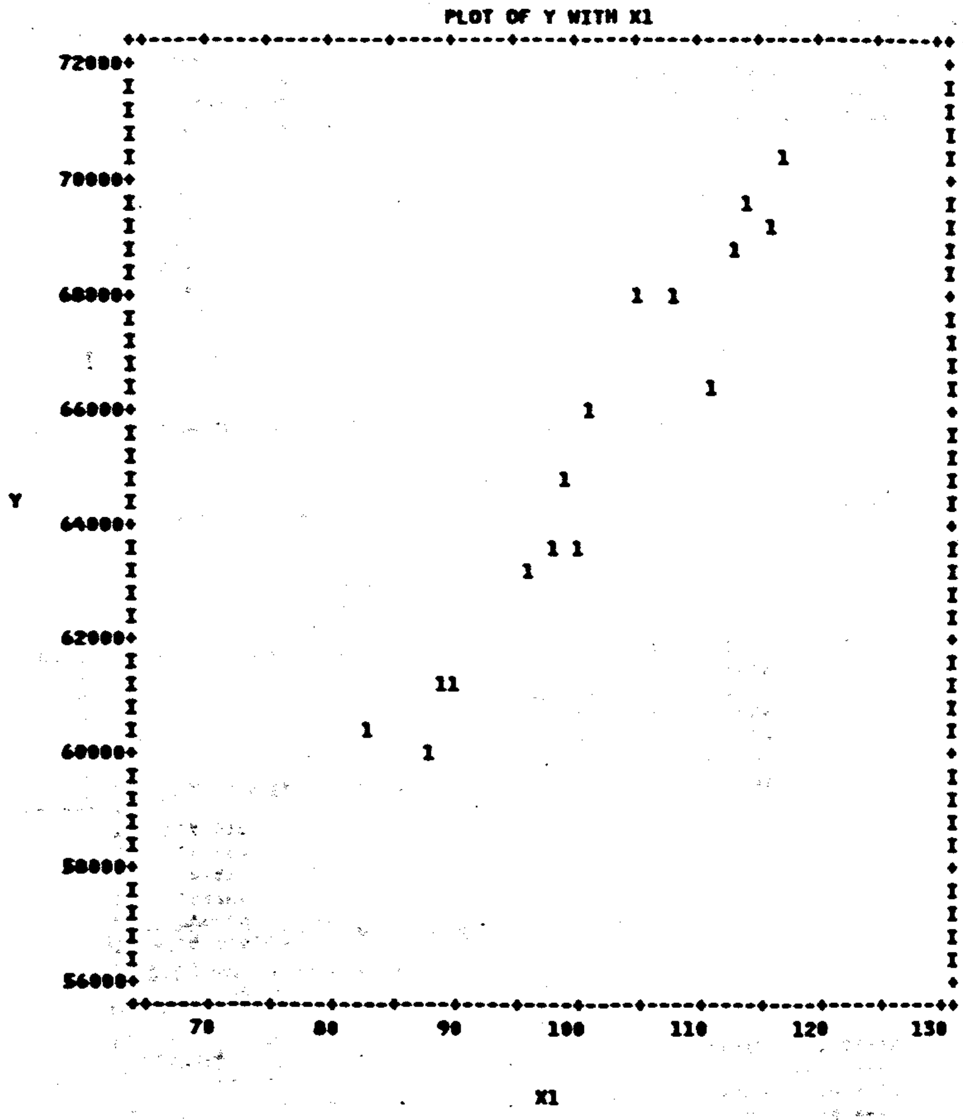
PEARSON CORRELATION COEFFICIENTS / PROB > |R| UNDER H0:RHO=0 / N = 16

	Y	X1	X2	X3	X4	X5
Y	1.00000 0.0000	0.97090 0.0001	0.98355 0.0001	0.50250 0.0473	0.45731 0.0749	0.96039 0.0001
X1	0.97090 0.0001	1.00000 0.0000	0.99159 0.0001	0.62063 0.0103	0.46474 0.0697	0.97916 0.0001
X2	0.98355 0.0001	0.99159 0.0001	1.00000 0.0000	0.60426 0.0132	0.44644 0.0830	0.99109 0.0001
X3	0.50250 0.0473	0.62063 0.0103	0.60426 0.0132	1.00000 0.0000	-0.17742 0.5109	0.68655 0.0033
X4	0.45731 0.0749	0.46474 0.0697	0.44644 0.0830	-0.17742 0.5109	1.00000 0.0000	0.36442 0.1652
X5	0.96039 0.0001	0.97916 0.0001	0.99109 0.0001	0.68655 0.0033	0.36442 0.1652	1.00000 0.0000
X6	0.97133 0.0001	0.99115 0.0001	0.99527 0.0001	0.66826 0.0047	0.41725 0.1078	0.99395 0.0001
PRED PREDICTED VALUE	0.99774 0.0001	0.97310 0.0001	0.98578 0.0001	0.50364 0.0467	0.45834 0.0742	0.96257 0.0001
RES RESIDUALS	0.06724 0.8046	-0.00000 1.0000	-0.00000 1.0000	-0.00000 1.0000	-0.00000 1.0000	-0.00000 1.0000

	X6	PRED	RES
Y	0.97133 0.0001	0.99774 0.0001	0.06724 0.8046
X1	0.99115 0.0001	0.97310 0.0001	-0.00000 1.0000
X2	0.99527 0.0001	0.98578 0.0001	-0.00000 1.0000
X3	0.66826 0.0047	0.50364 0.0467	-0.00000 1.0000
X4	0.41725 0.1078	0.45834 0.0742	-0.00000 1.0000
X5	0.99395 0.0001	0.96257 0.0001	-0.00000 1.0000
X6	1.00000 0.0000	0.97353 0.0001	-0.00000 1.0000
PRED PREDICTED VALUE	0.97353 0.0001	1.00000 0.0000	-0.00000 1.0000
RES RESIDUALS	-0.00000 1.0000	-0.00000 1.0000	1.00000 0.0000

FIGURE 7. SAS CORRELATION OF ERRORS AND INDEPENDENT VARIABLES
Use the correlation matrix to determine association between each independent variable and the residual from the multiple regression equation.

```
For IBM VM/XA           Southern Illinois University      License Number 19877
This software is functional through June 30, 1991.

Try the new SPSS-X Release 3.1 features:

* Interactive SPSS-X command execution        * The new RANK procedure
* Online Help                                 * Improvements in:
* Nonlinear Regression                        *    REPORT and TABLES
* Time Series and Forecasting (TRENDS)        *    Simplified Syntax
* Macro Facility                              *    Matrix I/O

See SPSS-X User's Guide, Third Edition, for more information on these features.

   1  0  DATA LIST FREE / Y X1 X2 X3 X4 X5 X6
   2  0  SET WIDTH =80
   3  BEGIN DATA
  19  END DATA

Preceding task required .02 seconds CPU time;  .03 seconds elapsed.

  20  PLOT PLOT=Y WITH X1/

There are 1,749,224 bytes of memory available.
The largest contiguous area has 1,749,224 bytes.

PLOT requires 15000 bytes of workspace for execution.


                                                                  Page    2
17-Oct-90 SPSS-X RELEASE 3.1 FOR IBM VM/CMS
09:47:37  Southern Illinois University   IBM 3081          IBM VM/XA

* * * * * * * * * * * * * * * * *   P L O T   * * * * * * * * * * * * * * * * *

Data  Information

      16 unweighted cases accepted.

Size of the plots

   Horizontal size is 65
     Vertical size is 40

Frequencies and symbols used (not applicable for control or overlay plots)

        1 - 1     11 - B     21 - L     31 - V
        2 - 2     12 - C     22 - M     32 - W
        3 - 3     13 - D     23 - N     33 - X
        4 - 4     14 - E     24 - O     34 - Y
        5 - 5     15 - F     25 - P     35 - Z
        6 - 6     16 - G     26 - Q     36 - *
        7 - 7     17 - H     27 - R
        8 - 8     18 - I     28 - S
        9 - 9     19 - J     29 - T
       10 - A     20 - K     30 - U
```



16 cases plotted.

FIGURE 1. SPSS^X. LINEARITY
Use bivariate scatterplots to assess linearity of predictor-criterion association.

```
17-Oct-90 SPSS-X RELEASE 3.1 FOR IBM VM/CMS                                   Page
09:47:37  Southern Illinois University    IBM 3081          IBM VM/XA

Preceding task required .05 seconds CPU time;  .44 seconds elapsed.

  21  REGRESSION VARS=Y X1 X2 X3 X4 X5 X6/STATISTICS=ALL F/DEP=Y
  22            /METHOD=ENTER
  23            /RESID=DEFAULTS
  24            /CASEWISE=ALL DEFAULTS
  25            /SCATTERPLOT=(*ZRESID,*ZPRED)
  26            /SAVE PRED (YHAT) RESID (ERR)

There are 1,749,184 bytes of memory available.
The largest contiguous area has 1,740,832 bytes.
```

*** * * *   M U L T I P L E   R E G R E S S I O N   * * * ***

Listwise Deletion of Missing Data

Equation Number 1   Dependent Variable..   Y

Beginning Block Number 1. Method: Enter

Variable(s) Entered on Step Number
1.. X6
2.. X4
3.. X3
4.. X1
5.. X5
6.. X2

Multiple R	.99774		
R Square	.99548	R Square Change	.99548
Adjusted R Square	.99247	F Change	330.28534
Standard Error	304.88407	Signif F Change	.0000

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	6	184172401.94449	30695400.32408
Residual	9	836420.05551	92936.00617

F =   330.28534   Signif F = .0000

AIC	187.82004
PC	.01155
CP	7.00000
SBC	193.23696

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal: Covariance   Above: Correlation

	X6	X4	X3	X1	X5	X2
X6	207460.663	-.54937	-.82418	.18628	.30816	-.80168
X4	-53.61674	.04591	.61857	-.34081	-.18091	.46860
X3	-183.32591	.06473	.23053	-.55500	-.75026	.94561
X1	7204.91263	-6.34671	-23.01719	7210.54462	.65918	-.64942
X5	39.96940	-.00915	-.08372	12.65424	.05111	-.83321
X2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

*** * * *   M U L T I P L E   R E G R E S S I O N   * * * ***

Equation Number 1   Dependent Variable..   Y

XTX Matrix

	X6	X4	X3	X1	X5	X2
X6	758.98060	-28.67217	-131.6399	59.74668	213.64593	-934.0346
X4	-28.67217	3.58893	6.79454	-7.69306	-7.15017	37.54356
X3	-131.6399	6.79454	33.61889	-37.46332	-87.83677	231.87218
X1	59.74668	-7.69306	-37.46332	135.53244	153.31803	-319.7367
X5	213.64593	-7.15017	-87.83677	153.31803	399.15102	-703.9908
X2	-934.0346	37.54356	231.87218	-319.7367	-703.9908	1788.5135
Y	2.47966	-.20474	-.53754	.04628	-.10122	-1.01375

	Y
X6	-2.47966
X4	.20474
X3	.53754
X1	-.04628
X5	.10122
X2	1.01375
Y	.00452

Equation Number 1   Dependent Variable.. Y

---------------------- Variables in the Equation ----------------------

Variable	B	SE B	95% Confdnce Intrvl B		Beta
X6	1829.151465	455.470499	798.780692	2859.514237	2.479664
X4	-1.033227	.214274	-1.517948	-.548506	-.204741
X3	-2.020230	.400400	-3.125066	-.915394	-.537543
X1	15.061872	84.914926	-177.028016	207.152560	.046282
X5	-.051104	.226073	-.562517	.460308	-.101221
X2	-.035819	.033491	-.111581	.039943	-1.015746
(Constant)	-3402258.635	090420.3836	-5496527.183	-1467990.086	

------------------------- Variables in the Equation -------------------------

Variable	SE Beta	Correl	Part Cor	Partial	Tolerance	VIF	F
X6	.617463	.971329	.090087	.001140	.001318	758.981	16.127
X4	.042460	.457307	-.100074	-.049004	.278635	3.589	23.252
X3	.129953	.502498	-.092709	-.009509	.029745	33.619	17.110
X1	.260926	.970099	.003975	.059022	.007378	135.532	.031
X5	.447780	.968391	-.005066	-.075137	.002505	399.151	.051
X2	.947855	.983662	-.023971	-.335004	5.591E-04	1788.513	1.144
(Constant)							15.294

------ in -------

Variable	Sig F
X6	.0030
X4	.0009
X3	.0025
X1	.8631
X5	.8262
X2	.3127
(Constant)	.0036

Number	Eigenval	Cond Index	Variance Proportions Constant	X1	X2	X3	X4	
1	6.86139	1.000	.00000	.00000	.00000	.00004	.00035	.0(
2	.08210	9.142	.00000	.00000	.00001	.01428	.09191	.0(
3	.04560	12.256	.00000	.00000	.00026	.00004	.06357	.0(
4	.01069	25.337	.00000	.00034	.00107	.06464	.42672	.0(
5	.00013	230.424	.00000	.45677	.01564	.00559	.11540	.0(
6	.00001	1048.080	.00015	.50456	.32839	.22534	.00000	.8:
7	.00000	43275.05	.99985	.03833	.65463	.60926	.30205	.1!

	X6
1	.00000
2	.00000
3	.00000
4	.00000
5	.00000
6	.00016
7	.99904

End Block Number 1   All requested variables entered.

■ ■

Summary table

Step	MultR	Rsq	F(Eqn)	SigF		Variable	BetaIn
1					In:	X6	.9713
2					In:	X4	.0630
3					In:	X3	-.3910
4					In:	X1	-.0224
5					In:	X5	-.5002
6	.9977	.9955	330.285	.000	In:	X2	-1.0137

FIGURE 2 SPSS^X. MULTICOLLINEARITY
Low tolerance value and high condition number with large variance proportion for two or more variables may indicate multicollinearity. Variables X_5 and X_6 may be multicollinear with previous terms in the model.

```
           * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   Y

Casewise Plot of Standardized Residual

*: Selected    M: Missing

            -3.0      0.0       3.0
 Case #     O:........:........:O            Y          *PRED         *RESID
      1     .         .   *    .      60323.00     60055.6600       267.3400
      2     .        *.        .      61122.00     61216.0139       -94.0139
      3     .         *        .      60171.00     60124.7128        46.2872
      4     .     *   .        .      61187.00     61597.1146      -410.1146
      5     .         .   *    .      63221.00     62911.2854       309.7146
      6     .       * .        .      63639.00     63888.3112      -249.3112
      7     .       * .        .      64989.00     65153.0490      -164.0490
      8     .         *        .      63761.00     63774.1004       -13.1004
      9     .         *        .      66019.00     66004.6952        14.3048
     10     .         .    *   .      67857.00     67401.6059       455.3941
     11     .         *        .      68169.00     68186.2689       -17.2689
     12     .         *        .      66513.00     66552.0550       -39.0550
     13     .       * .        .      68655.00     68810.5500      -155.5500
     14     .        *.        .      69564.00     69649.6713       -85.6713
     15     .         .   *    .      69331.00     68989.0685       341.9315
     16     .       * .        .      70551.00     70757.7578      -206.7578
 Case #     O:........:........:O            Y          *PRED         *RESID
            -3.0      0.0       3.0
```

```
           * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   Y

Residuals Statistics:

                 Min            Max        Mean    Std Dev    N

*PRED     60055.6563     70757.7500  65317.0000  3504.0206   16
*RESID     -410.1145       455.3940       .0000   256.1389   16
*ZPRED       -1.5015         1.5527       .0000     1.0000   16
*ZRESID      -1.3453         1.4958       .0000      .7746   16

Total Cases =        16

Durbin-Watson Test =    2.89949
```

```
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Outliers - Standardized Residual

   Case #         *ZRESID

       10         1.49581
        4        -1.34528
       15         1.12162
        5         1.01594
        1          .87694
        6         -.81781
       16         -.67822
        7         -.53812
       13         -.51024
        2         -.30839
```

```
Histogram - Standardized Residual

NExp N        (* = 1 Cases,     . : = Normal Curve)
 0  .01   Out
 0  .02  3.00
 0  .06  2.67
 0  .14  2.33
 0  .29  2.00
 0  .53  1.67 .
 1  .88  1.33 :
 3 1.29  1.00 :**
 0 1.70   .67  .
 0 2.00   .33  .
 5 2.12   .00 *:***
 2 2.00  -.33 *:
 4 1.70  -.67 *:**
 0 1.29 -1.00 .
 1  .88 -1.33 :
 0  .53 -1.67 .
 0  .29 -2.00
 0  .14 -2.33
 0  .06 -2.67
 0  .02 -3.00
 0  .01   Out
```

FIGURE 3. SPSS^X. OUTLIERS
Check STUDENTIZED RESIDUAL for values in excess of ±3.00 for univariate outliers. See Tabachnick & Fidell (1989), p. 67. ~~Also check~~ plot of STANDARDIZED RESIDUAL. Also note that the Durbin-Watson D statistic tests for nonindependence of errors when the order of cases is meaningful. Tables are found in Draper & Smith (1981), pp. 164-166. It is irrelevant for this data set.

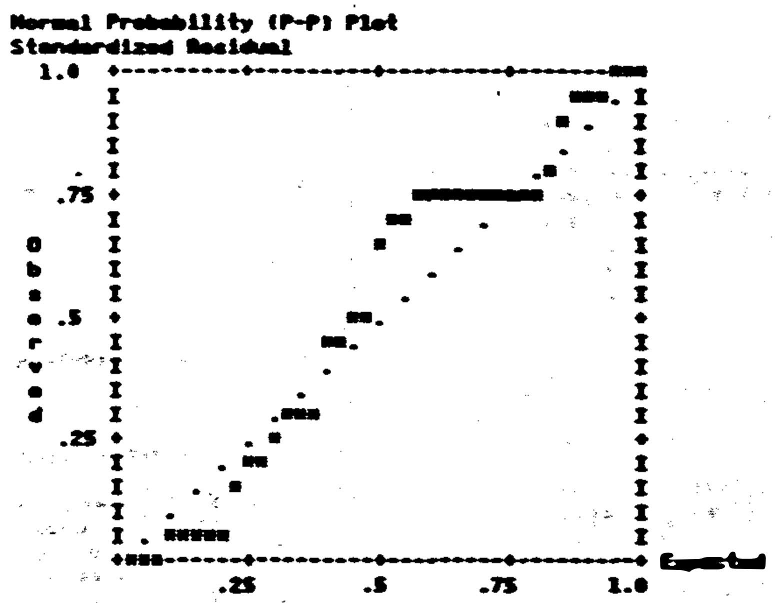


FIGURE 4.SPSS^X. NORMALITY OF RESIDUALS

If residuals are normally distributed, the dots (•'s) and asterisks (*'s) should coincide in the NORMAL PROBABILITY PLOT.

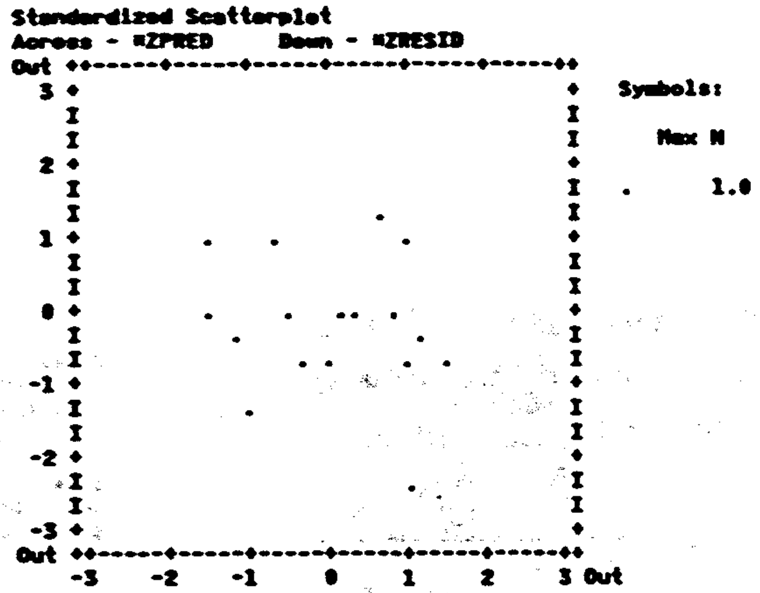


FIGURE 5.SPSS^X NORMALITY, LINEARITY, HOMOSCEDASTICITY, AND INDEPENDENCE OF RESIDUALS

Examine the standardized scatterplot of predicted and residual values to assess all four assumptions. All are met in this data set.

```
17-Oct-90 SPSS-X RELEASE 3.1 FOR IBM VM/CMS                          Page  14
09:47:39  Southern Illinois University   IBM 3081         IBM VM/XA

Preceding task required .19 seconds CPU time;  1.83 seconds elapsed.

  27  CONDESCRIPTIVE YHAT ERR/
  28  STATISTICS ALL

>Warning # 11003
>The new default column-style printing cannot be used for this DESCRIPTIVES, as
>there are too many statistics to print on one line per variable.  Old style
>printing will be used instead.


There are 1,748,296 bytes of memory available.
The largest contiguous area has 1,746,720 bytes.


148 bytes of memory required for the DESCRIPTIVES procedure.
4 bytes have already been acquired.
144 bytes remain to be acquired.


Number of valid observations (listwise) =        16.00

Variable  YHAT        Predicted Value

Mean         65317.000            S.E. Mean       876.005
Std Dev       3504.021            Variance   12278160.130
Kurtosis        -1.290            S.E. Kurt         1.091
Skewness         -.101            S.E. Skew          .564
Range        10702.090            Minimum     60055.65997
Maximum    70757.75783            Sum         1045072.000

Valid observations -      16      Missing observations -       0

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Variable  ERR         Residual

Mean              .000            S.E. Mean        59.035
Std Dev        236.139            Variance      55761.604
Kurtosis         -.299            S.E. Kurt         1.091
Skewness          .465            S.E. Skew          .564
Range          865.509            Minimum      -410.11462
Maximum      455.39409            Sum    -7.61614415980E-09

Valid observations -      16      Missing observations -       0
```

FIGURE 6.SPSS^X. NORMALITY OF RESIDUALS

SKEWNESS index of .465 indicates the residuals are positively skewed. The KURTOSIS value of -.299 indicates the residuals are platykurtic. Statistical tests of skewness and kurtosis are discussed in Tabachnick & Fidell (1989), pp. 72-73. For these data,

$$z_{SKEWNESS} = \frac{S - 0}{\sqrt{\frac{6}{N}}} = \frac{.464739}{\sqrt{\frac{6}{16}}} = .76$$

or, using reported SE_{SKEW},

$$z_{SKEWNESS} = \frac{.465}{.564} = .82;$$

$z_{SKEWNESS} < 1.96$ ∴ not skewed.

$$z_{KURTOSIS} = \frac{K - 0}{s_K} = \frac{K - 0}{\sqrt{\frac{24}{N}}} = \frac{-.298894}{\sqrt{\frac{24}{16}}} = -.24$$

or, using reported $SE_{KURTOSIS}$,

$$z_{KURTOSIS} = \frac{-.299}{1.091} = -.27$$

$z_{KURTOSIS} > -1.96$ ∴ no kurtosis

```
29  PEARSON CORR Y X1 X2 X3 X4 X5 X6 YHAT ERR/
30  OPTIONS 6
```

PEARSON CORR problem requires 1,872 bytes of workspace.

- - - - P E A R S O N   C O R R E L A T I O N   C O E F F I C I E N T S - - -

VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR	
Y	.9709	Y	.9836	Y	.5025	Y	.4573
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
X1	SIG .000	X2	SIG .000	X3	SIG .024	X4	SIG .037
Y	.9604	Y	.9713	Y	.9977	Y	.0672
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
X5	SIG .000	X6	SIG .000	YHAT	SIG .000	ERR	SIG .402
X1	.9916	X1	.6206	X1	.4647	X1	.9792
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
X2	SIG .000	X3	SIG .005	X4	SIG .035	X5	SIG .000
X1	.9911	X1	.9731	X1	.0000	X2	.6043
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
X6	SIG .000	YHAT	SIG .000	ERR	SIG .500	X3	SIG .007
X2	.4464	X2	.9911	X2	.9953	X2	.9858
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
X4	SIG .042	X5	SIG .000	X6	SIG .000	YHAT	SIG .000
X2	.0000	X3	-.1774	X3	.6866	X3	.6683
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
ERR	SIG .500	X4	SIG .255	X5	SIG .002	X6	SIG .002
X3	.5036	X3	.0000	X4	.3644	X4	.4172
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
YHAT	SIG .023	ERR	SIG .500	X5	SIG .083	X6	SIG .054
X4	.4583	X4	.0000	X5	.9940	X5	.9626
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
YHAT	SIG .037	ERR	SIG .500	X6	SIG .000	YHAT	SIG .000
X5	.0000	X6	.9735	X6	.0000	YHAT	.0000
WITH	N(16)	WITH	N(16)	WITH	N(16)	WITH	N(16)
ERR	SIG .500	YHAT	SIG .000	ERR	SIG .500	ERR	SIG .500

SIG IS 1-TAILED, "." IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 7.SPSS^X. CORRELATION OF ERRORS AND INDEPENDENT VARIABLES
Use the correlation matrix to determine association between each independent variable and the residual from the multiple regression equation.

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

Implementing Variable Selection Techniques in Regression

Jerome D. Thayer
Andrews University

Abstract

The most common methods of variable selection (forward, backward, all possible subsets) were considered. Criticisms and common misuses of stepwise methods were presented. Suggestions were made for each method concerning appropriate procedures to follow in running computer programs and the information that should be reported with the results. An example was presented which showed how proper selection should be done. When variables are selected for a regression model, the stepwise method can be helpful if the initial choice of variables is chosen as much as possible using theory, the defaults of the computer program used are not used automatically, more than one computer run is done using different variable selection methods, and the final model is chosen through an intelligent process, not automatically using the final model generated by the computer program. When the model is described, all subjective decisions made in the model selection process should be reported.

An earlier version of this paper was presented at the annual meeting of the American Educational Research Association in Boston, Massachusetts, April, 1990, as part of presidential address.

Multiple regression is one of the most popular statistical techniques used in behavioral science research. There are three ways in which it is typically used:

1) Testing a full model, interpreting the model and each of its components.

2) Adding components to a model and interpreting the value of the increment.

3) Using a stepwise method in which variables are added or deleted from a model in sequence to come up with a final "good" or "best" predictive model.

This paper deals with the third of these methods, the stepwise method.

Defining the term "stepwise"

In considering the stepwise method it is necessary to contrast the stepwise method used as a computer program with the stepwise method used as a methodological procedure and to note the different ways in which the stepwise method can be used.

Many computer programs are called "stepwise" programs because they can be used to build models using a stepwise method with default or user-specified alternatives controlling factors of the selection process including the criteria for entering and removing variables.

Stepwise computer programs can be used in four ways:

1) The program selects a model automatically using only the default values.

2) The program selects a model automatically using some or all user-specified values in place of default values.

3) The researcher uses the output of the program to help in selecting a model.

4) The program is used to make specified incremental tests by adding one or more variables to other variables.

Methods one and two are, almost without exception, the methods used in journal articles that claim to be using the "stepwise method". However, few of them specify what statistical criteria are used for adding and removing variables. In most cases the default values are probably used (method one). Critics of the stepwise method usually criticize the use of stepwise programs in either of the two automatic ways listed (methods one and two).

Since method three uses the professional judgment of the researcher in the selection of the final model, this procedure will be suggested as the appropriate use of the stepwise method in this paper. Method four uses the stepwise computer program, but it is not a use of the stepwise method so will not be considered here.

The stepwise method as a procedure can be used to describe at least four different variable selection strategies.

1) Forward method

   The selection begins with no variables in the model and variables are added one at a time if they meet the statistical criterion for entering variables.

2) Backward method

   The selection begins with all variables in the model and variables are removed one at a time if they meet the statistical criterion for removing variables.

3) Forward stepwise method

   This is a variation of the forward method in which at each step, before any variable is added, variables already in the model are considered for removal if they meet the statistical criterion for removing variables.

4) Backward stepwise method

   This is a variation of the backward method in which at each step, before any variable is removed, variables not in the model are considered for addition if they meet the statistical criterion for entering variables.

Usually in journal articles the method used is just called "stepwise" with no indication of which of the four methods or procedures is used. The method that is used in most cases is probably the forward stepwise method which is the default procedure for most stepwise computer programs.

Criticisms of the stepwise method

The stepwise method has been frequently criticized by methodologists (Davidson, 1988; Huberty, 1989; Thompson, 1989) and almost all authors of textbooks on multiple regression (i.e., Berenson et al., 1983; Chatterjee & Price, 1977; Cohen & Cohen, 1975; Draper & Smith, 1981; Freund & Minton, 1979; Gunst & Mason, 1980; Kleinbaum & Kupper, 1988; Morrison, 1983; Myers, 1986; Neter et al., 1983; Pedhazur, 1982; Wittink, 1988; Younger, 1979). The criticisms are both general and specific. Two 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)

Critics of the stepwise method suggest the following considerations for selecting a subset of predictors for a prediction model:

1) Selection of variables for a regression model should not be an automatic or mechanical process.
2) No one method will consistently select the "best" model.
3) There is no one "best" model according to any common criterion such as the maximum R^2.
4) The stepwise method should not be used to build models for explanatory research.
5) The stepwise method has limited usefulness when predictors are highly correlated, if a key set of variables work in combination, or when suppression exists.
6) The order in which variables enter the model should not be used as an indicator of the value of the variable as a predictor.

If a stepwise method is used to select a model in the automatic way that is most commonly found in the literature, it is quite likely that:

1) Other models with the same number of predictors may very well have a larger R^2.
2) Smaller models may very well predict an equivalent R^2.
3) Variables not included in the model may be just as good or better predictors than some of the variables in the model.
4) The variables will probably not enter the model in order of their importance in the final model.

Misuses of the stepwise method

In spite of these criticisms and suggestions, there are still many research studies reported in the recent literature in which these guidelines are violated. Most of these studies have the following characteristics:

1) Models selected by the computer were called the "best" or "optimum" model for maximizing the explained variance (R^2) with the minimum number of predictors (k).
2) No description was given of the process by which the model was selected other than the term "the stepwise method was used". In most cases an automatic forward stepwise process was probably used.

3) Explanatory interpretations were made by defining "good" predictors as those in the model and "poor" predictors as those not in the model.

4) The interpretation of the model included a ranking of the variables in the model in terms of importance based on order of entry.

5) No mention was made of the interrelationship of the variables in the description of the procedures used or in the interpretation of the final model selected.

Examples of misuses

Specific examples of these uses/misuses of stepwise regression found in the educational literature in 1988 and 1989 include:

1) We found the "most consistent variables that are most closely associated" with the criterion.

2) Variable A was picked as the "main predictor."

3) We wanted to find the "optimum equation."

4) The analysis yielded an "optimum predictor equation with as few predictors as possible."

5) "This allows the most consistent variables that are most closely associated with learning to be identified."

6) "The use of [variable A and variable B] as predictors revealed that [variable A] predicted [variable Y]. [Variable B] proved to lack significant predictive utility." In this article a table reported that the zero-order correlations between Y and variables A and B were .49 and .48. Variable B did not appear in the final model.

Should stepwise methods be used?

Although most, if not all statisticians would agree that stepwise methods should not be used when an explanatory model is desired, it is common to see research articles where explanatory interpretations are made to a model that is called a "prediction" model. Even if a predictive model is being selected, determining the value of each of the predictors in the prediction model requires more than what the stepwise method provides. Order of entry should not be used for this purpose. Stepwise methods should not be used to determine the number of variables in the final model. If multicollinearity exists in the data set, stepwise methods are especially suspect. In most cases, either the multicollinearity should be removed by removing variables, or other procedures should be used.

Since from the critics' point of view the stepwise methods are usually used in an inappropriate manner, the question then is whether the stepwise method should be a recommended technique for statistical analysis, and if so, how should it be used.

The objective of this paper is to consider the conditions under which variable selection procedures such as stepwise procedures can be used appropriately in educational research.

Value of using stepwise methods

The stepwise method is appropriate for situations in which a prediction model is desired, not an explanation model. In these situations, it is best used for exploratory analysis where little theory is available to guide in the selection of variables for the prediction model (Wittink, 1988).

Stepwise methods are very helpful if used properly when a subset of predictor variables is needed to be selected. A major advantage of stepwise methods is that by examining the output of each step of the model building process the researcher can see how each variable acts in different combinations which can be used to help the researcher to select the variables for the final model.

Observing the change in the partial correlations (and/or regression coefficients) as variables are added and deleted gives a feel for the variables that is difficult to get in any other way. If both forward and backward stepwise methods are used in conjunction with an all possible subsets program such as BMDP9R, a great variety of "good" models can be examined. An earlier study by Thayer (1986) showed that the backward stepwise and all possible subsets methods frequently gave different models than the forward stepwise method, in some cases with much higher R^2 values with the same or slightly more predictors.

The value of each potential predictor can be examined by comparing the zero-order correlations with the partial correlations at each step in the stepwise process. If the partial correlations remain high relative to the zero-order correlation, then the researcher can be confident of the stability of the variable in many prediction situations. If the partial correlations change markedly, then it will take some analysis to determine the dynamics involved, particularly noting which variables seem to be causing the changes.

How not to use stepwise methods

If stepwise methods are used the following procedures should be avoided:

1) Stepwise methods should not be used alone as the only procedure, especially if the researcher is looking for the "best" or "optimum" prediction model. An all possible subsets program such as BMDP9R should be used in conjunction with stepwise methods. It is also very desirable to use both the forward and

backward stepwise methods to examine alternative models. When one method is used the temptation is great to use the model that the computer selects as the final model. The final model should be selected as a result of many considerations, not only the statistical criterion used by the stepwise program.

2) Stepwise methods should not be used automatically using the default values. The default values of F(or p)-to-enter and F(or p)-to-remove are seldom appropriate for good model selection. Whether the default values are used or not, they should be specified in the reporting of the results.

3) The p values given for the increments at each step should not be taken at "face" value. Huberty (1989) suggests that "the tail probabilities . . . should not be taken too seriously. And one should certainly not refer such probabilities to conventional significance levels to determine the 'significance' of an entered or removed response variable."

<u>How to run stepwise programs</u>

If a stepwise program is used to provide data to the researcher for model selection, the following suggestions are offered:

1) Reduce the number of variables to work with to a size that will allow you to do an "all possible subsets" (BMDP9R) run. If computer memory permits, do a backward stepwise run to find the best 27 (or number that can be run by an all possible subsets program). If there are too many variables to do a backward run, then do a forward run with a very low F-to-enter, forcing in all theoretically important variables to find the "best" 27 (or so).

2) Allow theoretically important variables (variables that have been shown or are hypothesized to be "causal" variables) to be entered first by forcing them in the model or allowing them to be eligible for entrance if they satisfy the statistical criterion for entering variables.

3) Set low F-to-enter or high p-to-enter values, such as F = 0.00-2.00 or p = .10-1.00 (Myers, 1986; Wittink, 1988). This will allow the computer to enter more variables (if forward) or delete more variables (if backward) then desired for a final model, in order to consider more variables than you will use in the selected model. The major advantage of this is to allow more combinations of variables to be considered when the researcher selects the final model.

4) Run both forward and backward stepwise and all possible subsets procedures in order to consider alternative models and to examine the performance of the variables in different models.

The forward stepwise method frequently gives smaller models than the backward stepwise method and the researcher can observe changes occurring in the partial correlations (and/or regression coefficients) of variables which give a feel for the stability of the variable (Thayer, 1986).

The backward stepwise method has an advantage over the forward stepwise method because combinations of variables that work together but not singly are considered. The forward stepwise method will miss them (Thayer, 1986).

The all possible subsets method encourages examination of more than one model by providing statistics for many models of varying sizes, many of which are almost equivalent in statistical desirability.

5) Cross-validate alternative models suggested by the stepwise and all possible subsets runs. This can be done either by generating an equation from half of the data and cross-validating it on the other half, or by selecting another sample for the cross-validation.

6) Select the final model intelligently by using as many of the following criteria as possible:

Each variable in the model should contribute a meaningful amount to the total R^2 of the model (the incremental R squared of that variable in addition to the others in the model). With a large N the "best" model may be a smaller model than that suggested by considering only the p values of the variables.

The variables selected should as much as possible be theoretically meaningful variables.

The variables selected should as much as possible have partial correlations (and regression coefficients) which are relatively stable in the various steps or with different models. As variables are added or deleted in the stepwise process, if the sign of the partial correlation and regression coefficient for a predictor changes, that variable may not perform well in a cross-validation situation. If a partial correlation (and regression coefficient) becomes larger as the model increases in size, the variable should be studied closely to see whether there is some suppression or multicollinearity in the data that needs to be considered in the selection of the final model.

The variables selected should appear in many "good" models. Variables that only work in a few combinations would be unlikely to work well in a prediction model with new data.

The model should be one of the best models considered in terms of cross-validation.

How to report stepwise results

If stepwise procedures are used properly, many decisions must be made concerning how to run the stepwise program and how to select the variables for the final model. It is important that these decisions should be included in the final report.

The following procedures used should be reported:

1) F-to-enter/remove or p-to-enter/remove values used.

2) Stepping method used: forward, forward stepwise, backward, or backward stepwise.

3) Default or substitute values used.

4) Which alternative models were examined.

5) Results of stepwise methods compared to those of the all possible subsets method.

6) How subjective judgment (theory, etc.) was used in selecting variables for the model.

The following statistical results should be reported:

1) For each variable considered:

Zero-order correlations and partial correlations with the dependent variable at each step.

2) For each variable selected:

Why it was selected.

The stability of its regression coefficient, b or ß, (or its partial correlation) in different models.

3) For each variable not selected:

Why it was not selected.

Whether the variable was a good predictor in other combinations of variables tested or a good predictor alone.

Summary

When model selection is being done, the stepwise method can be helpful if the initial choice of variables is chosen as much as possible using theory, the defaults are not used automatically, more than one run is done using different variable selection methods, and the final model is chosen through an intelligent process, not automatically using the final model generated by the computer program.

Example

Appendices A-C report computer printouts and models of three variable selection computer runs: forward stepwise, backward stepwise and all possible subsets, using the BMDP2R and BMDP9R computer programs on data set A6 from Gunst and Mason (1980, pp 355, 363). These printouts, summary tables and comments inserted where appropriate, illustrate most of the points presented in this paper.

To make interpretation easier, the BMDP2R listing in Appendix A is a combined forward and backward run but the variables entered and removed are in the same order as they were with runs done using the forward stepwise and backward stepwise methods using F-to-enter/remove values of 2.00/1.99.

Table 1 reports a summary of the models selected by the forward and backward stepwise methods described in more detail in Appendix A. With the forward stepwise method, variable 2 was the first variable entered. If the default F-to-enter value of 4.00 had been used, variable 2 would not have entered and the 0-predictor model would have been selected.

Using F-to-enter/remove values of 2.00/1.99, the automatic forward stepwise method selected a 2-predictor model with an R^2 of .1495 which was the same as the best 2-predictor model found using the all possible subsets method. If the forward stepwise method would have been allowed to continue adding variables with F's below 2.00, the larger models selected became progressively worse compared to those identified as the "best" of the same model size by the all possible subsets method.

The model selected by the backward stepwise method was a 7-predictor model which was the same model chosen by the all possible subsets as the best model of any size. If smaller models had been chosen with the backward stepwise method, they would have become progressively poorer than the "best" model of the same size selected by the all possible subsets method. Forward stepwise gave better small models while backward stepwise gave better large models.

Table 1

Summary of Models Selected By Forward and Backward Stepwise Methods

Predictors	Step	Method	Variables In the Model	R^2
2	2	Forward	2 4	.1495
2	26	Backward	2 13	.0903
4	4	Forward	2 4 8 10	.2108
4	24	Backward	2 6 13 15	.1783
5	5	Forward	2 4 8 10 12	.2428
5	23	Backward	2 6 12 13 15	.2289
6	6	Forward	2 4 8 9 10 12	.2649
6	22	Backward	2 5 6 12 13 15	.2966
7	7	Forward	2 4 5 8 9 10 12	.2963
7	21	Backward	2 5 6 9 12 13 15	.3472

Appendix B reports the printout of the BMDP9R all possible subsets run on the A6 data. One 4-predictor model, four 5-predictor models, at least ten 6-predictor models, and at least ten 7-predictor models had C_p values lower than the recommended minimum value (k+1 where k is the number of predictors in the model). The model with the lowest C_p value was a 7-predictor model. Although this model was identified as the "best" model by BMDP9R, all of the models with lower than minimum C_p values could be considered to be "good enough" models. The best 2, 4, 5, 6, and 7 predictor models along with all other models with acceptable C_p values with these model sizes are reported in Appendices A-C.

Table 2 compares the models selected by the all possible subsets method with those of the forward stepwise and backward stepwise methods. The three methods never gave the same models with 2, 4, 5, 6, or 7 predictors.

The models selected by the forward stepwise method were identified by the all possible subsets methods as the best 2-predictor model, the 2nd best 4-predictor model, the 5th best 5-predictor model and not in the top ten 6 or 7-predictor models. The models selected by the backward stepwise method were identified by the all possible subsets methods as the 4th best 2-predictor model, not in the top ten 4 or 5-predictor model, the 2nd best 6-predictor model, and the best 7-predictor model.

Variables 4, 8 and 10 were three of the first four variables entered in the forward stepwise method but they were also three of the first five removed in the backward stepwise method. Using the order of entry criterion for importance would indicate that 4, 8 and 10 were some of the best variables if you used the forward stepwise method or some of the worst variables if you used the backward stepwise method.

If the forward stepwise method would have been used to select the "best" model and order of entry was used to indicate importance (which should not be done), variable 2 would be called the "most important" variable and variable 4 the "next most important". If better models had been used, such as those shown in Appendix C, and contribution to R^2 was used as the criterion for importance, variable 2 would have been the "most important" in every model, but variable 4 did not appear in any of the models.

Variable 5 was the second best variable in two of the four "best" models (with 5, 6, and 7 predictors) in Appendix C and Table 3 but the least important variable in the other two models. Since these models are good competitors for the "best" model as explained later, it can been seen that even using contribution to R^2 is likely to mislead in indicating the importance of variables, since it can be so heavily dependent on what other variables are in the model.

The listing of the partial correlations for each variable for each step gives an indication of the stability of the variables. Variable 2, the "most important" variable, is very stable, while other variables are shown to vary somewhat. None of the variables changes signs while in the model (indicated by as asterisk in the printout).

Table 2

Models Selected By All Possible Subsets, Forward Stepwise and Backward Stepwise Methods

Predictors	Method	Variables in the Model	R^2	All Possible Ranking
2	All Possible Subsets	2 4	.1495	
2	Forward	2 4	.1495	1st
2	Backward	2 13	.0903	4th
4	All Possible Subsets	2 4 5 9	.2278	
4	Forward	2 4 8 10	.2108	2nd
4	Backward	2 6 13 15	.1783	Not in top 10
5	All Possible Subsets	2 5 9 12 13	.2699	
5	Forward	2 4 8 10 12	.2428	5th
5	Backward	2 6 12 13 15	.2289	Not in top 10
6	All Possible Subsets	2 3 5 12 13 15	.3000	
6	Forward	2 4 8 9 10 12	.2649	Not in top 10
6	Backward	2 5 6 12 13 15	.2966	2nd
7	All Possible Subsets	2 5 6 9 12 13 15	.3472	
7	Forward	2 4 5 8 9 10 12	.2963	Not in top 10
7	Backward	2 5 6 9 12 13 15	.3472	1st

Disregarding theory in the selection of models, there were four models with 5, 6, and 7-predictors that appear to be worthy of selection as a "best" model are listed in Table 3. More complete information on the models is provided in Appendix C.

Variables 2, 5, 12, and 13 appear in all of these models, variable 15 in four of the models, variable 9 in three of the models, and variables 3 and 6 in only two of the models. In the list of partial correlations at each step in Appendix A it is clear that variable 6 is a better predictor in most situations and therefore would be expected to do better in cross validation. The best 5, 6, and 7 predictor models are then those with asterisks by the R^2 values in Table 3. The choice between these models could be done after cross-validation and consideration of other criteria not discussed in this paper.

Table 3

Candidates for "Best" Model

Number of Predictors	Variables in the Model	R^2
5	2 5 9 12 13	.269878*
6	2 3 5 12 13 15	.299565
6	2 5 6 12 13 15	.296645*
7	2 5 6 9 12 13 15	.347206*
7	2 3 5 9 12 13 15	.346599

* "best" models

References

Berenson, M. L., et al. (1983). Intermediate statistical methods and applications. Englewood Cliffs, NJ: Prentice-Hall

Chatterjee, S. & Price, B. (1977). Regression Analysis by example. New York: John Wiley & Sons.

Cohen, J. & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Davidson, B. M. (1988). The case against using stepwise research methods. Paper presented at the Annual Meeting of the Mid-South Educational Research Association.

Draper, N. R. & Smith, H. (1981). Applied regression analysis, 2nd edition. New York: John Wiley & Sons.

Freund, R. J. & Minton, P. D. (1979). Regression methods. New York: Marcel Dekker.

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

Huberty, C. J. (1989). Problems with stepwise methods--better alternatives. In Thompson, B. (Ed.), Advances in Social Science Methodology, vol 1, (pp. 43-70), Greenwich, CT: JAI Press.

Kleinbaum, D. G. & Kupper, L. L. (1988). Applied regression analysis and other multivariable methods (2nd Edition). Boston, MA: PWS-KENT.

Morrison, D. F. (1983). Applied linear statistical methods. Englewood Cliffs, NJ: Prentice-Hall.

Myers, R. H. (1986). Classical and modern regression with applications. Boston, MA: Duxbury.

Neter, J. et al. (1983). Applied linear regression models. Homewood, IL: Richard D. Irwin.

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

Thayer, J. D. (1986). Testing different model building procedures using multiple regression. Multiple Linear Regression Viewpoints, 16, 37-52.

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

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

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

Younger, M. S. (1979). A handbook for linear regression. North Scituate, MA: Duxbury Press.

Appendix A

Forward/Backward Stepwise Results

?2R - STEPWISE REGRESSION

:RAM INSTRUCTIONS

```
xut    file='a6'.
       format=free.
       variables=15.
ireaa  dependent=1.
       independent=2 to 15.
       enter =0.01,500.
       remove=0.00,499.
nt     part.
I
```

:ER OF CASES READ.   50

:RIABLE NAME	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SKEWNESS	KURTOSIS	SMALLEST VALUE	LARGEST VALUE	SMALLEST STD SCORE	LARGEST STD SCORE
X(1)	0.3162	0.0482	0.152482	0.5209	-0.3210	0.2210	0.4270	-1.9745	2.2981
X(2)	177.9060	6.7031	0.037678	0.0684	0.0559	163.1000	193.5000	-2.2088	2.3264
X(3)	78.3524	11.4684	0.146369	0.3066	-0.7693	54.2800	102.2600	-2.0990	2.0847
X(4)	40.9520	1.5861	0.038730	-0.2923	-0.3913	37.2000	43.8000	-2.3656	1.7956
X(5)	28.1060	1.4372	0.051134	0.0975	0.1848	24.2000	31.6000	-2.7178	2.4312
X(6)	90.6200	5.9709	0.065890	0.2624	-1.0669	80.5000	101.2000	-1.6949	1.7719
X(7)	16.1300	6.1019	0.378293	0.0703	-0.1614	3.5000	32.0000	-2.0699	2.6008
X(8)	73.0800	12.9107	0.176666	0.4533	-0.5277	50.0000	104.0000	-1.7877	2.3949
X(9)	74.6000	8.1140	0.108767	-0.2439	0.9651	48.0000	90.0000	-3.2783	1.8980
X(10)	6.2800	4.3801	0.697469	0.8519	0.4282	0.0000	20.0000	-1.4338	3.1323
X(11)	193.3400	25.4547	0.131658	0.7104	0.7753	146.0000	272.0000	-1.8598	3.0902
X(12)	114.9400	10.3990	0.090473	0.6146	1.4698	91.0000	147.0000	-2.3021	3.0830
X(13)	5.4800	0.3637	0.066360	-1.1897	3.7250	4.0000	6.0000	-4.0698	1.4299
X(14)	4.1740	0.5642	0.135163	0.1375	-0.2559	3.0000	5.5000	-2.0809	2.3503
X(15)	13.7680	7.3521	0.533998	0.6392	-0.2236	2.0100	32.6300	-1.5993	2.5655

:ELATIONS

	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	X(9)	X(10)	X(11)	X(12)	X(13)	X(14)
	2	3	4	5	6	7	8	9	10	11	12	13	14
2	1.000												
3	0.635	1.000											
4	0.654	0.666	1.000										
5	0.586	0.647	0.582	1.000									
6	0.426	0.889	0.554	0.522	1.000								
7	0.223	0.554	0.205	0.207	0.398	1.000							
8	-0.182	-0.264	-0.168	-0.323	-0.246	-0.006	1.000						
9	-0.187	-0.052	-0.145	0.148	-0.008	0.049	0.234	1.000					
10	-0.276	-0.576	-0.274	-0.158	-0.454	-0.670	0.155	0.054	1.000				
11	0.588	0.450	0.368	0.354	0.347	0.207	-0.064	-0.165	-0.358	1.000			
12	-0.151	-0.118	-0.042	-0.247	-0.065	0.244	0.528	0.155	-0.101	0.101	1.000		
13	0.216	-0.053	0.202	0.041	-0.162	-0.208	-0.300	-0.321	0.324	-0.031	-0.473	1.000	
14	-0.189	-0.368	-0.245	-0.229	-0.328	-0.336	0.007	0.134	0.213	-0.315	-0.018	-0.022	1.000
15	0.364	0.810	0.329	0.414	0.727	0.843	-0.093	0.047	-0.689	0.301	0.098	-0.205	-0.405
1	0.222	0.056	-0.094	-0.056	-0.032	0.132	0.163	0.147	-0.158	0.160	-0.076	-0.149	-0.053

	X(15)	X(1)
	15	1
15	1.000	
1	0.165	1.000

:ble 2 had the highest zero-order correlations with x(1), the dependent variable. Variables 7-11, 13 and 15 all had similar values.

```
?ING ALGORITHM. . . . . . . . . . . . . . . . .F
IDENT VARIABLE. . . . . . . . . . . . . . . .        1 X(1)
‹UM ACCEPTABLE F TO ENTER . . . . . . . . .    0.010, 500.000
‹UM ACCEPTABLE F TO REMOVE. . . . . . . . .    0.000, 499.000
‹UM ACCEPTABLE TOLERANCE. . . . . . . . . . 0.01000
```

STEP NO. 2
\-\-\-\-\-\-\-\-\-\-\-\-\-\-\-

VARIABLE ENTERED 4 X(4)

MULTIPLE R 0.3867
MULTIPLE R-SQUARE 0.1495
ADJUSTED R-SQUARE 0.1133

STD. ERROR OF EST. 0.0454

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.17033600E-01	2	0.8516802E-02	4.13
RESIDUAL	0.96874410E-01	47	0.2061158E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.20221)					
X(2)	2	0.00357	0.0013	0.496	0.57190	7.78	1
X(4)	4	-0.01272	0.0054	-0.418	0.57190	5.53	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	0.03034	0.48713	0.04	1
X(5)	5	-0.14554	0.58742	1.00	1
X(6)	6	-0.01457	0.68557	0.01	1
X(7)	7	0.11971	0.94419	0.67	1
X(8)	8	0.20199	0.96277	1.96	1
X(9)	9	0.19794	0.96429	1.88	1
X(10)	10	-0.15485	0.90860	1.13	1
X(11)	11	0.02914	0.65395	0.04	1
X(12)	12	-0.02049	0.97184	0.02	1
X(13)	13	-0.19154	0.94707	1.75	1
X(14)	14	-0.06840	0.93861	0.22	1
X(15)	15	0.14331	0.85272	0.96	1

This model was the model selected by forward stepwise using an F-to-enter of 2.00.

STEP NO. 21
\-\-\-\-\-\-\-\-\-\-\-\-\-\-\-

VARIABLE REMOVED 7 X(7)

MULTIPLE R 0.5892
MULTIPLE R-SQUARE 0.3472
ADJUSTED R-SQUARE 0.2384

STD. ERROR OF EST. 0.0421

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.39549630E-01	7	0.5649948E-02	3.19
RESIDUAL	0.74358380E-01	42	0.1770438E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.57824)					
X(2)	2	0.00423	0.0012	0.588	0.53012	11.78	1
X(5)	5	-0.01570	0.0061	-0.468	0.47030	6.62	1
X(6)	6	-0.00318	0.0016	-0.393	0.38895	3.87	1
X(9)	9	0.15129E-02	0.8388E-03	0.255	0.77996	3.25	1
X(12)	12	-0.16762E-02	0.6988E-03	-0.362	0.68420	5.75	1
X(13)	13	-0.04381	0.0208	-0.330	0.63093	4.43	1
X(15)	15	0.00253	0.0012	0.386	0.43940	4.21	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	-0.11216	0.07232	0.52	1
X(4)	4	-0.05438	0.36713	0.12	1
X(7)	7	-0.18517	0.18059	1.46	1
X(8)	8	0.14242	0.61554	0.85	1
X(10)	10	0.14081	0.42251	0.83	1
X(11)	11	0.10966	0.58982	0.50	1
X(14)	14	-0.09643	0.79427	0.38	1

This is the model that was selected by backward stepwise using an F-to-enter of 2.00.

› NO. 22
············
ABLE REMOVED 9 X(9)

IPLE R	0.5447
IPLE R-SQUARE	0.2966
STED R-SQUARE	0.1985
ERROR OF EST.	0.0432

YSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.33790250E-01	6	0.5631709E-02	3.02
RESIDUAL	0.80117760E-01	43	0.1863204E-02	

VARIABLES IN EQUATION FOR X(1) . VARIABLES NOT IN EQUATION

VARIABLE	COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL.	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
NTERCEPT	0.75673)											
2	0.00364	0.0012	0.506	0.56993	8.92	1 .	X(3)	3	-0.13162	0.07290	0.74	1
5	-0.01199	0.0059	-0.357	0.53053	4.14	1 .	X(4)	4	-0.08029	0.37123	0.27	1
6	-0.00352	0.0016	-0.436	0.39451	4.59	1 .	X(7)	7	-0.18857	0.18085	1.55	1
) 12	-0.15808E-02	0.7148E-03	-0.341	0.68815	4.89	1 .	X(8)	8	0.19456	0.64908	1.65	1
) 13	-0.05222	0.0208	-0.394	0.66426	6.30	1 .	X(9)	9	0.26812	0.77996	3.25	1
) 15	0.00261	0.0013	0.398	0.43996	4.26	1 .	X(10)	10	0.17102	0.43055	1.27	1
						.	X(11)	11	0.06474	0.60307	0.18	1
						.	X(14)	14	-0.04708	0.81680	0.09	1

This is the 6 predictor model that would have been selected by backward stepwise

NO. 23
··········
ABLE REMOVED 5 X(5)

PLE R	0.4784
PLE R-SQUARE	0.2289
TED R-SQUARE	0.1412
ERROR OF EST.	0.0447

SIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.26069730E-01	5	0.5213945E-02	2.61
RESIDUAL	0.87838280E-01	44	0.1996325E-02	

VARIABLES IN EQUATION FOR X(1) . VARIABLES NOT IN EQUATION

ARIABLE	COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL.	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
TERCEPT	0.62171)											
2	0.00250	0.0011	0.348	0.72136	4.99	1 .	X(3)	3	-0.19981	0.07841	1.79	1
6	-0.00425	0.0017	-0.526	0.41403	6.54	1 .	X(4)	4	-0.16324	0.40909	1.18	1
12	-0.12176E-02	0.7165E-03	-0.263	0.73387	2.89	1 .	X(5)	5	-0.29647	0.53053	4.14	1
13	-0.04737	0.0214	-0.357	0.67307	4.90	1 .	X(7)	7	-0.16202	0.18150	1.16	1
15	0.00245	0.0013	0.373	0.44178	3.51	1 .	X(8)	8	0.22815	0.66399	2.36	1
						.	X(9)	9	0.14120	0.87984	0.87	1
						.	X(10)	10	0.07498	0.46620	0.24	1
						.	X(11)	11	0.05971	0.60311	0.15	1
						.	X(14)	14	-0.02641	0.81998	0.03	1

This is the 5 predictor model that would have been selected by backward stepwise.

STEP NO. 24

VARIABLE REMOVED 12 X(12)

MULTIPLE R 0.4222
MULTIPLE R-SQUARE 0.1783
ADJUSTED R-SQUARE 0.1052

STD. ERROR OF EST. 0.0456

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.20304410E-01	4	0.5076102E-02	2.44
RESIDUAL	0.93603600E-01	45	0.2080080E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.33832)					
X(2)	2	0.00250	0.0011	0.348	0.72136	4.78	1
X(6)	6	-0.00361	0.0017	-0.447	0.43646	4.77	1
X(13)	13	-0.03077	0.0194	-0.232	0.85029	2.51	1
X(15)	15	0.00207	0.0013	0.315	0.45494	2.48	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	-0.13532	0.08226	0.82	1
X(4)	4	-0.21748	0.43907	2.18	1
X(5)	5	-0.21616	0.56578	2.16	1
X(7)	7	-0.20628	0.19016	1.96	1
X(8)	8	0.09303	0.81411	0.38	1
X(9)	9	0.14000	0.87999	0.88	1
X(10)	10	0.04373	0.47239	0.08	1
X(11)	11	0.01118	0.62425	0.01	1
X(12)	12	-0.24818	0.73387	2.89	1
X(14)	14	-0.01307	0.82206	0.01	1

This is the 4 predictor model that would have been selected by backward stepwise.

PARTIAL CORRELATIONS

IABLES 2 X(2)	3 X(3)	4 X(4)	5 X(5)	6 X(6)	7 X(7)	8 X(8)	9 X(9)	10 X(10)	11 X(11)
0.2223	0.0562	-0.0938	-0.0559	-0.0318	0.1324	0.1631	0.1473	-0.1585	0.1595
0.2223*	-0.1129	-0.3245	-0.2356	-0.1434	0.0871	0.2122	0.1971	-0.1036	0.0365
0.3768*	0.0303	-0.3245*	-0.1455	-0.0146	0.1197	0.2020	0.1979	-0.1548	0.0291
0.3975*	0.0704	-0.3183*	-0.0974	0.0223	0.1138	0.2020*	0.1632	-0.1804	0.0192
0.3828*	-0.0310	-0.3397*	-0.0846	-0.0460	-0.0057	0.2218*	0.1616	-0.1804*	-0.0300
0.3583*	-0.0544	-0.3269*	-0.1081	-0.0496	0.0388	0.2934*	0.1705	-0.2240*	0.0111
0.3754*	-0.0866	-0.3278*	-0.2068	-0.0762	0.0228	0.2693*	0.1705*	-0.2244*	0.0286
0.4204*	-0.0064	-0.2414*	-0.2068*	-0.0243	0.0658	0.2137*	0.2442*	-0.2151*	0.0606
0.4444*	-0.0245	-0.1903*	-0.2474*	-0.0717	0.1104	0.1864*	0.2220*	-0.1155*	0.0611
0.4462*	-0.2154	-0.1743*	-0.2824*	-0.2324	0.0115	0.1711*	0.2241*	0.0083*	0.0822
0.4536*	-0.0807	-0.0300*	-0.3116*	-0.2324*	-0.1459	0.1109*	0.2243*	0.0866*	0.1410
0.4489*	-0.0721	0.0063*	-0.3287*	-0.2721*	-0.1459*	0.0859*	0.2320*	0.0755*	0.1419
0.3634*	-0.0794	0.0392*	-0.3500*	-0.2936*	-0.1468*	0.0745*	0.2516*	0.1154*	0.1419*
0.3671*	-0.0794*	0.0607*	-0.3304*	-0.1897*	-0.1419*	0.0779*	0.2497*	0.1013*	0.1457*
0.3688*	-0.0785*	0.0549*	-0.3321*	-0.1896*	-0.1440*	0.0733*	0.2537*	0.0942*	0.1279*
0.3671*	-0.0794*	0.0607*	-0.3304*	-0.1897*	-0.1419*	0.0779*	0.2497*	0.1013*	0.1457*
0.3842*	-0.0645*	0.0607	-0.3326*	-0.1805*	-0.1323*	0.0930*	0.2442*	0.0867*	0.1354*
0.3951*	-0.0645	0.0392	-0.3731*	-0.3283*	-0.1417*	0.0862*	0.2487*	0.1088*	0.1367*
0.4118*	-0.0542	0.0586	-0.4059*	-0.3506*	-0.1521*	0.0862	0.2683*	0.1351*	0.1388*
0.3992*	-0.0872	0.0097	-0.3872*	-0.3470*	-0.1850*	0.1179	0.2791*	0.1351	0.1093*
0.4770*	-0.0818	-0.0050	-0.3809*	-0.3403*	-0.1852*	0.1135	0.2658*	0.1045	0.1093
0.4680*	-0.1122	-0.0544	-0.3691*	-0.2906*	-0.1852	0.1424	0.2681*	0.1408	0.1097
0.4146*	-0.1316	-0.0803	-0.2965*	-0.3105*	-0.1886	0.1946	0.2681	0.1710	0.0647
0.3191*	-0.1998	-0.1632	-0.2965	-0.3598*	-0.1620	0.2281	0.1412	0.0750	0.0597
0.3099*	-0.1353	-0.2175	-0.2162	-0.3097*	-0.2063	0.0930	0.1400	0.0437	0.0112
0.3321*	0.0332	-0.2502	-0.2058	-0.2167*	0.1024	0.1099	0.1484	-0.0946	0.0125
0.2636*	-0.1743	-0.3154	-0.2652	-0.2167	0.0333	0.1657	0.1459	-0.0212	-0.0051
0.2223*	-0.1129	-0.3245	-0.2356	-0.1434	0.0871	0.2122	0.1971	-0.1036	0.0365
Good Always		Good Small Not Alone	Low Alone	Good Only With X(15)					

PARTIAL CORRELATIONS

IABLES 12 X(12)	13 X(13)	14 X(14)	15 X(15)
-0.0756	-0.1493	-0.0525	0.1650
-0.0437	-0.2072	-0.0110	0.0925
-0.0205	-0.1915	-0.0684	0.1433
-0.1511	-0.1458	-0.0618	0.1508
-0.2016	-0.0700	-0.0343	0.0443
-0.2016*	-0.1488	-0.0312	0.0587
-0.2087*	-0.1071	-0.0517	0.0320
-0.2363*	-0.1746	-0.0978	0.1194
-0.2836*	-0.1746*	-0.1150	0.1444
-0.3028*	-0.1922*	-0.0745	0.1444*
-0.3513*	-0.2726*	-0.0776	0.2626*
-0.3280*	-0.2823*	-0.0848	0.2718*
-0.3523*	-0.2970*	-0.0481	0.2904*
-0.3599*	-0.2973*	-0.0467	0.3000*
-0.3554*	-0.2983*	-0.0467*	0.2910*
-0.3599*	-0.2973*	-0.0467	0.3000*
-0.3631*	-0.3008*	-0.0534	0.3119*
-0.3581*	-0.3122*	-0.0524	0.3059*
-0.3520*	-0.3308*	-0.0625	0.3274*
-0.3298*	-0.3048*	-0.0803	0.3080*
-0.3156*	-0.3138*	-0.1028	0.3064*
-0.3471*	-0.3090*	-0.0964	0.3020*
-0.3196*	-0.3573*	-0.0471	0.3004*
-0.2482*	-0.3166*	-0.0264	0.2717*
-0.2482	-0.2298*	-0.0131	0.2285*
-0.1992	-0.2622*	-0.0736	0.2285
-0.1587	-0.2072*	-0.0072	0.0298
-0.0437	-0.2072	-0.0110	0.0925
Good Large	Good Large		Good Large

SUMMARY TABLE

STEP NO.	VARIABLE ENTERED	VARIABLE REMOVED	MULTIPLE R	RSQ	CHANGE IN RSQ	F TO ENTER	F TO REMOVE	NO.OF VAR. INCLUDED
1	2 X(2)		0.2223	0.0494	0.0494	2.50		1
2	4 X(4)		0.3867	0.1495	0.1001	5.53		2
3	8 X(8)		0.4292	0.1842	0.0347	1.96		3
4	10 X(10)		0.4591	0.2108	0.0265	1.51		4
5	12 X(12)		0.4928	0.2428	0.0321	1.86		5
6	9 X(9)		0.5146	0.2649	0.0220	1.29		6
7	5 X(5)		0.5443	0.2963	0.0314	1.88		7
8	13 X(13)		0.5637	0.3178	0.0215	1.29		8
9	15 X(15)		0.5762	0.3320	0.0142	0.85		9
10	6 X(6)		0.6067	0.3681	0.0361	2.23		10
11	7 X(7)		0.6177	0.3815	0.0135	0.83		11
12	11 X(11)		0.6277	0.3940	0.0125	0.76		12
13	3 X(3)		0.6307	0.3978	0.0038	0.23		13
14	14 X(14)		0.6318	0.3991	0.0013	0.08		14
15		14 X(14)	0.6307	0.3978	-0.0013		0.08	13
16		4 X(4)	0.6289	0.3956	-0.0022		0.13	12
17		3 X(3)	0.6269	0.3930	-0.0025		0.15	11
18		8 X(8)	0.6233	0.3885	-0.0045		0.28	10
19		10 X(10)	0.6141	0.3771	-0.0114		0.73	9
20		11 X(11)	0.6079	0.3696	-0.0075		0.48	8
21		7 X(7)	0.5892	0.3472	-0.0224		1.46	7
22		9 X(9)	0.5447	0.2966	-0.0506		3.25	6
23		5 X(5)	0.4784	0.2289	-0.0678		4.14	5
24		12 X(12)	0.4222	0.1783	-0.0506		2.89	4
25		15 X(15)	0.3647	0.1330	-0.0453		2.48	3
26		6 X(6)	0.3004	0.0903	-0.0427		2.27	2
27		13 X(13)	0.2223	0.0494	-0.0408		2.11	1

Appendix B

All Possible Subsets Output

R - ALL POSSIBLE SUBSETS REGRESSION

AM INSTRUCTIONS

```
t     file='a6'.
      format=free.
      variables=15.
ess   dependent=1.
      independent=2 to 15.
```

DENT VARIABLE. . . . . 1 X(1)
R OF 'BEST' REGRESSIONS REPORTED 5
TION CRITERION . . . . CP

R OF CASES READ. 50

ACH SUBSET SELECTED BY YOUR CRITERION, THE R-SQUARED, ADJUSTED R-SQUARED, MALLOWS' CP, AND THE VARIABLE NAMES ARE ED. THE REGRESSION COEFFICIENTS AND T-STATISTICS ARE PRINTED TO THE RIGHT OF THE VARIABLE NAMES. MANY OTHER SUBSETS MAY ALSO PORTED THAT ARE NOT ACCOMPANIED BY REGRESSION COEFFICIENTS AND T-STATISTICS. SOME OF THESE SUBSETS MAY BE QUITE GOOD ALTHOUGH ARE NOT NECESSARILY BETTER THAN ANY SUBSET THAT HAS NOT BEEN PRINTED.

SUBSETS WITH 2 VARIABLES

UARED	ADJUSTED R-SQUARED	CP			
149539	0.113349	5.54	X(2)	X(4)	Forward Stepwise 2 Predictor (BEST) Model
102208	0.064004	8.29	X(2)	X(5)	
92249	0.053621	8.87	X(2)	X(8)	
90255	0.051543	8.99	X(2)	X(13)	Backward Stepwise 2 Predictor Model

SUBSETS WITH 4 VARIABLES

UARED	ADJUSTED R-SQUARED	CP					
227830	0.159193	4.98	X(2)	X(4)	X(5)	X(9)	
210777	0.140623	5.97	X(2)	X(4)	X(8)	X(10)	Forward Stepwise 4 Predictor Model

Backward Stepwise 4 Predictor Model
Not Listed In Best Ten 4 Predictor Models

SUBSETS WITH 5 VARIABLES

UARED	ADJUSTED R-SQUARED	CP						
269878	0.186910	4.53	VARIABLE	COEFFICIENT	T-STATISTIC			
			2 X(2)	0.00425862	3.43			
			5 X(5)	-0.0169603	-2.88			
			9 X(9)	0.00166493	1.94			
			12 X(12)	-0.00139223	-1.96			
			13 X(13)	-0.0408527	-1.98			
			INTERCEPT	0.294943				
252855	0.167953	5.52	X(2)	X(3)	X(12)	X(13)	X(15)	
247058	0.161496	5.86	X(2)	X(4)	X(5)	X(9)	X(13)	
244840	0.159026	5.99	X(2)	X(5)	X(8)	X(12)	X(13)	
242845	0.156804	6.10	X(2)	X(4)	X(8)	X(10)	X(12)	Forward Stepwise 4 Predictor Model

SUBSETS WITH 6 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP							
0.299565	0.201830	4.80	VARIABLE	COEFFICIENT	T-STATISTIC				Backward Stepwise 5 Predictor Model Not Listed In Best Ten 5 Predictor Models
			2 X(2)	0.00469572	3.44				
			3 X(3)	-0.00285065	-2.19				
			5 X(5)	-0.0103034	-1.69				
			12 X(12)	-0.00168170	-2.33				
			13 X(13)	-0.0488981	-2.38				
			15 X(15)	0.00369285	2.30				
			INTERCEPT	0.404160					
0.296645	0.198502	4.97	X(2)	X(5)	X(6)	X(12)	X(13)	X(15)	Backward Stepwise 6 Predictor Model
0.291666	0.192828	5.26	X(2)	X(3)	X(8)	X(12)	X(13)	X(15)	
0.290071	0.191011	5.35	X(2)	X(4)	X(5)	X(9)	X(12)	X(13)	
0.288310	0.189004	5.45	X(2)	X(5)	X(8)	X(9)	X(12)	X(13)	
0.287008	0.187520	5.53	X(2)	X(5)	X(9)	X(12)	X(13)	X(15)	
0.285008	0.185242	5.65	X(2)	X(3)	X(5)	X(9)	X(12)	X(15)	
0.282375	0.182242	5.80	X(2)	X(5)	X(9)	X(12)	X(13)	X(14)	
0.282028	0.181846	5.82	X(2)	X(5)	X(7)	X(9)	X(12)	X(13)	
0.281706	0.181479	5.84	X(2)	X(5)	X(6)	X(9)	X(12)	X(13)	

Forward Stepwise 6 Predictor Model
Not Listed In Best Ten 6 Predictor Models

SUBSETS WITH 7 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP				
0.347206	0.238407	4.02	VARIABLE	COEFFICIENT	T-STATISTIC	Backward Stepwise 7 Predictor (BEST) Model
			2 X(2)	0.00422728	3.43	
			5 X(5)	-0.0156974	-2.57	
			6 X(6)	-0.00317676	-1.97	
			9 X(9)	0.00151294	1.80	
			12 X(12)	-0.00167624	-2.40	
			13 X(13)	-0.0438120	-2.11	
			15 X(15)	0.00253197	2.05	
			INTERCEPT	0.578242		
0.346599	0.237699	4.06	VARIABLE	COEFFICIENT	T-STATISTIC	
			2 X(2)	0.00513442	3.78	
			3 X(3)	-0.00251951	-1.96	
			5 X(5)	-0.0141462	-2.23	
			9 X(9)	0.00146501	1.74	
			12 X(12)	-0.00173720	-2.49	
			13 X(13)	-0.0409863	-1.99	
			15 X(15)	0.00345443	2.19	
			INTERCEPT	0.267485		

Forward Stepwise 7 Predictor Model
Not Listed In Best Ten 7 Predictor Models

Model 1 -- All Possible Subsets Best 7 Predictor Model & Backward Stepwise 7 Predictor Model

ɪTICS FOR 'BEST' SUBSET

ʌS' CP	4.02
ED MULTIPLE CORRELATION	0.34721
PLE CORRELATION	0.58924
ŧED SQUARED MULT. CORR.	0.23841
JAL MEAN SQUARE	0.001770
ARD ERROR OF EST.	0.042077
ISTIC	3.19
ATOR DEGREES OF FREEDOM	7
INATOR DEGREES OF FREEDOM	42
FICANCE (TAIL PROB.)	0.0084

ŧHAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES
LECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

IABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ	
ŧERCEPT	0.578242	0.275920	11.993	2.10	0.042			
(2)	0.00422728	0.00123163	0.588	3.43	0.001	0.530117	0.18310	
(5)	-0.0156974	0.00609680	-0.468	-2.57	0.014	0.470301	0.10297	2nd Highest Contribution to R-SQ
(6)	-0.00317676	0.00161419	-0.393	-1.97	0.056	0.388951	0.06020	
(9)	0.00151294	0.000838828	0.255	1.80	0.078	0.779956	0.05056	
(12)	-0.00167624	0.000698809	-0.362	-2.40	0.021	0.684202	0.08943	
(13)	-0.0438120	0.0208095	-0.330	-2.11	0.041	0.630935	0.06890	
(15)	0.00253197	0.00123339	0.386	2.05	0.046	0.439403	0.06550	

ONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE
ED FROM THE REGRESSION EQUATION.

Model 2 -- All Possible Subsets Best 6 Predictor Model

STICS FOR 'BEST' SUBSET

ED MULTIPLE CORRELATION	0.29956
PLE CORRELATION	0.54733
TED SQUARED MULT. CORR.	0.20183
UAL MEAN SQUARE	0.001855
ARD ERROR OF EST.	0.043075
TISTIC	3.07
ATOR DEGREES OF FREEDOM	6
INATOR DEGREES OF FREEDOM	43
IFICANCE (TAIL PROB.)	0.0138

IABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ	
NTERCEPT	0.404160	0.248288	8.383	1.63	0.111			
X(2)	0.00469572	0.00136419	0.653	3.44	0.001	0.452848	0.19300	
X(3)	-0.00285065	0.00130277	-0.678	-2.19	0.034	0.169636	0.07799	
X(5)	-0.0103034	0.00608455	-0.307	-1.69	0.098	0.495200	0.04671	Lowest Contribution to R-SQ
X(12)	-0.00168170	0.000721680	-0.363	-2.33	0.025	0.672333	0.08845	
X(13)	-0.0488981	0.0205573	-0.369	-2.38	0.022	0.677556	0.09216	
X(15)	0.00369285	0.00160697	0.563	2.30	0.026	0.271283	0.08602	

Model 3 -- Backward Stepwise 6 Predictor Model

SQUARED MULTIPLE CORRELATION	0.29664
MULTIPLE CORRELATION	0.54465
ADJUSTED SQUARED MULT. CORR.	0.19850
RESIDUAL MEAN SQUARE	0.001863
STANDARD ERROR OF EST.	0.043165
F-STATISTIC	3.02
NUMERATOR DEGREES OF FREEDOM	6
DENOMINATOR DEGREES OF FREEDOM	43
SIGNIFICANCE (TAIL PROB.)	0.0148

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ	
	INTERCEPT	0.756732	0.264225	15.695	2.86	0.006			
2	X(2)	0.00364014	0.00121855	0.506	2.99	0.005	0.569933	0.14597	
5	X(5)	-0.0119911	0.00589071	-0.357	-2.04	0.048	0.530530	0.06778	Lowest Contribution to R-SQ
6	X(6)	-0.00352246	0.00164422	-0.436	-2.14	0.038	0.394513	0.07507	
12	X(12)	-0.00158083	0.000714826	-0.341	-2.21	0.032	0.688146	0.08000	
13	X(13)	-0.0522187	0.0208053	-0.394	-2.51	0.016	0.664260	0.10304	
15	X(15)	0.00261129	0.00126449	0.398	2.07	0.045	0.439962	0.06976	

Model 4 -- All Possible Subsets Best 5 Predictor Model

SQUARED MULTIPLE CORRELATION	0.26988
MULTIPLE CORRELATION	0.51950
ADJUSTED SQUARED MULT. CORR.	0.18691
RESIDUAL MEAN SQUARE	0.001890
STANDARD ERROR OF EST.	0.043476
F-STATISTIC	3.25
NUMERATOR DEGREES OF FREEDOM	5
DENOMINATOR DEGREES OF FREEDOM	44
SIGNIFICANCE (TAIL PROB.)	0.0138

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ	
	INTERCEPT	0.294943	0.253049	6.117	1.17	0.250			
2	X(2)	0.00425862	0.00124108	0.592	3.43	0.001	0.557368	0.19538	
5	X(5)	-0.0169603	0.00589221	-0.506	-2.88	0.006	0.537929	0.13748	2nd Highest Contribution to R-SQ
9	X(9)	0.00166493	0.000859501	0.280	1.94	0.059	0.793122	0.06226	
12	X(12)	-0.00139223	0.000709751	-0.300	-1.96	0.056	0.708118	0.06385	
13	X(13)	-0.0408527	0.0205961	-0.308	-1.98	0.054	0.687627	0.06529	

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

Alternatives in Analyzing the Solomon Four Group Design

Isadore Newman and Carolyn Benz
The University of Akron

John Delane Williams
University of North Dakota

Abstract

This paper dealt with an alternative approach of a Solomon four group design. Earlier writings of Solomon and others have indicated that there should be a more sophisticated approach to the statistical analysis of this research design. The suggested approach presented in this paper allows one to take advantage of pre-test scores when they exist, thereby reducing the error term and making the analysis more powerful.

Introduction

Solomon (1949) first introduced the Four Group Design, citing the paradoxical situation presented by the experimental group-control group comparison strategy in use at that time; i.e., that comparisons of posttest scores on an experimental group having taken a pretest with one control group which has taken the pretest and a second control group which has not had the pretest actually may reduce the treatment effects as they were being measured. Solomon noted that "more sophisticated statistical procedures, such as an adaptation of the analysis of covariance...in particular the mathematical nature of...the interaction term, needs to be investigated" (p. 146). Thus he suggested what has come to be known as the Solomon Four Group Design, diagrammed below:

R	O_1	X	O_2
R	O_3		O_4
R		X	O_5
R			O_6

Campbell and Stanley (1963) cite this design as the first consideration of external validity factors, and that "both the main effects of testing and the interaction of testing and X are determinable" (p. 25). This very powerful design has become frequently used, and often referenced. It would appear that there has tended to be more written and discussed on the design than on the statistical analysis utilized to answer the questions that can be reflected by this design.

Purpose

The purpose here is to demonstrate alternative strategies to analyzing the four group design that can add to the questions researchers may wish to investigate. For example, when (only) a two way analysis of variance is used to analyze Solomon type data there is much information available that is not being statistically addressed.

Alternative approaches are herein shown that utilize more of the information and may be able to reflect questions not considered previously. The analyses presented are based upon a conceptual work completed earlier by these authors (Newman, Benz & Williams, 1980). Solomon's 1949 statement is perhaps even more relevant today; i.e., that the

> Control group design seems to have awaited the development of statistical concepts which allow for the characterization of group performances in terms of measures of central tendency; and, psychologists seem to have been slow to combine statistical sophistication with experimental design. (p. 137)

Perhaps a more "statistically sophisticated" (in Solomon's terms) analysis can be suggested that adds to both the utility and the effectiveness of this research design.

Newman et al. (1980) earlier considered a repeated measures design while conducting t-tests among subjects, some of whom who had been pretested and some of whom who were not pretested. That research demonstrated an increase in power using what was termed the "independent-dependent simultaneous t-test." While this presentation is not concerned with t-tests, conceptually there is a similarity with the Solomon Four Group Design strategies, including writing models that reflect the research question using more of the available information than has typically been done. Williams and Newman (1982) earlier considered the Solomon Four Group Design to be a three-way analysis of variance with two empty cells.

It is useful to address the data as both a two way analysis of variance (experimental/control and pretested/ not pretested) and also as a psuedo-analysis of covariance, albeit the covariate is missing for two of the groups. The data in Table 1 is used in both analyses.

TABLE 1

Data for Analyzing Solomon Type Data for Two Way Analysis of Covariance and a Pseudo-Analysis of Covariance

Pre	Y	X_1	X_2	X_3	X_4	X_5	X_6
5	15	1	0	0	0	1	1
7	12	1	0	0	0	1	1
5	10	1	0	0	0	1	1
12	17	1	0	0	0	1	1
6	11	1	0	0	0	1	1
5	8	0	1	0	0	0	1
4	7	0	1	0	0	0	1
4	8	0	1	0	0	0	1
6	6	0	1	0	0	0	1
6	6	0	1	0	0	0	1
0	11	0	0	1	0	1	0
0	8	0	0	1	0	1	0
0	10	0	0	1	0	1	0
0	9	0	0	1	0	1	0
0	12	0	0	1	0	1	0
0	9	0	0	0	1	0	0
0	8	0	0	0	1	0	0
0	6	0	0	0	1	0	0
0	3	0	0	0	1	0	0
0	4	0	0	0	1	0	0

Where

Pre = the pretest score if present; 0 if no pretest score;

Y = the posttest score;

X_1 = 1 if a member of the experimental group that is pretested, 0 otherwise;

X_2 = 1 if a member of the control group that is pretested, 0 otherwise;

X_3 = 1 if a member of the experimental group that is pretested, 0 otherwise;

X_4 = 1 if a member of the control group that is not pretested, 0 otherwise;

X_5 = 1 if a member of either experimental group, 0 otherwise; and

X_6 = 1 if a member of either pretested group, 0 otherwise.

One of the various ways of accomplishing a two way analysis of variance is to use four linear models:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e_1; \quad [1]$$

$$Y = b_0 + b_5X_5 + e_2; \quad [2]$$

$$Y = b_0 + b_6X_6 + e_3; \text{ and} \quad [3]$$

$$Y = b_0 + b_5X_5 + b_6X_6 + e_4, \quad [4]$$

where the b_i^* are regression coefficients and are unique to each equation.

Focusing on the sums of squares, SS_1 = 150.00; SS_2 = 125.00; SS_3 = 20.00; and SS_4 = 145.00. Also SS_T = 224.00 and SS_W = 74.00. The interaction sum of squares is given by $SS_1 - SS_4$ = 150.00 - 145.00 = 5.00. These results can easily be incorporated into a summary table; see Table 2.

TABLE 2

Summary Table for the Two Way Analysis of Variance of Posttest Data in a Solomon Design

Source of Variation	df	SS	MS	F
Experimental-Control	1	125.00	125.00	27.03
Pretested-Not Pretested	1	20.00	20.00	4.32
Interaction	1	5.00	5.00	1.08
Within	16	74.00	4.625	

The thrust of the Solomon design is focused on testing the second and third listed sources of variation, whether or not a group was pretested and the interaction. Some might claim that the interaction effect may even be the more important test in a Solomon design. It is worthwhile to focus on the hypothesis tested as the interaction: $\bar{Y}_1 - \bar{Y}_2 = \bar{Y}_3 - \bar{Y}_4$. A reparameterization of equation 1 (a full model) is given:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1, \qquad [5]$$

then the hypothesis, in terms of the regression coefficients in equation 5 is:

$$b_1 - b_2 = b_3 - b_4 \text{ or } b_1 = b_3 + b_2 - b_4.$$

Imposing this restriction on equation 5 yields:

$$Y = (b_3 + b_2 - b_4)X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_5;$$

$$Y = b_2(X_2 + X_1) + b_3(X_3 + X_1) + b_4(X_4 - X_1) + e_5.$$

Letting $V_1 = X_2 + X_1$, $V_2 = X_3 + X_1$ and reparameterizing by letting $b_4 = 0$, $Y = b_0 + b_2V_1 + b_3V_2 + e_5$. [6]

The use of equation 6 yields $SS_6 = 145.00$, so that the interaction sum of squares would be $SS_1 - SS_6 = 150.00 - 145.00 = 5.00$, yielding the same sum of squares as previously found for interaction.

Considering a Psuedo-Analysis of Covariance

One approach to simultaneously using all the data is to use the pretest as a covariate for those individuals when a pretest is available. The linear model can be given as

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_pPre + e_6. \qquad [7]$$

What are the outcomes of using this psuedo-analysis of covariance? The pretest-posttest effect is partially nested in the covariate. If interest is centered on the adjusted means, adjusting for covariate differences for the groups that are pretested, but having the non-pretested group left alone, the adjusted means are identical to the adjusted means were the non-pretested groups completely eliminated from the analysis; in either case, the within regression coefficient is .55264. In making these covariate adjustments, care must be taken to avoid mechanically assuming that those who have not been pretested have a pretest score of zero and adjust accordingly (some computer programs in fact might do this). Any multiple comparison of interest can be done in the presence of the covariate for those pretested. If the interaction hypothesis is of interest, $\bar{Y}_1 - \bar{Y}_2 = \bar{Y}_3 - \bar{Y}_4$ which as before, translates to $b_1 = b_3 + b_2 - b_4$. A reparameterized full model is given in equation 8:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_pPre + e_6. \qquad [8]$$

When the restriction $b_1 = b_3 + b_2 - b_4$ is imposed,

$$Y = (b_3 + b_2 - b_4)X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_pPre + e_7, \text{ or}$$

$$Y = b_2(X_2 + X_1) + b_3(X_3 + X_1) + b_4(X_4 - X_1) + b_pPre + e_7.$$

Letting $V_1 = X_2 + X_1$ and $V_2 = X_3 + X_1$ and reparameterizing by letting $b_4 = 0$ (all as before)

$$Y = b_0 + b_2V_1 + b_3V_2 + b_pPre + e_7. \qquad [9]$$

The hypothesis for overall experimental-control differences is given by $\bar{Y}_1 + \bar{Y}_3 = \bar{Y}_2 + \bar{Y}_4$; in terms of the regression coefficients, $b_1 + b_3 = b_2 + b_4$ or $b_1 = b_2 + b_4 - b_3$. Imposing this restriction on equation 8 yields $Y = (b_2 + b_1 - b_3)X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_pPre + e_8$, or $Y = b_2(X_2 + X_1) + b_3(X_3 - X_1) + b_4(X_4 + X_1) + b_pPre + e_8$. Letting $V_1 = X_2 + X_1$ and $V_3 = X_3 - X_1$, and reparameterizing by letting $b_4 = 0$,

$$Y = b_0 + b_2V_1 + b_3V_3 + b_pPre + e_8. \qquad [10]$$

To address the pretested-not pretested effect, the restriction, $b_1 + b_2 = b_3 + b_4$, or $b_1 = b_3 + b_4 - b_2$, corresponding to the hypothesis $\bar{Y}_1 + \bar{Y}_2 = \bar{Y}_3 + \bar{Y}_4$, can be placed on equation 8, yielding $Y = (b_3 + b_4 - b_2)X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_pPre + e_9$, and $Y = b_2(X_2 - X_1) + b_3(X_3 + X_1) + b_4(X_4 + X_1) + b_pPre + e_9$; letting $V_4 = X_2 - X_1$, $V_2 = X_3 + X_1$ and reparameterizing by letting $b_4 = 0$,

$$Y = b_0 + b_2V_4 + b_3V_2 + b_pPre + e_9. \qquad [11]$$

It should be pointed out that, though this test can be accomplished for the data at hand, a more useful test of this hypothesis could be completed if an independent covariate or covariates are available; if the pretest is used as a covariate, the pretesting effect is partially nested in the pretest scores used as a covariate. A model for the covariate can also be given:

$$Y = b_0 + b_P Pre + e_{10} \qquad [12]$$

A summary table for this psuedo-analysis of covariance can be formed; see Table 3. In Table 3, $SS_W = 62.39$ from the use of the full model (equation 7); $SS_7 = 161.61$. For the interaction, $SS_{INTERACTION} = SS_7 - SS_9$ (which yields 161.61 - 160.72, or $SS_{INTERACTION} = .89$. For the experimental control difference, $SS_{EXP/CONTROL} = SS_7 - SS_{10}$; $SS_{EXP/CONTROL} = 161.61 - 74.21 = 87.40$. The pretesting effect is given by $SS_7 = SS_{11} = 161.61 - 160.10 = 1.51$. The sum of squares for the covariate is given by $SS_{12} = 54.04$. These results are shown in Table 3.

TABLE 3

Summary Table for the Psuedo-Analysis of Covariance with a Solomon Design

Source of Variation	df	SS	MS	F
Covariate	1	54.04		
Pretest-No Pretest	1	1.51	1.51	.36
Experimental-Control	1	87.40	87.40	21.01
Interaction	1	.89	.89	.22
Within	15	62.39	4.16	

It should be clear that the summary table for this psuedo-analysis of covariance is not additive. Finally the adjusted means for the pretested groups can be found:

$\bar{Y}_1(adj) = \bar{Y}_1 - b_W(\bar{X}_1 - \bar{X}_T)$ or

$Y_1(adj) = 13 - .55264(7 - 6)$ or 12.45; for

$\bar{Y}_2(adj) = 7 - .55264(5 - 6)$ or 7.55.

Discussion and Conclusions

An essential issue for the Solomon Four Group Design is in regard to the experimenter's expectations in choosing the design. Is the design chosen as a panacea to rid the analysis of unwanted alternative interpretations, i.e., doesn't this design come with certain "warranties?" If so, choosing this design (or any other) is just another misstep in searching for the "holy grail." Alternative interpretations of literally any data analysis would seem not only to be a constant, but also a welcome constant, particularly to those who subscribe to Popper's view (as cited in Griffin, 1988) of scientists who actively seek evidence to refute their pet theories. Our own recommendation regarding data analysis (including the Solomon Four Group Design) is to first formulate the research process so that the precise questions of interest can be answered. Then state hypotheses and linear models that precisely address those questions. Beyond this, also recognize that a myriad of other issues can distort interpretations. In addition to the issues addressed in Campbell and Stanley (1963) and Cook and Campbell (1979), other concerns that may have different readings by other diligent investigators have to be considered, including issues regarding the criterion (or criteria)--do they in fact measure what they are claimed to measure? Do those who disagree with the use of a particular measure of a given construct as a measure of that construct have any validity in their arguments? Similar issues regarding experimental groups or definitions of the independent variables also come into play. In a more relativistic vein than is our practice, there probably are no final solutions; data and their interpretations would seem always to be subject to reanalysis and reinterpretation.

References

Campbell, D., T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally College Publishing Company.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation design and analysis issues for field settings. Chicago, IL: Rand McNally College Publishing Company.

Griffin, D. R. (1988). Of minds and molecules: Postmodern medicine in a psychosomatic universe. In Griffin, D. (Ed.), The re-enchantment of science. Albany, NY: SUNY Press.

Newman, I., Benz, C., & Williams, J. D. (1980). Repeated measures: Suggested approach for simultaneous independent and dependent measures. Paper presented at the meeting of the American Psychological Association, Montreal, Canada.

Solomon, R. L. (1949). An extension of control group design. Psychological Bulletin, 46, 137-150.

Williams, J. D., & Newman, I. (1982). Using linear models to simultaneously analyze a Solomon four group design. Multiple Linear Regression Viewpoints, 11(3), 77-90.

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

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

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

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

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

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

THE UNIVERSITY OF AKRON
AKRON, OH 44325-4205

Title

ISSN 0195-7171