Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

How Many Subjects Does It Take To Do A Regression Analysis

Samuel B. Green

Multivariate Behavioral Research, *26* (3), 499-510

How Many Subjects Does It Take To Do A Regression Analysis?

Samuel B. Green
University of Kansas

Numerous rules-of-thumb have been suggested for determining the minimum number of subjects required to conduct multiple regression analyses. These rules-of-thumb are evaluated by comparing their results against those based on power analyses for tests of hypotheses of multiple and partial correlations. The results did not support the use of rules-of-thumb that simply specify some constant (e.g., 100 subjects) as the minimum number of subjects or a minimum ratio of number of subjects (N) to number of predictors (m). Some support was obtained for a rule-of-thumb that $N \geq 50 + 8\,m$ for the multiple correlation and $N \geq 104 + m$ for the partial correlation. However, the rule-of-thumb for the multiple correlation yields values too large for N when $m \geq 7$, and both rules-of-thumb assume all studies have a medium-size relationship between criterion and predictors. Accordingly, a slightly more complex rule-of-thumb is introduced that estimates minimum sample size as function of effect size as well as the number of predictors. It is argued that researchers should use methods to determine sample size that incorporate effect size.

"How many subjects does it take to do a regression analysis?" It sounds like a line delivered by a comedian at a nightclub for applied statisticians. In fact, the line would probably bring jeers from such an audience in that applied statisticians routinely are asked some variant of this question by researchers and are forced to respond with questions of their own, some of which have no good answers. In particular, because many researchers want a sample size that ensures a reasonable chance of rejecting null hypotheses involving regression parameters, applied statisticians are likely to present their responses within a power analytic framework. From this perspective, sample size can be determined if three values are specified: alpha, the probability of committing a Type I error (i.e., incorrectly rejecting the null hypothesis); power, one minus the probability of making a Type II error (i.e., not rejecting a false null hypothesis); and effect size, the degree to which the criterion variable is related to the predictor variables in the population. Although alpha by tradition is set at .05, the choice of values for power and effect size is less clear and, in some cases, seems rather arbitrary.

As an alternative to determining sample size based on power analytic techniques, some individuals have chosen to offer rules-of-thumb for regression analyses. These rules-of-thumb come in various forms. One form indicates that the number of subjects, N, should always be equal to or greater than some constant,

A (i.e., $N \geq A$), while a second form stipulates a recommended minimum ratio B of subjects-to-predictors (i.e., $N \geq B\,m$ where m is the number of predictors). Finally, a third form is a more general rule that encompasses the first two (i.e., $N \geq A + B\,m$).

Tabachnick and Fidell (1989) suggest, with some hesitancy, in their multivariate text that the minimum number of subjects for each predictor or independent variable (IV) in a regression analysis should be 5-to-1. They state the following:

> If either standard multiple or hierarchical regression is used, one would like to have 20 times more cases than IVs. That is, if you plan to include 5 IVs, it would be lovely to measure 100 cases. In fact, because of the width of the errors of estimating correlation with small samples, power may be unacceptably low no matter what the cases-to-IVs ratio if you have fewer than 100 cases. However, a bare minimum requirement is to have at least 5 times more cases than IVs — at least 25 cases if 5 IVs are used. (pp. 128-129)

In the first edition of a multivariate primer by Harris (1975), he recommended that the number of subjects $N > 50 + m$. In the revision of his text, he states,

> ... not too different from the *Primer's* earlier suggestion that $N - m$ be > 50. More common is a recommendation that the ratio of N to m be some number, for example, 10. I know of no systematic study of the ratio versus the difference between N and m as the important determinant of sampling stability of the regression weights. However, ratio rules break down for small values of m ... and there are hints in the literature that the difference rule is more appropriate. (1985, p. 64.)

Harris' difference rule is an example of the general rule-of-thumb of $N \geq A + B\,m$ with $A = 50$ and $B = 1$. Harris' call for a study to compare the ratio and difference rules can be expanded and recast within the context of the present paper as a request for a study to evaluate whether the A constant is a necessary component of the general rule-of-thumb. If A is unnecessary, the general rule-of-thumb simplifies to the ratio rule-of-thumb, $N \geq B\,m$.

Nunnally (1978) makes slightly different recommendations based on an examination of an equation for determining an unbiased estimate of the population squared multiple correlation coefficient ($\hat{R}^2$) from the sample squared multiple correlation coefficient (R^2):

$$\hat{R}^2 = 1 - (1 - R^2)[(N - 1)/(N - m)]. \quad (1)$$

He states,

> If there are only 2 or 3 independent variables and no preselection is made among them, 100 or more subjects will provide a multiple correlation with little bias. In that case,

> if the number of independent variables is as large as 9 or 10, it will be necessary to have from 300 to 400 subjects to prevent substantial bias. (p. 180)

Other individuals, based on a variety of justifications specify somewhat different rules-of-thumb. For example, Marks (1966), as cited in Cooley and Lohnes (1971), recommended a minimum of 200 subjects for any regression analysis, while Schmidt (1971) suggested a minimum subject-to-predictor ratio ranging in value from 15-to-1 to 25-to-1. Still other individuals, for example, Pedhazur (1982), discuss rules-of-thumb, but make no general set of recommendations themselves.

An important question is whether researchers who use these rules-of-thumb have designed studies with adequate power. The answer to this question is unknown; however, it is known that many empirical studies do have insufficient power (for a review of this literature, see Cohen, 1988). Perhaps some portion of these studies have used rules-of-thumb for regression analysis (e.g., 5-to-1 subject-to-predictor ratio). The purpose of the present study is to determine if researchers who apply such rules-of-thumb are designing studies with low power. Sample sizes for regression analyses will be determined following recommendations made by Cohen (1988) in the second edition of his book on power analysis. These results will be compared against various rules-of-thumb to judge their adequacy. It was anticipated that none of the reviewed rules-of-thumb would be satisfactory in that all of them ignore effect size and, accordingly, recommendations for more complex rules-of-thumb would be required. It is hoped that this presentation will encourage researchers to struggle with the difficult decisions required by power analysis with the realization that simplistic rules-of-thumb ignore the idiosyncratic characteristics of research studies. In addition, it is hoped that the discussion will encourage applied statisticians to develop methods for the determination of sample size that researchers find less esoteric and more useful (see Harris, in press, for one such alternative).

The methods employed in this study impose a few limitations on the conclusions. The sample sizes were based on power tables presented in Cohen (1988) and are slightly different from those that would be obtained using other power tables (for a more in-depth discussion, see Gatsonis & Sampson, 1989). Also, the power analyses assume that the regression analyses include all predictors and do not allow for the preselection of predictor variables (e.g., stepwise regression). More generally, in order to have a focused presentation, it was necessary to develop methods which suggest that decision making with regression analysis is simpler than what should occur in practice. For example, researchers need to consider carefully prior to determining sample size whether they can reduce the number of predictor variables by forming a priori linear combinations of two or more predictors and whether the variables are reliable and are highly interrelated. Most importantly, researchers need to attend seriously to the issue of estimating effect size based on their knowledge of a research area and the methods of their study rather than assume a value for an effect size offered by Cohen or discussed in this paper (see Cohen, 1988 for a detailed discussion).

Sample Sizes Required to Evaluate Multiple Correlation Coefficients with a Power of .80

I began by considering what sample size is required to evaluate the hypothesis that the multiple correlation between the predictors and the quantitative variable (y) is equal to zero with a power of .80. To conduct power analyses, choices of values for alpha, power, and effect size were made:

1. Alpha was set at .05, the traditional level of significance.

2. Power was set at .80, a value proposed by Cohen (1988) as appropriate for a wide range of behavioral research areas. He argued that the setting of power, the probability of not committing a Type II error, is to some extent arbitrary, but should be dependent on the loss associated with making this error. Inasmuch as power is partially a function of alpha, it also should be a function of the loss associated with a Type I error. He suggested that typically across the behavioral sciences, a 4 to 1 ratio reflects the relative seriousness of a Type I error to a Type II error. Consequently, when alpha is set equal to .05, the probability of a Type II error should be $4 \times .05 = .20$ and power would be $1 - .20 = .80$.

3. Cohen (1988) stresses two indexes of effect size for regression analysis, f^2 and the better known R^2. The two indexes are directly related:

$$f^2 = R^2 / (1 - R^2). \quad (2)$$

Although Cohen argues that the choice of values for effect size (R^2 or f^2) should depend on the research area, he proposes, as a convention, R^2s of .02, .13, and .26 (f^2 of .02, .15, and .35) to serve as operational definitions for the descriptors small, medium, and large, respectively. Cohen discusses these selections rather extensively and indicates that they agree with his subjective judgment of small, medium, and large effect sizes obtained in behavioral sciences. These three values are used in the current study.

Calculations as outlined by Cohen (1988) were performed to determine sample sizes for the selected values of alpha, power, and effect size using tables (pp. 448-455) he provided for this purpose. The tables have entries in which the number of predictors for regression analyses may assume 23 different values, ranging from 1 predictor to 120 predictors. While sample sizes were determined for all 23 values, for sake of simplicity, only those for 15 of the 23 values are presented in the first three columns of Table 1. Sample sizes are not presented in

Table 1
Sample Size Required to Test the Hypothesis that the Population Multiple Correlation Equals Zero with a Power of .80 (Alpha = .05)

Number of predictors	Sample sizes based on power analysis			Sample sizes based on new rule-of-thumb		
	Effect size			Effect size		
	Small	Medium	Large	Small	Medium	Large
1	390	53	24	400	53	23
2	481	66	30	475	63	27
3	547	76	35	545	73	31
4	599	84	39	610	81	35
5	645	91	42	670	89	38
6	686	97	46	725	97	41
7	726	102	48	775	103	44
8	757	108	51	820	109	47
9	788	113	54	860	115	49
10	844	117	56	895	119	51
15	952	138	67	1045	139	60
20	1066	156	77	1195	159	68
30	1247	187	94	1495	199	85
40	1407	213	110	1795	239	103

Table 1 for regression analyses with greater than 40 predictors because data analyses with this many predictors were assumed to be relatively infrequent.

Harris' (1975) rule-of-thumb that $N \geq 50 + m$ reflects to some extent the results given in Table 1. A rule-of-thumb, by definition, should require minimal complexity and, therefore, in the determination of sample size, it might be argued that the rule-of-thumb should be developed to give accurate answers for typical studies rather than for all studies. According to Cohen (1988), a typical study in the behavioral sciences would have a medium effect size, and the minimum number of subjects required for studies with this effect size according to Table 1 is 53 (when $m = 1$), similar to the minimum number based on the rule-of-thumb ($50 + m = 51$). Also, both the tabled sample sizes and those based on Harris' rule-of-thumb indicate that the number of subjects increases as the number of predictors increases. However, as the number of predictors increases, the differences in the sample sizes from Harris' rule-of-thumb and those from power analyses become larger. One possible revision would be that the minimum number of subjects should be 50 + 8 *m*. This rule-of-thumb is fairly accurate for medium effect-size studies with a small number of predictors ($m < 7$), but becomes increasingly more conservative with additional predictors (i.e., the rule-of-thumb overestimates the required sample size according to power analyses).

Even though an argument can be made for the rule-of-thumb that $N \geq 50 + 8m$, it has its drawbacks. This rule, in particular, and traditional rules-of-thumb, in general, have two problems associated with them. First, their mathematical simplicity does not offset their inability to take into account effect size. The number of subjects required for conducting a regression analysis with a small effect size is dramatically different from the number of subjects required for an analysis with a large effect size. Accordingly, it is important that researchers consider the magnitude of the effect they want to detect when determining the number of subjects to include in their study rather than view their study as a typical study in the behavioral sciences and base their sample size on a medium effect size. Second, even if effect size is held constant, the number of subjects is not linearly related to the number of predictors as indicated by the general form of the rule-of-thumb. Instead, the number of subjects required for each additional predictor is greater with few predictors than with many predictors.

In order to better reflect the results of power analyses, a new two-step rule-of-thumb was developed based on Cohen's (1988) power analytic approach. Cohen's procedure requires the determination of lambda, which is obtained by entering tables (pp. 448-455) in his book. With the new rule-of-thumb, lambda is approximated by L in step 1. Once L is determined, N can be computed in step 2 using a simple equation given by Cohen. The two steps of the rule-of-thumb are as follows:

1. Compute L. L is 8 with a single predictor. For regression analyses with 2 through 10 predictors, L increases for each additional predictor by 1.5, 1.4, 1.3, 1.2, 1.1, 1.0, .9, .8, and .7, respectively. Algebraically, for $m < 11$,

$$L = 6.4 + 1.65m - .05m^2. \tag{3}$$

For each additional predictor past 10, L increases .6.

2. Compute required minimum sample size, N. $N \geq L / f^2$ where $f^2 = R^2 / (1 - R^2)$.

For example, with 7 predictors, $L = 8.0 + 1.5 + 1.4 + 1.3 + 1.2 + 1.1 + 1.0 = 15.5$ (or using Equation 3, $6.4 + 1.65 (7) - .05 (7^2) = 15.5$). For a medium effect size (R^2 of .13 and f^2 of .15), $N \geq 15.5 / .15 = 103$. This answer is close to the value of 102, the sample size derived from conducting a power analysis (see Table 1).

If this new rule-of-thumb is to replace current rules-of-thumb, it needs to be simple mathematically and easy to remember, as well as relatively accurate. Although the rule is more complicated than traditional rules-of-thumb, it is not complex. Values for L in step 1 do progress in a systematic fashion and the equation in step 2 is simple. Also, the results of the new rule-of-thumb agree moderately well with the sample sizes determined by power analytic methods, as can be observed by comparing the last three columns of Table 1 with the first three columns. It is most accurate for moderate effect sizes; the sample sizes based on the rule-of-thumb and on power analyses never differed by more than 5% from each other if $m \leq 20$. These results compare favorably to those based on the rule-of-thumb that $N \geq 50 + 8m$; the sample sizes based on this traditional rule-of-thumb and on power analyses differed by more than 5% from each other when $m = 1$ and when $m \geq 8$. For small effect sizes, the new rule-of-thumb is reasonably accurate when m is small; however, it produces an N that is consistently larger than the one based on power analyses as m increases, with the two differing by 10% or more when $m \geq 20$. For a large effect size, the new rule-of-thumb always underestimates the sample sizes based on power analyses, although the degree of underestimation is not great when m is small.

In summary, no specific minimum number of subjects or minimum ratio of subjects-to-predictors was supported. The general rule-of-thumb of $N \geq 50 + 8m$ was seen as more accurate than these simpler rules-of-thumb. However, a new rule-of-thumb was found to have even greater accuracy, although it requires that researchers estimate the effect size that they wish to detect. An argument could be made that this new two-step rule-of-thumb is sufficiently complex that researchers could almost as easily do a power analysis to determine their sample sizes. Researchers are encouraged to conduct such power analyses.

Sample Sizes Required to Evaluate Partial Correlation Coefficients with a Power of .80

When conducting multiple regression analyses, researchers typically evaluate not only hypotheses that population multiple correlations are equal to zero, but also hypotheses that population semi-partial correlations are equal to zero and/or population partial correlations are equal to zero (Cohen & Cohen, 1983). The test statistics to evaluate these latter two hypotheses are identical; therefore, discussion will center around a test for just one of these coefficients, the partial correlation. Researchers try to understand their multiple regression results in part by evaluating partial correlations between a criterion variable, y, and a set of predictor variables, partialling out the effect of a second set of predictor variables from both y and the first set of predictors. For example, in investigating the relationship between college grade point average (CGPA) and the predictors of high school grade point average (HSGPA) and Scholastic Aptitude Tests (SATs), researchers might focus their attention on three correlations: the multiple correlation of CGPA with HSGPA and SATs, the partial correlation between CGPA and HSGPA holding constant (partialling out) SATs, and the partial correlation between CGPA and SATs holding constant HSGPA. The latter two partial correlations are evaluated to determine whether one of the predictors enhances the predictability of CGPA if the other predictor is held constant or whether a single predictor is sufficient. Because tests of partial correlations are frequently conducted, the power of these tests should be considered when evaluating sample-size rules-of-thumb for regression analyses.

How to incorporate the partial correlation into our assessment of these rules-of-thumb was problematic. For any multiple regression analysis, a researcher might calculate a number of partial correlations. For example, with three predictors, twelve partial correlations could be computed: six correlations relating y to one of the predictors holding a second predictor constant, three correlations relating y to two of the predictors holding a third predictor constant, and three correlations relating y to one of the predictors holding the other two predictors constant. The number of subjects required for these analyses based on power analysis differs.

It was decided to focus on one type of partial correlation, the correlation between y and a predictor holding all other predictors constant, because the significance test for this correlation is often performed by researchers. However, the test may not be labeled as evaluating this partial correlation in that tests of other statistics from regression analyses produce the same results as the one for this partial correlation. Specifically, the test that the partial correlation between y and a predictor x_1 holding constant all other predictors (x_2 through x_m) is equal to zero is equivalent to a test that the weight for x_1 in a multiple regression equation containing all m predictors is equal to zero or that the variance of y accounted for by x_1 over and above x_2 through x_m (the squared semipartial or unique variance) is equal to zero. Most computer regression programs include tests of one or more of these statistics. For example, the General Linear Models Procedure (PROC GLM) in the Statistical Analysis System (SAS, 1985) gives tests for the regression weights and the unique variances (referred to as Type III sums of squares). Undoubtedly, researchers report these tests in part because they are standard output of regression packages.

Sample Sizes Based on Effect Sizes Given by Cohen

To determine the sample size required to reject the hypothesis that the partial correlation is equal to zero, values for alpha, power, and effect size that were used to determine sample size for the multiple correlation were initially chosen. Alpha was set at .05; power was set at .80; and the squared partial correlation was set at .02, .13, or .26. Cohen (1988) indicates that .02, .13, and .26 represent small, medium, and large effect sizes for not only the squared multiple correlation, but also for the squared multiple partial correlation, of which the squared partial correlation is the special case being considered here. Using Cohen's (1988) power analysis approach, the sample sizes for the three effect sizes were determined. The obtained sample sizes may be found using the following simple rule:

1. Determine the number of subjects required to test the hypothesis that the population correlation between y and a single predictor is equal to 0 (first row of Table 1).

2. Add m - 1 to the result of step 1.

When the squared partial correlation between y and a predictor holding (m - 1) other predictors constant is .02, the required number of subjects is 390 + (m - 1); when it is .13, $N \geq 53 + (m - 1)$; and when it is .26, $N \geq 24 + (m - 1)$.

The rule-of-thumb recommended by Harris (1975) offers a good approximation to the sample sizes obtained from power analyses when the squared partial correlation is .13, Cohen's definition of a medium effect size. In particular, Harris suggested that $N > 50 + m$, while the power analysis for an effect size of .13 indicated that $N \geq 52 + m$. However, within this perspective, the rule is rather inflexible in that it allows for only a single effect size, a squared partial correlation of .13. Also, as discussed below, this rule-of-thumb may not be appropriate even for the typical behavioral science study because .13 may be too large a value for a medium effect size.

Sample Sizes Based on Revised Effect Sizes

Cohen (1988) suggested that effect sizes should be greater for correlational analyses which may have multiple predictors, such as multiple and partial correlations, than for zero-order correlational analyses. Therefore, he suggested small, medium and large effect sizes of .02, .13, and .26 for squared multiple and multiple partial correlations and .01, .09, and .25 for squared zero-order correlations. However, the partial correlation of interest here is not a multiple partial correlation, but is a partial correlation between y and a single predictor, holding all other predictors constant. Even the values for a zero-order correlation of .01, .09 and .25 might be too large in that a partial correlation between y and x controlling for all other predictors frequently will be somewhat smaller than a zero-order correlation between y and x. The zero-order correlation may exceed the partial correlation if suppression occurs among the predictors (for a discussion of suppression, see Cohen & Cohen, 1983; Pedhazur, 1982) or if the correlation between x and the other predictors is close to zero (particularly if the multiple correlation is large, for example, .70 or greater). However, some researchers (e.g., Nunnally, 1978) have suggested that suppression occurs rarely in practice. Given that suppressors are rare and given that in many research studies predictor variables are substantially intercorrelated, a partial correlation between two variables is likely to be smaller than the zero-order correlation between them. Therefore, a medium effect size for a squared partial correlation might be redefined to be .07 rather than .13, Cohen's suggested value for a squared partial correlation, or .09, his suggested value for a squared zero-order correlation.

If the squared partial correlation of interest is .07, the required sample size for regression analyses using Cohen's power tables is $N \geq 104 + m$. It should be noted that a similar value may be obtained by initially using the two-step rule-of-thumb presented for multiple correlations: (a) $L = 8$ with a single predictor and (b) $f^2 = .07 / (1 - .07) = .075$, substituting partial correlations for multiple correlations in the f^2 equation, and $L / f^2 = (8 / .075) = 107$. Based on these calculations, $N \geq 107 + (m - 1) = 106 + m$. Regardless of the approach, the choice of values for an effect size has a dramatic effect. Based on power calculations, the minimum sample size is $52 + m$ if a medium effect size is .13 or $104 + m$ if it is .07.

Although the resulting equations for sample size with the partial correlation all have the general rule-of-thumb form that $N \geq A + m$, no one value for A is satisfactory in that A is a function of effect size (as well as alpha and power). Researchers should be allowed to determine the value for the partial correlation and use it to establish sample size using power analysis or, if unable to access power tables, the rule-of-thumb that $N \geq (8 / f^2) + (m - 1)$.

Conclusion

Researchers who use a rule-of-thumb rather than power analyses are trading simplicity of use for accuracy and specificity of response. Traditional rules-of-thumb that give a minimum number of subjects or a minimum ratio of subjects-to-predictors are the simplest to use, but sample sizes based on them are rarely congruent with power analyses. Slightly more complex rules-of-thumb of the form that $N \geq A + B\,m$ show slightly better agreement with power analysis results, but only if researchers are interested in a typical effect size and hypotheses concerning a multiple correlation and/or one type of partial correlation (as well as an alpha of .05 and power of .80). More specifically, the power for a test of a multiple correlation with a medium effect size is approximately .80 or greater if $N \geq 50 + 8\,m$ and the power for a test of a medium-sized partial correlation between y and a predictor holding all other predictors constant is approximately .80 if $N \geq 104 + m$. A more complex rule-of-thumb is necessary in order for minimum sample size to be a function of effect size as well as the number of predictors for both the multiple correlation (i.e., $N \geq L / f^2$) and the partial correlation (i.e., $N \geq (8 / f^2) + (m - 1)$). Greater accuracy and flexibility can be gained beyond these rules-of-thumb by researchers conducting power analyses. Because researchers should estimate effect size based on the characteristics of their study as opposed to assuming some conventional value presumed to be appropriate for the typical study in behavioral sciences, it is recommended that researchers use either the more complex rules-of-thumb that directly incorporate effect size or conduct power analyses.

Researchers who evaluate hypotheses concerning partial correlations in addition to hypotheses about multiple correlations should determine minimum sample sizes for both types of tests and select the largest of these minimum sample sizes as the N for their study. For example, if a researcher was planning to conduct a regression analysis with 5 predictors and an estimated multiple correlation of .50 (f^2 = .33) and estimated partial correlations of .40 (f^2 = .19), he/she might use the rule-of-thumb for the multiple correlation that $N \geq L / f^2 = 13.4 / .33 = 40.2$ and the rule-of-thumb for the partial correlation that $N \geq 8 / f^2 + (m - 1) = 8 / .19 + (5 - 1) = 46$. Consequently, this researcher should conduct the study with a minimum of 46 subjects.

In conclusion, researchers who use traditional rules-of-thumb are likely to design studies that have insufficient power because of too few subjects or excessive power because of too many subjects. For example, researchers who use the rule-of-thumb of 5 subjects for each predictor (Tabachnick and Fidell, 1989) are conducting studies that have a high probability of not yielding significance unless the effect size is extremely large. In contrast, researchers who follow the recommendations of Nunnally (1978) and collect data on a minimum of 300 or 400 subjects have likely collected more data than necessary if the number of predictors are few and the effect size of medium value or greater. Of course, larger sample sizes might be justified on issues unrelated to power. These other issues must be considered on their own merits. For example, Nunnally (1978) made his recommendation based on the amount of bias (shrinkage) in sample multiple correlations rather than power. However, it is unclear why researchers should collect data from 300 or 400 subjects to minimize shrinkage rather than determine sample size based on power analyses which may yield Ns substantially less than 300 or 400. These researchers may account for the bias that occurs with smaller sample sizes by using a shrinkage equation such as Equation 1.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic.

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

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. New York: Wiley.

Gatsonis, C., & Sampson, A. R. (1989). *Multiple correlation: Exact power and sample size calculations*. *Psychological Bulletin, 106*, 516-524.

Harris, R. J. (1975). *A primer of multivariate statistics*. New York: Academic.

Harris, R. J. (1985). *A primer of multivariate statistics* (2nd ed.). Orlando, FL: Academic.

Harris, R. J. (In press). MIDS (Minimally Important Difference Significant) criterion for sample size. *Journal of Educational Statistics*.

Marks, M. R. (1966, September). *Two kinds of regression weights that are better than betas in crossed samples*. Paper presented at the meeting of the American Psychological Association.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

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). Cary, NC: SAS Institute Inc.

Schmidt, F. L. (1971). The relative efficiency of regression in simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement, 31*, 699-714.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd. ed.). Cambridge, MA: Harper & Row.