COMPUTER STUDIES

STATISTICAL POWER COMPUTATIONS FOR DETECTING DICHOTOMOUS MODERATOR VARIABLES WITH MODERATED MULTIPLE REGRESSION

HERMAN AGUINIS
University of Colorado at Denver

CHARLES A. PIERCE
Montana State University

A revised and improved version of Aguinis, Pierce, and Stone-Romero's (1994) program for estimating the statistical power of moderated multiple regression to detect dichotomous moderator variables is described. The QuickBASIC program runs on IBM and IBM-compatible personal computers and estimates power based on user-provided values for (a) total sample size, (b) sample sizes across the two moderator-based subgroups, (c) correlation coefficients between the predictor and criterion for each of the two moderator-based subgroups, (d) correlation coefficient between the predictor and hypothesized moderator, and (e) sample and population standard deviations for the predictor. Program-generated power estimates for typical research situations in education, psychology, and management indicate that hypothesis tests of moderating effects are typically conducted at insufficient levels of statistical power.

Numerous theories in education, psychology, and management posit the operation of moderator variables; that is, variables that interact with a predictor (X) in explaining variance in a criterion (Y). Variable Z is defined as a moderator of an X-Y relationship when the nature of this relationship varies across values or levels of Z (Saunders, 1956; Zedeck, 1971).

Given ongoing theoretical advancements, researchers are increasingly interested in complex relationships that go beyond main effects. Accordingly,

A previous version of this article was presented at the meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO, April 1997. We thank the members of the Behavioral Science Research Group for comments on the research reported herein. Address correspondence to Herman Aguinis, College of Business and Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217-3364; e-mail: haguinis@castle.cudenver.edu; World Wide Web: http://www.cudenver.edu/~haguinis and http://www.montana.edu/wwwpy/cppage.html.

Educational and Psychological Measurement, Vol. 58 No. 4, August 1998 668-676
© 1998 Sage Publications, Inc.

interest in moderator variables is increasing substantially in education, psychology, management, and related disciplines (e.g., Aguinis, Bommer, & Pierce, 1996; Aguinis & Pierce, in press; MacCallum & Mar, 1995; Sagie & Koslowsky, 1993).

Moderated multiple regression (MMR) has long been considered an appropriate statistical technique for estimating the effects of continuous and categorical moderator variables (Aguinis & Pierce, 1998; Cohen & Cohen, 1983; Saunders, 1956). Conducting an MMR analysis entails forming a sample-based least squares regression equation that tests the additive model for predicting a criterion Y from predictors X and Z and the interaction between X and Z (i.e., moderating effect of Z), represented by the $X \times Z$ product term as follows:

$$\hat{Y} = a + b_1X + b_2Z + b_3X \times Z, \quad (1)$$

where $\hat{Y}$ is the predicted value for Y, a is the least squares estimate of the intercept, b_1 is the least squares estimate of the population regression coefficient for X, b_2 is the least squares estimate of the population regression coefficient for Z, and b_3 is the least squares estimate of the population regression coefficient for the product term that carries information about the interaction between X and Z (Cohen & Cohen, 1983). Rejecting the null hypothesis of $\beta_3 = 0$ indicates the presence of an interaction or moderating effect. Stated differently, rejecting this null hypothesis indicates that the regression of Y on X is unequal across levels of Z (e.g., minority and nonminority subgroups, male and female subgroups).

Statistical Power Problems With MMR

MMR is a routinely used method for estimating and interpreting the effects of dichotomous moderators such as ethnicity (minority vs. nonminority) and gender (male vs. female) (Sackett & Wilk, 1994; Stone, 1988). Unfortunately, attempts to detect hypothesized moderating effects using MMR are often unsuccessful due to insufficient statistical power (Aguinis, 1995). Thus, to identify conditions that adversely affect the power of MMR, Aguinis and Stone-Romero (1997) conducted a Monte Carlo simulation that examined the main and interactive effects of the following methodological and statistical factors on the power of MMR: (a) range restriction on a continuous predictor variable X (i.e., ratio of sample to population standard deviation scores; RR), (b) total sample size (N), (c) relative proportion of cases in each of the two dichotomous moderator variable-based subgroups Z ($p_1 = n_1/N$, where $n_1 + n_2 = N$), (d) predictor-moderator intercorrelation (i.e., multicollinearity; ρ_{XZ}), and (e) magnitude of the population moderating effect (i.e., absolute difference between ρ_{XY} levels for the two moderator-based subgroups, $|\rho_{XY(1)} - \rho_{XY(2)}|$; effect size, ES). To more easily interpret the results of the simulation, the empirically obtained power rates were regressed onto values of the manipulated methodological and statistical artifacts. Results of Aguinis and Stone-Romero's simulation showed that the manipulated factors had both main and interactive effects on the power of MMR.

Regarding the main effects, results showed that predictor variable range restriction, total sample size, sample sizes across the two moderator-based subgroups, and moderating effect size each had profound effects on the power of MMR. As an example of the effects of range restriction, for a medium moderating effect size, equal number of cases in the two moderator-based subgroups, and total sample size of 300, predictor range restriction values of .80, .40, and .20 reduced power to .94, .80, and .67, respectively.

Regarding the interactive effects, results indicated that the two-way interactions between the manipulated methodological and statistical artifacts accounted for nearly 17% of the variance in the empirically generated MMR power rates above and beyond the influence of the main effects. Converting this value to Cohen's (1988) metric yielded an effect size (f^2) of 3.25, which is nearly 10 times greater than a "large" effect (i.e., $f^2 = .350$; Aiken & West, 1991). The practical implication of such sizable interactive effects is that even if conditions in research designed to detect moderating effects are favorable with respect to one or more factors that affect the power of MMR, the presence of an unfavorable level of one or more of the other factors that influence power will result in power levels that are far below Cohen's recommended .80 standard.

In short, results of Aguinis and Stone-Romero's (1997) Monte Carlo simulation indicated that the power of MMR is influenced by main and interactive effects of factors such as predictor variable range restriction, total sample size, and the number of cases in each moderator-based subgroup. It should be noted that in many research situations in education, psychology, and management, range restriction may affect the distribution of predictor scores (Hattrup & Schmitt, 1990), total sample size may be less than 100 (Russell et al., 1994), and the number of cases across levels of the hypothesized moderator often differ markedly (Trattner & O'Leary, 1980). In these common situations, failure to ascertain the presence of a dichotomous moderator variable using MMR can be explained by one of two competing hypotheses:

Hypothesis 1: The population moderating effect is equal to zero, or
Hypothesis 2: The population moderating effect is greater than zero but is not detected due to low statistical power.

In light of these two hypotheses, researchers would benefit from being able to estimate the power of MMR having knowledge of predictor range restriction, total sample size, number of cases in each of the two moderator-based subgroups, between-subgroup differences in the relationship between

variables X and Y, and predictor-moderator intercorrelation. If the resulting power estimate is greater than the recommended .80 level, MMR users could rule out Hypothesis 2 and more confidently conclude that support has been found for Hypothesis 1. Alternatively, if the resulting estimate is less than .80, null results should be interpreted as representing inconclusive findings and as an indication that further research with a greater level of power is warranted (Westermann & Hager, 1986).

The Present Article

Aguinis, Pierce, and Stone-Romero (1994) developed a computer program that provides estimates of the power to detect the effects of dichotomous moderator variables using MMR. However, the Aguinis et al. program suffers from three limitations: (a) It does not consider the effects of predictor range restriction on the power of MMR, (b) it does not include interactive effects between some factors known to affect the power of MMR (e.g., predictor range restriction × moderating effect magnitude), and (c) it was developed based on Stone-Romero, Alliger, and Aguinis's (1994) equations that assumed a linear relationship between methodological and statistical factors and the empirically obtained power rates. It has been shown, however, that the power function is nonlinear (McClelland & Judd, 1993).

To overcome these three aforementioned limitations, a revised and improved computer program was developed. The program (a) includes the effects of predictor range restriction, (b) considers the interactive effects of methodological and statistical factors ascertained by Aguinis and Stone-Romero (1997), and (c) is based on results of the Monte Carlo simulation by Aguinis and Stone-Romero who, in contrast to Stone-Romero et al. (1994), linearized the power function using an arcsin square root transformation before regressing empirically obtained power rates onto values of the manipulated parameters.

The Program

The program was written in QuickBASIC (release 4.5) and runs on IBM and IBM-compatible personal computers. It uses the intercept and b weights from Aguinis and Stone-Romero's (1997, Table 4) empirically derived regression equation for predicting statistical power based on values for the manipulated methodological and statistical artifacts:

$$\begin{aligned}\text{Estimated Power} = {} & 1.082 + (.5338 \times RR) + (.0025 \times N) + (.9708 \times p_1) \\ & + (.0317 \times Z_{XZ}) + (1.6527 \times ES) + (1.212 \times RR \times ES) + (.001 \times RR \times N) \\ & - (.102 \times RR \times Z_{XZ}) + (.166 \times RR \times p_1) + (.006 \times N \times ES) + (2.3 \times ES \times p_1) \\ & + (.087 \times ES \times Z_{XZ}) + (.002 \times N \times p_1) + (.178 \times Z_{XZ} \times p_1),\end{aligned} \tag{2}$$

where RR is the range restriction ratio for predictor X (standard deviation in sample / standard deviation in population), N is the total sample size, p_1 is the proportion of cases in moderator-based Subgroup 1 (i.e., n_1 / N), Z_{XZ} is the Fisher's Z transformation of the sample-based correlation between the predictor X and the moderator Z (i.e., r_{XZ}), and ES is the moderating effect size (i.e., the absolute value of the difference between the Fisher's Z transformations of the sample-based correlations between X and Y for Subgroup 1 and Subgroup 2).

Before implementing the equation, the program centers the user-provided values because the Aguinis and Stone-Romero (1997) equation is based on centered predictors. Then, after the estimated power is computed, the program implements an inverse of the arcsin square root transformation to the obtained power estimate. This transformation is needed because it is the inverse of the transformation used by Aguinis and Stone-Romero to linearize the power function (Winer, Brown, & Michels, 1991, p. 356, case ii). Thus, the obtained estimated power displayed by the program is expressed in the typical proportion metric.

Finally, the program also computes a 95% confidence interval (CI) for the statistical power point estimate. The 95% CI is computed by first obtaining the standard error of estimate $S_{Y'}$:

$$S_{Y'} = S_Y \times \sqrt{1 - R^2}, \qquad (3)$$

where S_Y is the standard deviation of the power values obtained in the Aguinis and Stone-Romero (1997) simulation expressed in arcsin square root metric (i.e., .817), and R^2 is the squared multiple correlation coefficient for the model reported by Aguinis and Stone Romero (i.e., .948, see Table 4). The lower and upper limit values for the 95% CI around the point estimate are obtained using Equation 4:

$$\text{Point Estimate} \pm S_{Y'} \times 1.96. \qquad (4)$$

As in the case of the point estimate, the lower and upper limits of the CI are converted using the inverse of the arcsin square root transformation. Thus, the program reports both the point estimate and its 95% CI in the typical proportion metric.

Input

The user is prompted interactively to input (a) the sample-based correlation coefficient between X and Y for moderator-based Subgroup 1 ($r_{XY(1)}$), (b) the sample-based correlation coefficient between X and Y for moderator-based Subgroup 2 ($r_{XY(2)}$), (c) the sample-based correlation coefficient between the predictor X and the hypothesized moderator Z (r_{XZ}), (d) the sample size for Subgroup 1 (i.e., n_1), (e) the sample size for Subgroup 2 (i.e., n_2), and

(f) the sample and population standard deviations for predictor X (that are used to compute RR). After entering these values, the program displays the estimated power of MMR to detect a dichotomous moderator variable for the specified conditions based on a nominal Type I error rate of .05.

Note that Equation 2 uses the value for p_1 and not the actual sample sizes in the subgroups (i.e., n_1 and n_2). Also, note that Aguinis and Stone-Romero's (1997) simulation used values of .1, .3, and .5 for p_1. Accordingly, when entering values into the present program, the user should always treat the subgroup with the smaller n as Subgroup 1. In cases in which $n_1 = n_2$, the user can treat either subgroup as Subgroup 1.

Most of the aforementioned information needed by the program is easily obtainable given that a researcher usually has access to the raw data. First, information regarding N, n_1, and n_2 is readily available. Second, $r_{XY(1)}$, $r_{XY(2)}$, and r_{XZ} are also easily obtainable using statistical software packages. Third, the only potentially missing piece of information is predictor X range restriction. Degree of predictor range restriction is typically available in personnel selection research because variability information is known for the entire pool of applicants and the subset of applicants who are selected (Aguinis & Whitehead, 1997).

In situations in which the standard deviation for population X scores is not readily available and, moreover, is presumed to differ from the standard deviation in the sample, we recommend that MMR users obtain two power estimates. The first estimate is computed assuming severe range restriction. Using this severe RR figure, MMR users would obtain a *conservative* power estimate. The range restriction figure can be obtained from previous studies using similar populations or can be based on an educated (and conservative, i.e., severe RR) guess. The second estimate is computed assuming no range restriction (i.e., $RR = 1.00$, when the sample and population standard deviations are equal) or some other RR figure considered liberal (i.e., moderate RR). Using this second RR figure leads to a *liberal* power estimate. This procedure provides MMR users with two power estimates, one (conservative and lower) based on a more severe RR figure and the second (liberal and higher) based on a more moderate RR figure. Thus, the resulting two values represent the lower and upper limits of an interval that includes the estimated power of MMR.

Output

Table 1 displays program-generated power values for situations considered typical in research studies aimed at estimating the moderating effects of dichotomous moderator variables in education, psychology, and management. For example, the median total sample size in a meta-analysis that included 138 validation studies was 103 (Russell et al., 1994). Thus, Table 1 shows power values for situations in which $N = 100$.

Table 1
Power Estimates to Detect the Moderating Effect of a Dichotomous Variable as a Function of Moderating Effect Magnitude (difference in subgroup correlation coefficients), Multicollinearity, Subgroup Sample Sizes, and Predictor Range Restriction

					Power Estimates		
$r_{XY(1)}$	$r_{XY(2)}$	r_{XZ}	n_1	n_2	No *RR* (1.0)	Moderate *RR* (.6)	Severe *RR* (.2)
.2	.4	.1	50	50	.17	.13	.09
.2	.4	.2	50	50	.17	.13	.09
.2	.6	.1	50	50	.42	.30	.19
.2	.6	.2	50	50	.43	.30	.19
.2	.8	.1	50	50	.82	.62	.39
.2	.8	.2	50	50	.82	.62	.40
.2	.4	.1	40	60	.16	.12	.08
.2	.4	.2	40	60	.16	.12	.08
.2	.6	.1	40	60	.38	.26	.16
.2	.6	.2	40	60	.38	.26	.16
.2	.8	.1	40	60	.74	.52	.30
.2	.8	.2	40	60	.74	.53	.31
.2	.4	.1	20	80	.13	.10	.07
.2	.4	.2	20	80	.13	.10	.07
.2	.6	.1	20	80	.28	.18	.10
.2	.6	.2	20	80	.28	.18	.10
.2	.8	.1	20	80	.55	.34	.15
.2	.8	.2	20	80	.56	.34	.16

Note. $r_{XY(1)}$ = correlation between predictor *X* and criterion *Y* for moderator-based subgroup 1 (i.e., *Z* = 1); $r_{XY(2)}$ = correlation between predictor *X* and criterion *Y* for moderator-based subgroup 2 (i.e., *Z* = 2); r_{XZ} = correlation between predictor *X* and moderator *Z* (i.e., multicollinearity); n_1 = sample size of moderator-based subgroup 1; n_2 = sample size of moderator-based subgroup 2; *RR* = range restriction on predictor *X*.

As shown in Table 1, the statistical power of hypothesis tests of moderating effects using MMR is usually inadequate. For instance, unequal sample sizes decrease the power of MMR quite notably. Note, however, that this finding regarding the detrimental effects of unequal subgroup sample sizes on statistical power is not new. The presence of unequal subgroup sample sizes attenuates observed relationships not only in the context of MMR (Stone-Romero et al., 1994) but also when computing other statistics such as point-biserial correlations (Kemery, Dunlap, & Griffeth, 1988).

A perusal of Table 1 indicates that power levels rarely reach the recommended .80 value. Moreover, for most situations presented in Table 1, the power level is .50 or lower, which suggests that the rejection of false null hypotheses regarding the effects of dichotomous moderator variables can be predicted as accurately or more accurately by a flip of a coin. Given these results, it is not surprising that social scientists express concerns that moderators are elusive (Zedeck, 1971) and that further research should be conducted to improve methods to detect hypothesized moderators (Cronbach, 1987).

Finally, as a check of the accuracy of the program, we compared values generated by the program with those reported by Aguinis and Stone-Romero (1997, Tables 2-3). Results did not indicate any unexpected discrepancies.

Program Availability

The executable (MMRPWR.EXE) and source code (MMRPWR.BAS) versions of the program are available at no cost on a 3.5-inch diskette. Users who wish to obtain the program should send a blank formatted diskette and a self-addressed, stamped envelope to Herman Aguinis, Ph.D., College of Business and Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217-3364, USA. Alternatively, the program can be sent electronically as an attachment file by e-mailing a request to haguinis@castle.cudenver.edu or to capierce@montana.edu.

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management, 21*, 1141-1158.

Aguinis, H., Bommer, W. H., & Pierce, C. A. (1996). Improving the estimation of moderating effects by using computer-administered questionnaires. *Educational and Psychological Measurement, 56*, 1043-1047.

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods 1*, 296-314.

Aguinis, H., & Pierce, C. A. (in press). Testing moderator variable hypotheses meta-analytically. *Journal of Management.*

Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement, 54*, 690-692.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192-206.

Aguinis, H., & Whitehead, R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology, 82*, 528-538.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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

Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102*, 414-417.

Hattrup, K., & Schmitt, N. (1990). Prediction of trades apprentices' performance on job sample criteria. *Personnel Psychology, 43*, 453-466.

Kemery, E. R., Dunlap, W. P., & Griffeth, R. W. (1988). Correction for variance restriction in point-biserial correlations. *Journal of Applied Psychology, 73*, 688-691.

MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118*, 405-421.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376-390.
Russell, C. J., Settoon, R. P., McGrath, R. N., Blanton, A. E., Kidwell, R. E., Lohrke, F. T., Scifres, E. L., & Danforth, G. W. (1994). Investigator characteristics as moderators of personnel selection research: A meta-analysis. *Journal of Applied Psychology, 79*, 163-170.
Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of source adjustment in preemployment testing. *American Psychologist, 49*, 929-954.
Sagie, A., & Koslowsky, M. (1993). Detecting moderator variables with meta-analysis: An evaluation and comparison of techniques. *Personnel Psychology, 46*, 629-640.
Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16*, 209-222.
Stone, E. F. (1988). Moderator variables in research: A review and analysis of conceptual and methodological issues. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management* (Vol. 6, pp. 191-229). Greenwich, CT: JAI.
Stone-Romero, E. F., Alliger, G. M., & Aguinis, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects for dichotomous variables. *Journal of Management, 20*, 167-178.
Trattner, M. H., & O'Leary, B. S. (1980). Sample sizes for specified statistical power in testing for differential validity. *Journal of Applied Psychology, 65*, 127-134.
Westermann, R., & Hager, W. (1986). Error probabilities in educational and psychological research. *Journal of Educational Statistics, 11*, 117-146.
Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.
Zedeck, S. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin, 76*, 295-310.

MEG: MEGACLUSTER ANALYTIC STRATEGY FOR MULTISTAGE HIERARCHICAL GROUPING WITH RELOCATIONS AND REPLICATIONS

PAUL A. McDERMOTT
University of Pennsylvania

An analytic and computer strategy is introduced and demonstrated for multistage Euclidean grouping (MEG). The procedure sequentially produces first-stage clusters for independent data blocks; second-stage, higher order clusters based on a full similarity matrix for first-stage clusters; and third-stage clusters that allow case migration to relocate prior misassignments and to optimize within-cluster homogeneity. The process is facilitated by special SAS computer codes and, in addition to conventional SAS cluster output, produces special fusion statistics, plots of all fusion statistics, and indices of homogeneity within clusters and within profile variables. The program also reports replication rates for final clusters.

Cluster analysis has gained immense popularity in recent years. Thousands of published applications are now found in biologic and social science literature (see Blashfield & Aldenderfer, 1988; Borgen & Barnett, 1987; Overall, Gibson, & Novy, 1993). Most applications favor the use of hierarchical agglomeration strategies where N multiple attribute profiles are sorted into distinct clusters as defined by profile shape, level, and dispersion. The resulting clusters are used in turn to discover latent subpopulations, to distinguish normative and pathognomonic trait variation, and to confirm quantitatively the pattern phenotypes suggested by contemporary theory.

In some respects, the popularity of cluster analysis has outpaced the technology. This is particularly manifest as it pertains to methodology that would validate the results of clustering processes. In their comprehensive review of cluster technique, Blashfield and Aldenderfer (1988) emphasized

Correspondence concerning this article should be addressed to Paul A. McDermott, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104.

Educational and Psychological Measurement, Vol. 58 No. 4, August 1998 677-686
© 1998 Sage Publications, Inc.