Multiple Linear Regression Viewpoints

A Publication sponsored by the American Educational
Research Association's Special Interest Group on
Multiple Linear Regression: The General Linear Model

MLRV

Volume 34 • Number 2 • Spring 2008

Table of Contents

Multiple Linear Regression Viewpoints

Editorial Board

Multiple Linear Regression Viewpoints (ISSN 0195-7171) is published by the AERA Special Interest Group on Multiple Linear Regression: General Linear Model through the **University of Alabama at Birmingham**.

Subscription and SIG membership information can be obtained from:
Cynthia Campbell, MLR:GLM/SIG Executive Secretary
Department of Educational Technology, Research & Assessment
Northern Illinois University
DeKalb, IL 60115-2854.
ccampbell@niu.edu

MLRV abstracts appear in CIJE, the ERIC system, and microform copies are available from University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106. *MLRV* is listed in the *EBSCO Librarians Handbook.*

Testing Interactions in Classification Problems

Bernadine Beard **Valerie Bryan**
John D. Morris **Mary G. Lieberman**
Florida Atlantic University

The purpose is to demonstrate a procedure for testing the increment to classification accuracy afforded by an interaction term using predictive discriminant analysis (PDA) and logistic regression (LR). The PSAT (Math) and ethnicity were employed to predict students at risk of passing or failing the Florida Comprehensive Assessment Test (FCAT) Math. Results favored PDA over LR for the failure group. For PDA, at an a priori alpha of .05, the moderation of ethnicity was not significant for the failure group, but was for both passing and the total sample, as were hit-rates for total sample and separate groups by Ethnicity.

Researchers employing multiple linear regression frequently use the well-known technique of hypothesis testing through contrasting the predictive variance attributable to full versus restricted models. This method's power, generality, and applicability to a very wide range of questions in science form a theoretical umbrella under which most univariate inferential statistical tests can be viewed.

In multivariate statistics, a test of the so-called additional information hypothesis (AIH) was suggested by C. R. Rao over 50 years ago (see Hand, 1981, p. 149). The context of this hypothesis is that of one-factor multivariate analysis of variance. The research question associated with this hypothesis pertains to an assessment of the difference between the intergroup distance when all response variables are analyzed and the intergroup distance for a subset of response variables. The hypothesis, then, is that the omitted set of variables adds no information (in the sense of intergroup distance) to that yielded by the subset included. The research question considered in this situation is different from that which occurs in a predictive context because the same criteria are not appropriate. For Rao's AIH, the criterion is intergroup difference, whereas, for the prediction problem, the criterion is classification accuracy (see Huberty & Wisenbaker, 1992) – total, or that obtained within each of the separate groups.

The same type of model contrast explanatory increment question can be asked, and seems to be of at least as much potential interest, in classification questions. Specifically, the question arises in studies using predictive discriminant analysis (i.e., classification), logistic regression, as well as other methods of classification. In this case, the criterion for model accuracy is some form of classification accuracy. The test concerns the difference in proportion of correct classifications (hit-rate) between full and restricted models, just as is done using the R2 in multiple regression. The appropriate test statistic is McNemar's (1947) contrast between correlated proportions, and was introduced by Morris and Huberty (1991; 1995) for the purpose of full versus restricted model testing in predictive discriminant analysis for specific planned contrasts using the total group hit-rate as the criterion.

One of the difficulties with the application of the McNemar statistic to full versus restricted model classification questions is that it requires tallying the number of subjects classified correctly and incorrectly and summarizing the results in a fourfold table corresponding to the full and restricted models. To obtain the entries for that table, one needs more than a knowledge of hit-rates for each model; one must count the number of subjects who were correctly and incorrectly classified in both the full and restricted models in turn. Thus, the total and separate-group hit-rates that are available from standard discriminant analysis and logistic regression package programs are not sufficient information to complete the comparison between full and restricted models. For each individual case, one must tally whether the subject was classified correctly or incorrectly jointly for the full and reduced models. Moreover, if one considers cross-validated classification to be the appropriate metric of model accuracy, then these classifications/misclassifications that are to be tallied must be cross-validation estimates. A computer program to accomplish this otherwise difficult task has been made available in the case of discriminant analysis (Morris & Huberty, 1991; 1995) and logistic regression (Lieberman, Morris & Huberty, 2000).

One typical use of a full vs. restricted model test in multiple regression is in the consideration of an interaction, tested in multiple regression by considering the contribution of a standard multiplicative term (often with variables centered to avoid collinearity problems). Our purpose in this paper is to extend this procedure and illustrate an example thereof in a classification model. Note that these variable importance tests are based on the increment to classification accuracy and are quite different than tests of an interaction term β available in standard computer packages. Those tests are appropriate when considering partial influence on group separation, but not on hit-rate – the criterion of importance in a classification analysis. As this is simply an extension of the full vs. restricted model testing paradigm to the interaction question, the same aforementioned FORTRAN computer program is applicable.

One may argue, however, that because of the positive bias of estimation of hit-rate classification of the calibration sample, a cross-validated estimate of accuracy should be used. A nonparametric approach to estimating cross-validated hit-rate, which has a wide following in the discriminant analysis literature, is the leave-one-out procedure (Huberty & Olejnik, 2006; Huberty & Mourad, 1980; Lachenbruch & Mickey, 1968; Mosteller & Tukey, 1968). In this method, a subject is classified by applying the rule derived from all subjects except the one being classified. This process is repeated round-robin for each subject, with a count of the overall classification accuracy used to estimate the cross-validated accuracy. We show how interaction tests using full versus restricted model testing, parallel to that used in multiple regression, can be extended to classification studies. We illustrate the interaction test for total as well as separate-group Leave-One-Out classification accuracies for both predictive discriminant analysis (PDA) and Logistic Regression (LR).

Method

An example provided herein regards predicting a high-stakes state mandated, "pass or fail," test from a prior low-stakes test hoped to be diagnostic thus aiding in remediation. Specifically the Florida "FCAT" is predicted from the PSAT. Specifically, the question examined is whether the accuracy with which the PSAT Math can classify subjects correctly in regard to passing or failing the FCAT Math is moderated by ethnicity. In this case both the PDA and LR model were created using PSAT (centered on its mean), Ethnicity, and their product predicting FCAT success. The contrast of interest was that between the hit-rate for all three variables and that afforded by excluding the cross-product term, thus testing the increment to classification accuracy afforded by the moderator variable.

Results and Discussion

Table 1 illustrates the cross-validated (leave-one-out) hit-rates from a linear discriminant function with equal priors for the total sample (N=533), as well as by ethnicity. First, for both ethnicities prediction is most accurate for those who fail, which, if there is to be a separate group accuracy difference is in the desirable direction. Also, hit-rate is more accurate for the total sample as well as for separate groups for Ethnicity 1 than Ethnicity 2. The question of interest is whether hit-rate is significantly moderated by ethnicity.

Table 1 shows the results. Although not of primary interest in this study, the difference in cross-validated hit-rate for LR than PDA is of interest. Because of the poor predictive performance of the LR model (using both ethnicities) for the "Failed" group, a better choice would be the PDA model. As primary interest was in correct classification of the subjects who fail the high stakes FCAT, only the interaction tests for the PDA model will be discussed herein.

For the PDA results, if the researcher posited an alpha of .05, the aforementioned moderation tests would lead the researcher to conclude that the moderation of ethnicity on hit-rate was not significant for the FCAT failure group ($p > .05$), but was for both the Passing group (McNemar $z = 3.32$, $p < .01$) and Total sample (McNemar $z = 2.89$, $p < .05$). Thus one could assert that the model was significantly more accurate at predicting group membership for subjects who passed, and for the combination of subjects passing and failing, for the Majority ethnicity than for the corresponding Minority. As well, as the hit-rates are simple ratio statistics, their simple difference (e.g. 82.4% - .66.7%= 15.7%, and 83.3%-69.9= 13.4%) or a proportional increment [e.g. (82.4%/66.7% -1) = 23.5%, and (83.3%/69.9% -1) = 19.2%], depending on notions of purpose and whether group size is a good estimate of the corresponding parameter, might serve as effect size estimators.

Of course, as in multiple regression, the procedure is not limited to categorical moderators. The relevant computer program is available from the authors.

Table 1. PDA and (LR) hit-rate predicting FCAT from PSAT for all Subjects and by two Ethnicities

	Hit-Rate for:					
	Failed FCAT		Passed FCAT		Total (of row)	
All Subjects	88.9%	(44.4%)	79.4%	(97.1%)	80.9%	(89.1%)
Majority	91.9%	(56.8%)	82.4%	(97.5%)	83.3%	(93.3%)
Minority	79.5%	(47.7%)	66.7%	(93.0%)	69.9%	(81.5%)

References

Allen, D. A. (1971). *The prediction sum of squares as a criterion for selecting predictor variables* (Tech. Rep. No. 23). Lexington: University of Kentucky, Department of Statistics.

Hand, D. J. (1981). *Discrimination and classification*. New York: Wiley.

Huberty, C. J & Mourad, S. A. (1980). Estimation in multiple correlation/prediction. *Educational and Psychological Measurement*, *40*, 101-112.

Huberty, C. J and Olejnik, S. (2006). *Applied MANOVA and discriminant analysis*. New York: Wiley.

Huberty, C. J & Wisenbaker, J. M. (1992). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 2, pp. 169-208). Greenwich, CT: JAI Press.

Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, 10, 1-11.

Lieberman, M. G., Morris, J. D., & Huberty, C. J (2000, April). *Full vs. restricted model testing in logistic regression*. Paper presented at the meeting of the American Educational Research Association, New Orleans.

Looney, S. W. (1988). A statistical technique for comparing the accuracies of several classifiers. *Pattern Recognition Letters*, *8*, 5-9.

Morris, J. D., & Huberty, C. J (1991, April). *Full vs. restricted model testing in discriminant analysis*. Paper presented at the meeting of the American Educational Research Association, Chicago.

Morris, J. D., & Huberty, C. J (1995). Full versus restricted model testing in predictive discriminant analysis. *Journal of Experimental Education*, *63*, 161-165.

McNemar, Q. (1947). Note on the sampling error of the differences between correlated proportions or percentages. *Psychometrika*, *12*, 153-157.

Mosteller, F. & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.). *Handbook of social psychology* (Vol. 2, pp. 80-203). Reading, MA: Addison-Wesley.

Send correspondence to: Mary G. Lieberman
Florida Atlantic University
Email: mlieberm@fau.edu

A Method for Choosing Weights to Predict College Grades for Admission Decisions and to Assess their Fairness by Race/Ethnicity

Steven Andrew Culpepper University of Colorado Denver
Ernest C. Davenport University of Minnesota
Mark L. Davison University of Minnesota

Previous research suggests that equal weights tend to outperform statistically optimal weights in cross-validation studies. This paper argues that the findings from the equal weights literature are relevant for researchers that predict college grades and/or assess differential prediction of college grades by student characteristics. An application of the criterion profile methodology (CPM) is presented to demonstrate how to examine individual criterion profiles. This study showed how to use the CPM to determine the extent to which equal and statistically optimal coefficients differentially predicted college grades for minority and majority students. The results support previous findings, in that, 92.5% of the explained variance in college grades was attributed to equal weights, where standardized test scores and high school rank were weighted equally, and 7.5% of the explained variance was accounted for by statistically optimal coefficients that weighted ACT Math scores less than ACT English and high school rank. Additionally, equally weighting admission information was more accurate for predicting Asian Americans' future academic performance than European Americans.

Prediction is an important aspect of scientific endeavors. For instance, educators predict student achievement, psychologists classify clients into different diagnoses, university personnel place students into developmentally appropriate courses, and economists forecast stock prices and economic conditions. Multiple regression is an important statistical tool for making these types of predictions.

One classic application of multiple regression is the prediction of college grades with standardized test scores and measures of high school academic success (Hills, 1964; McKelpin, 1965; Munday, 1965; Richards & Lutz, 1968; Sassenrath & Pugh, 1965; Stanley & Porter, 1967). Many applications of multiple regression for predicting college grades are designed to create an equation (which is later referred to as a selection equation) for admission officers to objectively select or sort applicants based upon predicted academic performance. However, practitioners interested in prediction should be cautious about using regression equations to predict applicants' grades. Specifically, using an estimated regression equation to select or sort applicants assumes the intercept and slope coefficients are statistically equal (i.e., invariant) across different demographic groups (Dorans, 2004).

Differential prediction, which occurs when the intercept and/or slope coefficients differ between two or more groups (Cleary, 1968; Humphreys, 1952), can impact admissions decisions that use quantitative information and multiple regression predictions. For instance, intercept differences indicate that one group's academic performance is over-predicted and another group's performance is under-predicted. Previous research suggests that racial/ethnic minority group performance is consistently over-predicted with a common regression equation (Breland, 1979; Burton & Ramist, 2001; Cleary, Humphreys, Kendrick, & Wesman, 1975; Duran, 1983; Linn, 1973; Stanley & Porter, 1967; Wilson, 1983; Young, 2001). In this case, admission decisions are biased in favor of minorities; using one equation to predict academic performance will tend to over-predict minorities' actual performance. In contrast, female's performance is consistently under-predicted (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Chou & Huberty, 1990; Elliott & Strenta, 1988; Noble, 2003; Pennock-Román, 1994; Ramist, Lewis, & McCamley-Jenkins, 1993; Young, 1994).

Slope differences provide evidence about the accuracy of different variables for predicting group outcomes. For example, Young's (2001) review of differential prediction studies suggests that minority students tend to have smaller slope coefficients for academic predictors, such as test scores and high school grades, than their European American counterparts. Slope differences by race/ethnicity are often evidence that test scores and/or high school grades are less valuable indicators of minority students' future academic performance and should be considered differently and perhaps with less value in admissions decisions.

Researchers and practitioners need to assess the presence of differential prediction to ensure responsible multiple regression-based predictions. While multiple regression is used to assess differential prediction (the application of regression to assessing differential prediction is also referred to as moderated multiple regression, or MMR, see for example Saunders, 1956), research suggests that it is not a perfect tool for uncovering differential prediction. Aguinis and Stone-Romero (1997) note that the power to detect slope coefficient differences is smaller when certain artifacts are present (e.g., small sample sizes, relatively low representation of minorities in the sample, measurement error in the predictor and criterion, and range restriction in the predictor). That is, the likelihood of uncovering true differences in slope coefficients is less likely when the aforementioned artifacts are present. This translates into researchers being less likely to conclude that test scores and/or high school academic achievement are less valuable indicators for a given minority group (e.g., comparisons by race/ethnicity, gender, first-generation status, etc.) even when such an inference is true in the population.

Dana and Dawes (2004) offer an additional caution for statisticians and practitioners who make predictions with multiple regression. Through the use of statistical simulations, Dana and Dawes (2004) contributed evidence to an existing body of research (Davis & Sauser, 1991; Dawes, 1979; Dawes & Corrigan, 1974; Dawes, Faust, & Meehl, 1989) that equal weights (i.e., slope coefficients that are equal for every predictor) tend to outperform statistically optimal weights (i.e., standardized slope coefficients derived with ordinary least squares, or OLS) in cross-validation samples. Dana and Dawes (2004) boldly conclude that multiple regression should not be used for prediction purposes when the total variance explained is small (i.e., R^2 is less than 0.25). Similarity, Einhorn and Hogarth (1975) suggest that regression should not be used when $R^2 < 0.50$.

High school grades and standardized test scores tend to account for less than 50% of the total variation in college grades (and sometimes less than 25% of variation). This poses a challenge for researchers who wish to predict college grades and/or conduct differential prediction studies. In particular, differential prediction studies use MMR to test the extent to which subgroups' differ in statistically optimal slope coefficients. Dana and Dawes' (2004) findings provide evidence that comparing the equivalence of statistically optimal weights may be inadvisable when R2 is small. Indeed, if equal weights account for the majority of variance in college grades rather than statistically optimal weights, it may be more appropriate to determine the extent to which equal weights differentially predict college grades for different subgroups.

The goal of this paper is to show how an external profile analysis technique can be used to assess differential prediction by race, while concurrently examining the value of equal and statistically optimal weights. A simple application of a criterion profile methodology (CPM; Davison & Davenport, 2002) is presented to test the extent to which statistically optimal and equal weights differentially predict the college grades of two minority groups (African; Asian Americans) when compared to European Americans.

One goal of this study was to assess differential prediction in a way that accommodates Dana and Dawes' (2004) concerns. Accordingly, the first section of this paper introduces the mathematical formulation of the CPM and describes how the CPM can address Dana and Dawes' (2004) concerns.

The CPM is an external profile analysis technique (Davison & Davenport, 2002) that yields a predictor profile that differentiates between subjects with high and low scores on a criterion, such as college grades. So, another goal of this study was to introduce researchers to the CPM, since it is applicable to other research endeavors. The profile analysis feature of the CPM was described to clearly articulate the model. In this study, high school grades and ACT test scores were used to demonstrate individual and sub-group profile differences. The third section discusses the results of the differential prediction analysis as it relates to using the CPM and the last section provides concluding remarks.

Description of the Criterion Profile Methodology

Profiles patterns identified with cluster analysis or multidimensional scaling have been criticized for not exhibiting criterion-related validity evidence (Watkins, 2000). The advantage of the CPM is that the identified profiles are explicitly related to a criterion, such as college grades in this study, and exhibit some degree of validity as determined by the strength of the relationship between the predictors and the criterion (Davison & Davenport, 2002). That is, the identified predictor patterns distinguish those low and high on the criterion variable.

The CPM parses the variation in a criterion variable explained by a set of independent variables into two components: a level effect, which is characterized by an equally weighted linear composite of the predictors, and a pattern effect, which is the covariance between a subject's predictor profile and the regression coefficients. The relationship between the level effect and equal weights versus the pattern effect and optimal weights is discussed below.

To formally define these two effects we start with the usual regression model in Equation 1 below:

$$Y_p = \sum_{v=1}^{V} \beta_v X_{pv} + a + e_p \tag{1}$$

where Y_p represents the criterion score for person p, β_v represents the regression coefficient for variable v ($v = 1$ to V where V is the number of predictors and $V \geq 2$), and X_{pv} is the score for subject p on predictor v. Finally, a is the intercept of the regression equation and e_p is a random error term. The criterion profile is defined as the set of slope coefficients, $\beta_1, \beta_2, \dots$, and β_v, for the p predictors.

Davison and Davenport (2002) prove that the regression model in equation 1 is equivalent to the following model:

$$Y_p = \gamma_1 Cov_p + \gamma_2 \overline{X}_p + a + e_p \tag{2}$$

where, $\overline{X}_p$ is referred to as level, Cov_p is referred to as pattern, and γ_1 and γ_2 are their respective slope coefficients (these coefficients are standardized if Cov_p and $\overline{X}_p$ represent z-scores). The first term on the right of Equation 2 constitutes the pattern effect, and the second term is the level effect. Regardless of the original number of predictors, the original regression equation can be reduced to three terms, a, Cov_p, and $\overline{X}_p$. It is important to note that Y_p and e_p are the same in Equations 1 and 2, so that the level and pattern variables together account for the same proportion of variation in the criterion as the original variables.

The equations for level and pattern are presented below in Equations 3 and 4:

$$\overline{X}_p = \frac{\sum_{v=1}^{V} X_{pv}}{V} \quad ; \tag{3}$$

$$Cov_p = \frac{\sum_{v=1}^{V} (\beta_v - \overline{\beta})(X_{pv} - \overline{X}_p)}{V} \quad ; \tag{4}$$

where the only new variable, $\overline{\beta}$, in Equation 4 represents the average of the regression coefficients. Equation 3 shows that level, $\overline{X}_p$, is person p's unweighted average on the independent variables, which suggests that individuals who tend to have larger (smaller) standardized values on the V predictors will also tend to have larger (smaller) values for level.

Pattern (Cov_p) is the covariance between person p's predictor scores and the weights from the original regression, and it is a measure of the match between the observed score profile of person p and the pattern that distinguishes people with high scores on the criterion. Individuals with predictor scores whose pattern matches the configuration of the regression weights will have larger profile match statistics Cov_p and therefore higher predicted values.

Because it is a covariance measure, pattern is positive for subjects whose scores are consistent with the criterion profile (i.e., the configuration of the slope coefficients, β_v) and negative for subjects whose scores are consistent with the mirror image of the criterion profile. The mirror image profile is defined by slope coefficients with the exact opposite configuration as the criterion profile. The coefficients for the mirror image profile (ψ_v) can be found with the following expression: $\psi_v = \overline{\beta} - (\beta_v - \overline{\beta}) = 2\overline{\beta} - \beta_v$. Subjects with predictor profiles corresponding to the mirror image pattern tend to have lower predicted values after controlling for level. The criterion and mirror image profiles are discussed later during the application of the CPM with admissions data.

Value of the CPM for Differential Prediction Studies

The CPM is useful for identifying profiles that differentiate between individuals with high and low criterion values, such as first year cumulative grade point average in our case. However, the criterion profile is only useful for differentiating between high and low CGPA scorers when the pattern effect is statistically significant after controlling for level, which occurs when the statistically optimal weights add predictive value in addition to equal weights. In fact, pattern is generally statistically significant when variability exists among the standardized regression weights. In cases where pattern provides no additional prediction in the criterion over that of level, the equal weight profile differentiates between low and high scoring subjects, subjects with larger criterion scores tend to have high values on all the predictors rather than a configuration of predictor scores.

The pattern variable is also important, since it represents the extent to which practitioners should employ statistically optimal weights in decisions. Referring to Equation 4 again, we see that pattern is particularly important for assessing differential prediction. The relative size of the regression weights vary along two extremes: the weights are either close to being equal or they differ substantially in magnitude. Pattern will account for less variation in a criterion when the regression weights are similar or equal to each other, which would lead to equal weights outperforming statistically optimal weights. Conversely, statistically optimal weights are important for prediction purposes when pattern accounts for relatively more variation in a criterion than level, or equal weights. Pattern is useful to the extent that the optimal weights (betas) vary and this variance in weights accounts for differences in the criterion.

This study assesses the value of using equal vs. statistically optimal weights by estimating the amount of variation that is accounted for by level and pattern. More formally, the hypothesis is:

$$H_0\text{: } R^2_{\text{Level}} = R^2_{\text{Level + Pattern}} \text{ and } H_1\text{: } R^2_{\text{Level}} \neq R^2_{\text{Level + Pattern}};$$

which is testable with the traditional F-test comparing parametric regression models with V-1 and N-V-1 degrees of freedom where V is the number of independent variables and N is the sample size (Davison & Davenport, 2002). Substantively, this test will provide evidence for whether or not statistically optimal weights provide predictive value above and beyond equal weights.

Methods

Sample

This study used data from the entering class of 2000 at a public research university. The data was collected from each student during the pre-college admissions process and provided to the researchers by the Office for Institutional Research. The sample consisted of 2,035 students who enrolled in the College of Liberal Arts (CLA) fall 2000 and persisted through one year of academic study. These 2,035 students were disaggregated by self-reported race/ethnicity. Of the 2,035 students, 68 were African American (AFA), 11 were American Indian, 186 were Asian American (ASA), 1,683 were European American (EA), 38 were Hispanic, 5 were International, and 44 were unidentified. Only the AFA, ASA and EA groups were included in the analyses, since there were small numbers of American Indians, Hispanics, International, and unidentified students in the sample. The final sample size included 1,933 students (four students had missing scores on at least one of the predictors).

Variables

In this application of the CPM, the predictors of interest were students' ACT English sub-score (ACTE), ACT math test sub-score (ACTM), and high school percentile rank (HSR) and the criterion was first-year cumulative grade point average (CGPA). The regression and CPM analyses were conducted by standardizing the predictors and criterion across racial subgroups onto a z-score scale with a mean of zero and variance of one. Table 1 reports descriptive statistics of the variables. Certainly, one could argue that ACTE, ACTM, and HSR predict first-year grades differently depending upon the type of coursework or degree program in which a student engages. It is important to note that the students in the CLA were chosen to reduce the potential heterogeneity in regression equations across different colleges within the university.

Point of Caution

With respect to the CPM, there are no limitations of formulating the general linear model in terms of level and pattern. In fact, the reconfiguration of the general linear model into the CPM accounts for the same proportion of variance in the criterion. However, it is important to consider

Table 1. Multiple Regression and Criterion Profile Methodology Summary.

	Std. Weight	Sig.	Part Corr.	*M*	*SD*
Regression Model[a]					
ACT English	0.306	***	0.261	24.3	4.30
ACT Math	0.087	***	0.074	24.6	4.08
High School Rank	0.306	***	0.302	80.8	11.50
CPM Model[a]					
Level	0.496	***	0.493	-0.0004	0.71
Pattern	0.138	***	0.137	-0.00003	0.05

Cross-Validation Summary	Sample 1	Sample 2
R^2, Level Only	0.233	0.229
R^2, Level + Pattern	0.259	0.242

Note: Std Weight = Standardized slope coefficient, Sig. = Level of significance, Part Corr. = Part Correlation, *M* = average, *SD* = standard deviation.
The average CGPA was 3.01 with a standard deviation equal to 0.59.
[a]The model $R^2 = 0.250$ for both the regression and CPM models.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

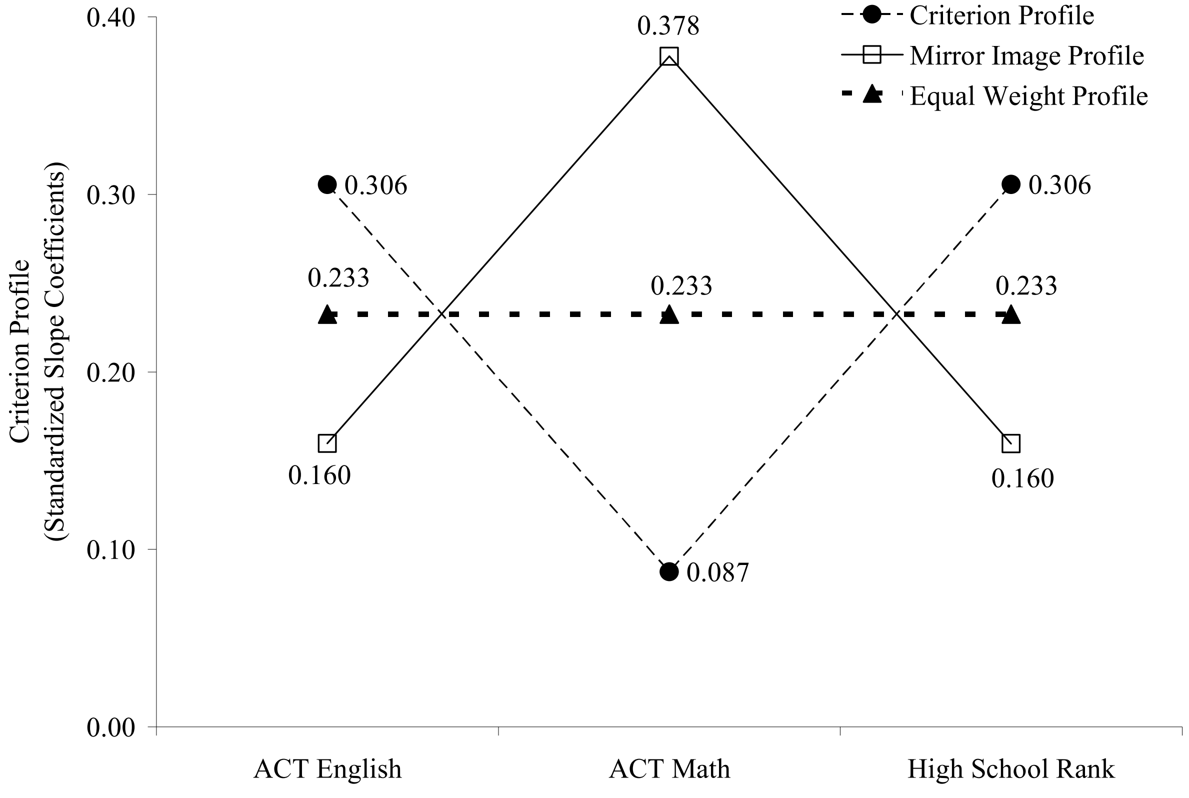


Figure 1. Pooled, within race/ethnicity, and equal weight criterion profiles.
Note. $R^2 = 0.250$ for the criterion profile and $R^2 = 0.231$ for the equal weight profile.

one issue to ensure meaningful CPM analyses, which is that the independent variables need to be on either a substantively meaningful scale, such as the number of credit hours in various mathematics courses (Davison & Davenport, 2002), or the same scale, such as z-scores, to yield regression coefficients that are comparable in the criterion profile. Failure to address the scaling of the predictors may produce misleading or substantively uninformative results (Davison & Davenport, 2002).

Results

The results section consists of two subsections. In an effort to further articulate and demonstrate the CPM, the first section applies the CPM to identify a profile pattern that distinguishes those high and low on the criterion variable. The second section uses the CPM to assess the extent to which equal and statistically optimal weights differentially predict CGPA.

Profile Application of the Criterion Profile Methodology: Understanding Individual Differences.

Table 1 presents the regression summary of ACTE, ACTM, and HSR as predictors of CGPA. The regression model accounted for approximately 25.0% of the variation in CGPA. Furthermore, all the predictors were positively related to CGPA and were statistically significant at the 0.001 level. The standardized slope coefficients equaled 0.306, 0.087, and 0.306, for ACTE, ACTM, and HSR, respectively. The standardized slope coefficients define the criterion profile, which was characterized by larger weights for ACTE and HSR than for ACTM. The mirror image profile consisted of weights with an exact opposite configuration of the criterion profile. Figure 1 plots the criterion and mirror image profiles, in addition to the equal weight profile.

Figure 2 plots three subjects' standardized predictor profiles to demonstrate how the CPM can be used to describe individual differences. Figure 2 shows that subject 261 more closely matched the criterion profile and subject 144 matched the mirror image profile. The average pattern was approximately zero with a corresponding standard deviation of 0.05. Subject 144 had a pattern value about two standard deviations below the mean ($Cov_{144} = -0.13$) and subject 261 had a pattern value about two standard deviations above the mean ($Cov_{261} = 0.12$). Subject 122's standardized predictor profile resembled the equal weight profile, since the profile was nearly flat and the values for the three predictors were within one standard deviation of each other. Additionally, subject 122's profile did not match either the criterion or mirror image profiles, as indicated by $Cov_{122} = -0.001$. Moreover, subject 122 had the largest level value and subject 261 had the smallest value for level.

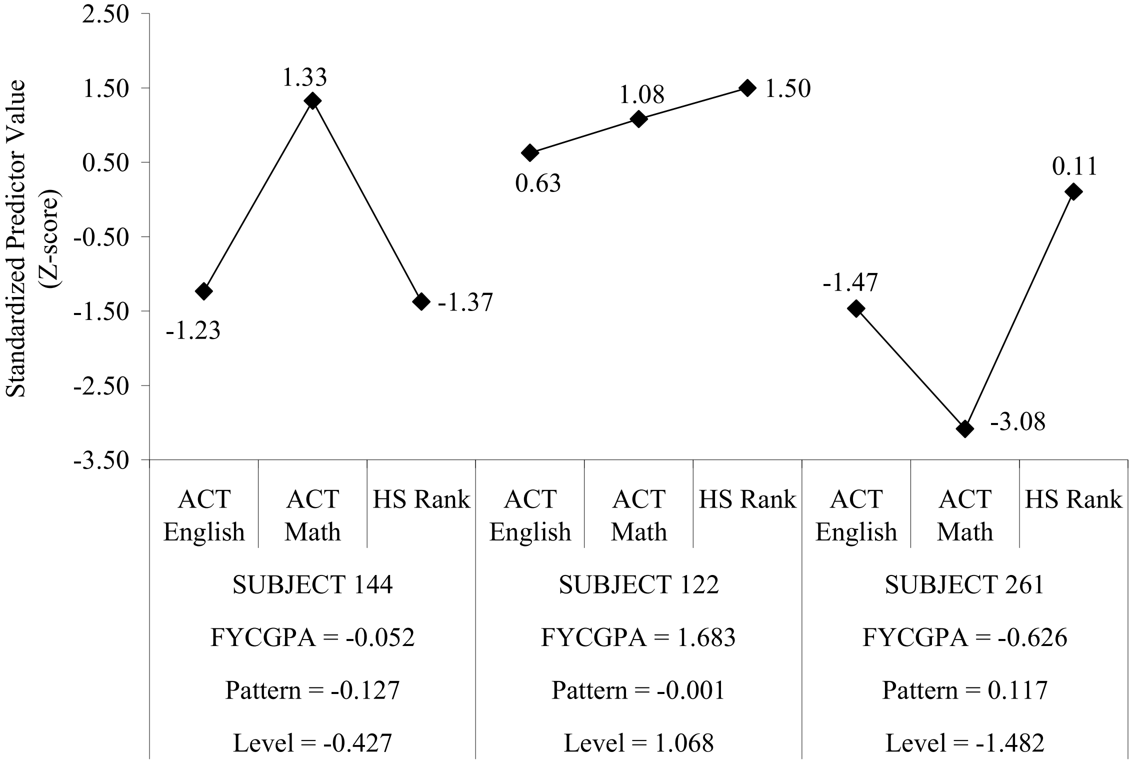


Figure 2. Three subjects' predictor profiles.

Figure 2 also includes information pertaining to the subjects' CGPA, which were standardized to z-scores. Of these three subjects, 122 performed the best academically (1.68) followed by Subject 144 (-0.05), and Subject 261 (-0.63). It is important to determine the extent to which higher academic performance was associated with individual differences in level, individual differences in pattern, or

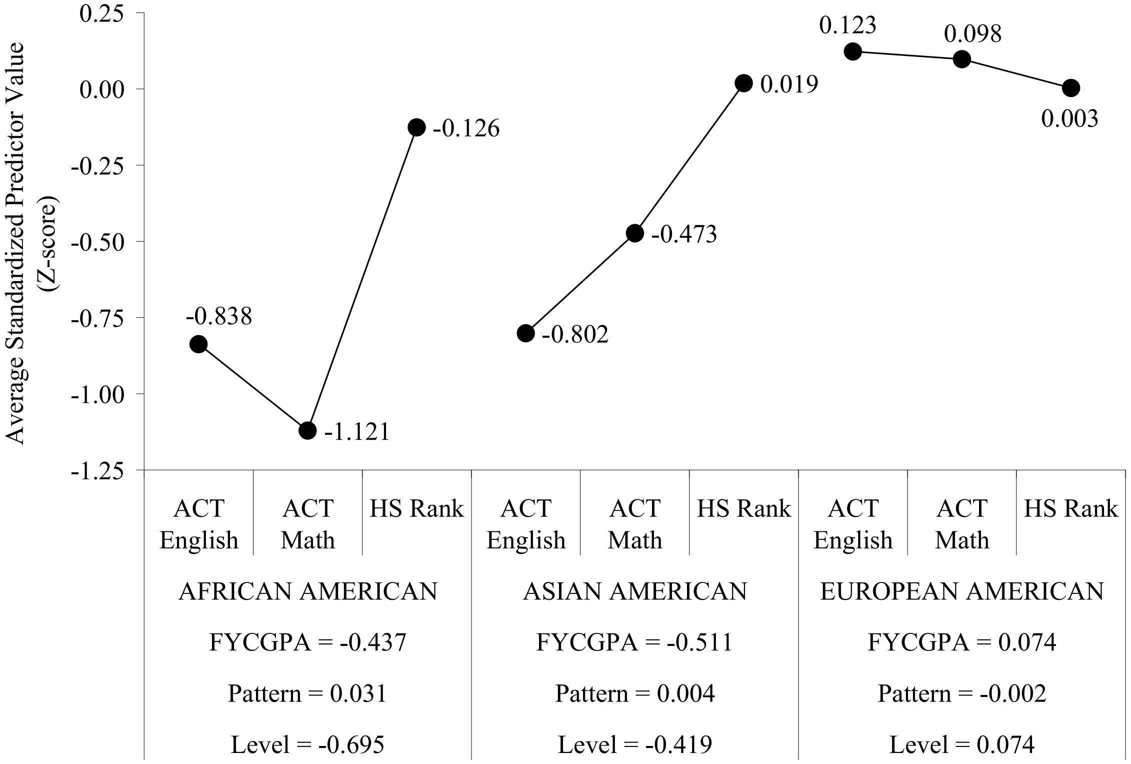


Figure 3. Average predictor profiles by race.

individual differences in both level and pattern. To answer this question we need to understand the amount of variation that was captured by level and pattern independently (note that the bivariate correlation between level and pattern equaled -0.11, so the effects were nearly orthogonal). Table 1 presents the standardized slope coefficients (later denoted as β), p-values, and part-correlations for level ($\beta = 0.496$; $p < 0.001$) and pattern ($\beta = 0.138$; $p < 0.001$). A significant F-test ($F(2, 1{,}929) = 23.6$; $p < 0.001$) provided evidence that pattern accounted for variation in college grades after controlling for level.

The F-test results suggest that statistically optimal weights provided some predictive value that was not captured by equal weights. Still, the results suggest that the vast majority of variance accounted for in CGPA was attributed to level, or an equal weighting scheme. Squaring the part correlations in Table 1 yields the change in R^2 effect-size. Level alone accounted for 23.1% of the total variance in CGPA or 92.5% of the explained variance in CGPA (e.g., 0.231/0.250 = 0.925). Pattern accounted for 1.9% additional variance in CGPA or 7.5% of the explained variance in CGPA. This evidence suggests that individual differences in CGPA were more associated with differences in level than with pattern. From a prediction perspective, equal weights captured most of the variance in CGPA when compared to statistically optimal weights.

Davison and Davenport (2002) note that it is also important to cross-validate CPM findings to assess the value of level and pattern in a different sample. The total sample was randomly divided into two groups. Standardized test scores and high school rank were regressed onto CGPA and the resulting standardized slope coefficients were used to create pattern variables in the omitted sample. The bottom portion of Table 1 summarizes the cross-validation findings. In particular, level accounted for 23.3% and 22.9% of the total variance in CGPA within each sample. Pattern accounted for an additional 2.6% and 1.3% after controlling for level. The cross-validation results provide additional evidence that level accounted for the vast majority of variance in CGPA. Note that this cross-validation is especially important given sample fluctuations of regression weights and the fact that practitioners may use them as if they are stable.

This finding suggests that the differences among CGPA scores for the three subjects in Figure 2 were largely due to individual differences in level rather than pattern; i.e., subjects with higher academic performance in college tended to have higher scores on all of the predictors rather than a configuration of predictor values that matched the criterion profile. For instance, Figure 2 shows that the subject who had the highest values on all of the predictors (Subject 122) also had the largest CGPA. Conversely, the two subjects that exhibited variability in their standardized predictor values had lower level values and lower academic performance.

Differential Prediction Application of the Criterion Profile Methodology

The previous section presented a profile application of the CPM. This section focuses upon the use of the CPM as a means of comparing the predictive value of statistically optimal weights and/or equal weights in differential prediction studies. Previous research suggests that an equally weighted linear composite of the independent variables provides predictive power comparable to or better than statistically optimal weights (Dana & Dawes, 2004; Davis & Sauser, 1991; Dawes, 1979; Dawes & Corrigan, 1974; Dawes et al., 1989). The use of the CPM for differential prediction studies offers a way to examine whether equal weights and/or statistically optimal weights are differentially valid for different groups simultaneously.

The CPM statistical results in Table 1 provided evidence that equally weighting the independent variables accounted for nearly all of the variation in CGPA. Thus, the value in using statistically optimal weights after controlling for an equally weighted linear composite was limited. In this instance, where level accounts for the majority of variation in CGPA, it may not be appropriate to assess differential prediction of subgroups by comparing the statistically optimal regression equations. Instead, a more meaningful differential prediction study should independently compare the extent to which level, a composite that equally weights the independent predictors, and pattern differentially predicts CGPA for students of different races/ethnicities.

Given the relative value of equal and statistically optimal weights in this study, it may be statistically appropriate to exclude pattern from the MMR model and only estimate whether level differential predicts college grades. Instead, pattern was included in the differential prediction model to demonstrate how researchers can use the CPM to address situations where level and pattern each account for a significant amount of variation in the criterion.

Table 2 presents CPM results for comparing the equivalence of subgroup regression equations. The estimated CPM-MMR model is shown below in Equation 5:

$$\begin{aligned} FYCGPA_p &= b_0 + b_1(\mathrm{AFA}_p) + b_2(\mathrm{ASA}_p) + b_3\bar{X}_p + b_4 Cov_p + \\ &b_5(\mathrm{AFA}_p)\bar{X}_p + b_6(\mathrm{ASA}_p)\bar{X}_p + b_7(\mathrm{AFA}_p)Cov_p + b_8(\mathrm{ASA}_p)Cov_p + e_p \end{aligned} \quad (5)$$

where the dummy variables: AFA_p and ASA_p, equal one for African American and Asian American students, respectively, and zero otherwise. Additionally, CGPA, level, and pattern were standardized. b_0 represents the average EA CGPA in standard deviation units, since EA was the reference group. b_1 and b_2 represent the average difference in CGPA between AFA and EA and ASA and EA, respectively. In other words, these two parameters represent necessary intercept adjustments to better predict these two minority groups relative to the majority group. Furthermore, the EA coefficient for level was b_3 and b_5 and b_6 denote the amount that the slope coefficient for level differed between AFA and EA and ASA and EA, respectively. Just as b_1 and b_2 represented adjustments, so does b_5 and b_6. b_4 represents the contribution of pattern to the prediction equation for EA. b_7 and b_8 represent the corresponding adjustments to pattern for AFA and ASA, respectively.

Table 2. Criterion Profile Methodology Moderated Multiple Regression Summary

	Slope	Sig.
Intercept	0.031	
African American	-0.040	
Asian American	-0.181	*
Level	0.459	***
African American * Level	0.066	
Asian American * Level	0.182	*
Pattern	0.140	***
African American * Pattern	-0.015	
Asian American * Pattern	0.043	

Note. The model $R^2 = 0.259$.
Level and Pattern were standardized.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The extent to which all adjustments are non-significant indicates the degree to which one regression equation would be equally predictive for the three racial/ethnic groups. Statistically significant b_5 and/or b_6 parameters would indicate group differences in the accuracy of equal weights for predicting college grades. Similarly, statistically significant b_7 and/or b_8 parameters would suggest that the ethnic groups were not equal relative to using optimal weights as represented by the beta coefficients.

The evidence in Table 2 suggests that AFA and EA did not statistically differ in regression equations. None of the adjustments for the AFA group were significant which indicates that the estimates provided by the referent group, EA, were sufficient for AFA. That is, level ($b_5 = 0.066$; $p > 0.05$) and pattern ($b_7 = -0.015$; $p > 0.05$) predicted the same for AFA and EA and there were no intercept differences ($b_1 = -0.040$; $p > 0.05$), which indicated that the AFA and EA groups performed similarly academically after controlling for the three measures of pre-collegiate academic success. Conversely, ASA exhibited a statistically lower intercept than EA ($b_2 = -0.18$; $p < 0.05$) and a larger slope coefficient for level ($b_6 = 0.18$; $p < 0.05$). The former finding suggests that the ASA group tended to earn lower CGPA than the EA group. The slope coefficients in equation 5 represent partial correlations, so the latter finding suggests that equal weights demonstrated more criterion-related validity for ASA students than for their EA counterparts. Moreover, in additional analyses, level accounted for 22.6% and 21.0% of the variance in CGPA for ASA and EA, respectively, which suggests that equal weights may have been slightly more accurate for ASA students. EA and ASA did not differ in the extent to which using statistically optimal weights related to the subsequent quality of the prediction.

Conclusion

This paper used the CPM to: 1) conduct profile analysis to differentiate between individuals and groups who earn high and low college grades; and 2) explore differential predictability in the use of equal and statistically optimal weights. The results provided evidence that equal weights, or level, accounted for more variance in CGPA than statistically optimal weights, pattern. The criterion profile, or statistically optimal weights, provided little additional predictive ability for differentiating between students with high and low CGPA scores. Therefore, the best weighting scheme was one that treated test scores and high school rank equally rather than the statistically optimal weighting scheme that gave more weight to ACT English and high school rank than to ACT Math. Perhaps intuition or research would suggest that ACT Math is not as good of a predictor of academic performance for students in liberal arts, since their coursework may include less mathematics. The findings of this study suggest that using equal weights for all applicants will capture most of the variability in first-year grades; approximately 90% of the accounted for variance. From a practitioner's perspective, the evidence suggests that ACT Math scores should be treated with the same weight or importance as ACT English scores, and high school rank in decisions for admitting applicants to the College of Liberal Arts.

This finding had direct relevance for assessing differential prediction. That is, statistically optimal weights accounted for very little variance in addition to equal weights. Thus, it was more appropriate to test the extent to which equal and statistically optimal weights differentially predicted grades for different racial groups. In fact, an equal weighting scheme was more valid for ASA than for EA. There was no evidence to suggest that AFA and EA equations differed, so the equal and statistically optimal weighting schemes provided similar predictive accuracy.

The use of equal or statistically optimal weights poses another methodological challenge for assessing differential prediction. Future differential prediction studies should determine the value of equal and statistically optimal weights by computing the variance accounted for by level and pattern in college grades in the full sample and in cross-validation samples. Failure to determine the relative value of equal and statistically optimal weights may result in researchers comparing the equivalence of groups in statistically optimal coefficients when equal weights account for the vast majority of variance in a criterion. This study demonstrated that the CPM is an appropriate method for addressing this methodological issue.

References

Breland, H. M. (1979). *Population validity and college entrance measures* (Research Monograph No. 8). New York: College Entrance Examination Board.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test* (Research Report No. 2000-1). New York: College Entrance Examination Board.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT® studies of classes graduating since 1980* (Research Report No. 2001-1). New York: College Entrance Examination Board.

Chou, T., & Huberty, C. J. (1990). *A freshman admissions prediction equation: An evaluation and recommendation* (No. ERIC Document Reproduction Service No. ED 333 081). Athens, Georgia: University of Georgia.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115-124.

Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30*, 15-41.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics, 29*, 317-331.

Davis, K. R., & Sauser, W. I. (1991). Effects of alternative weighting methods in a policy-capturing approach to job evaluation: A review and empirical investigation. *Personnel Psychology, 44*, 85-127.

Davison, M. L., & Davenport, E. C., Jr. (2002). Identifying criterion-related patterns of predictor scores using multiple regression. *Psychological Methods, 7*, 468-484.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.

Dawes, R., M. & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95-106.

Dawes, R., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

Dorans, N. J. (2004). Using subpopulation invariance to assess test score equity. *Journal of Educational Measurement, 41*, 43-68.

Duran, R. P. (1983). *Hispanics' education and background: Predictors of college achievement* (Research Report). New York: College Entrance Examination Board.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171-192.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement, 25*, 333-347.

Hills, J. R. (1964). Prediction of college grades for all public colleges of a state. *Journal of Educational Measurement, 1*, 155-159.

Humphreys, L. (1952). Individual differences. *Annual Review of Psychology, 3*, 131-150.

Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 139-161.

McKelpin, J. P. (1965). Some implications of the intellectual characteristics of freshmen entering a liberal arts college. *Journal of Educational Measurement, 2*, 161-166.

Munday, L. (1965). Predicting college grades in predominately negro colleges. *Journal of Educational Measurement, 2*, 157-159.

Noble, J. (2003). *The effects of using ACT composite score and high school average on college admission decisions for racial/ethnic groups* (ACT Research Report Series No. 1). Iowa City: ACT Program.

Pennock-Román, M. (1994). *College major and gender differences in the prediction of college grades* (College Board Report No. 2). New York: College Entrance Examination Board.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1993). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Report No. 1). New York: College Entrance Examination Board.

Richards, J. M., Jr., & Lutz, S. W. (1968). Predicting student accomplishment in college from the ACT assessment. *Journal of Educational Measurement, 5*, 17-29.

Sassenrath, J. M., & Pugh, R. (1965). Relationships among CEEB Scholastic Aptitude Test and American College test scores and grade point average. *Journal of Educational Measurement, 2*, 199-205.

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16*, 209-222.

Stanley, J. C., & Porter, A. C. (1967). Correlation of scholastic aptitude test scores with college grades for negroes versus whites. *Journal of Educational Measurement, 4*, 199-218.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15*, 465-479.

Wilson, K. M. (1983). *A review of research on the prediction of academic performance after the freshman year* (College Board Research Report No. 83-2 and Educational Testing Service Research Report No. 83-11). New York: College Entrance Examination Board.

Young, J. W. (1994). Differential prediction of college grades by gender and by ethnicity: A replication study. *Educational and Psychological Measurement, 54*, 1022-1029.

Young, J. W. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis* (Research Report No. 2001-6). New York: College Entrance Examination Board.

Send correspondence to: Steven Andrew Culpepper
University of Colorado Denver
Email: steve.culpepper@ucdenver.edu

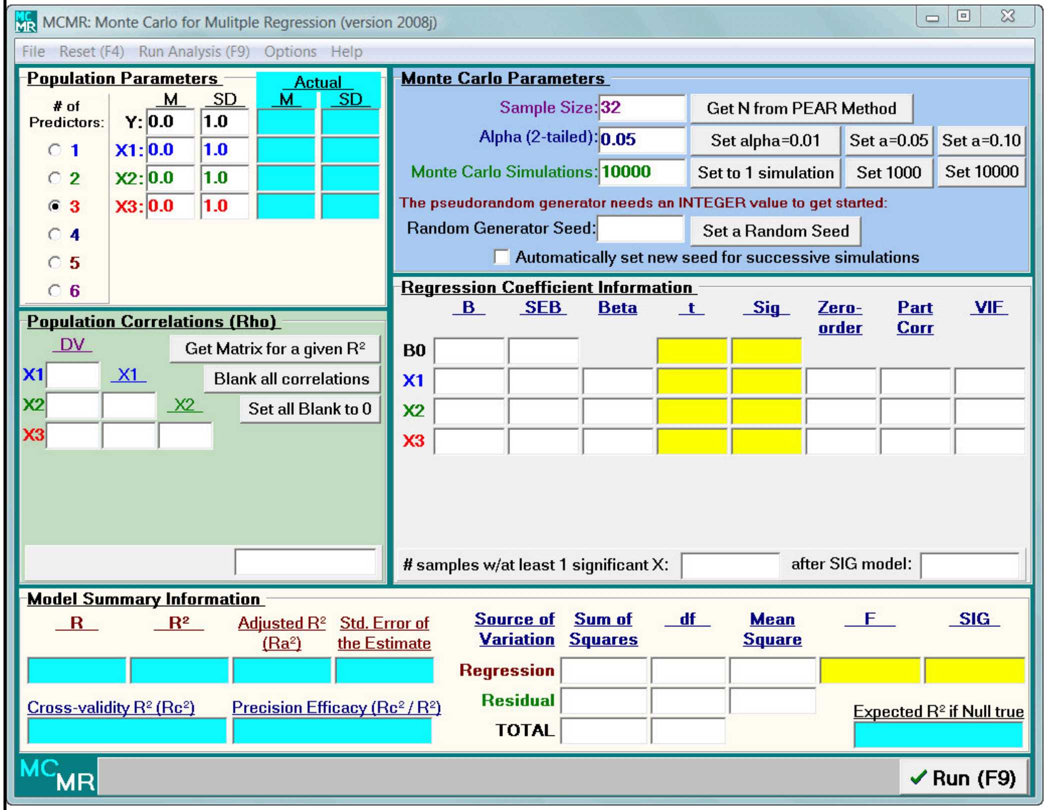
A Monte Carlo Program for Multiple Linear Regression

Gordon P. Brooks
Ohio University

The primary purpose of this presentation is to demonstrate a new computer program that statistics instructors can use to help teach certain regression topics in their courses. In particular, a computer program was written in Borland Delphi 2007 and will run under most recent versions of the Microsoft Windows operating system, including XP and Vista. The program may be downloaded free of charge.

The MCMR: Monte Carlo for Multiple Regression program performs Monte Carlo simulations of ordinary least squares multiple linear regression with up to 6 predictors. The program runs single sample analyses in addition to Monte Carlo simulations. For single samples, data can be saved and imported in comma-delimited text format. For Monte Carlo analyses, sampling distribution data can be saved for several regression statistics for further analyses elsewhere. The on-screen results from any analysis can be saved to a file and printed. The summary results provided from the Monte Carlo simulations include R-squared statistics, shrinkage statistics, regression coefficients, standard errors, and other relevant statistical results. Suggestions for use will be provided to help users understand how the program can be used effectively in intermediate statistics courses.

The MCMR Program is available at: http://oak.cats.ohiou.edu/~brooksg/software.htm



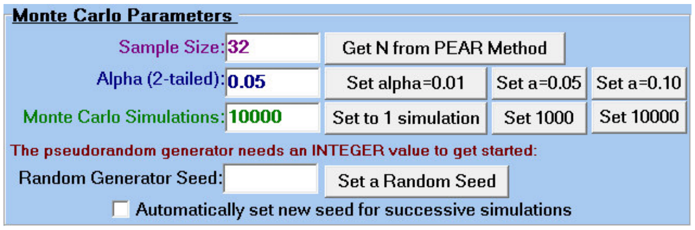
Y: 0.0 1.0
X1: 0.0 1.0
X2: 0.0 1.0
X3: 0.0 1.0

Sample Size: 32
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 10000

This is the Opening Screen that appears when the program is started (or after the "Reset (F4)" menu option is chosen).

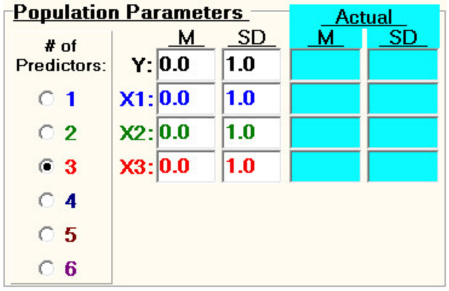
<u>3 sections require user input</u>
This is where we describe the population from which samples will be drawn in the Monte Carlo process. That is, the Monte Carlo process randomly generates samples of data that could come from the particular population described (using means, standard deviations, and correlations).

Click "Run" (bottom right) or press F9 to begin the Monte Carlo analysis.



Sample Size: 32
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 10000

Set sample size, alpha, number of simulations, and maybe a seed for the random number generator (if you use the same seed, you get the same results).

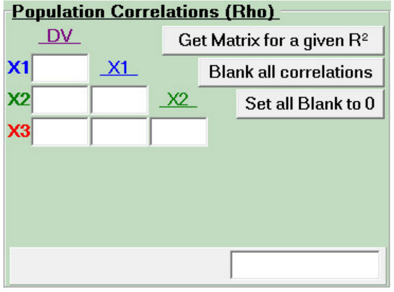


Population Parameters

# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

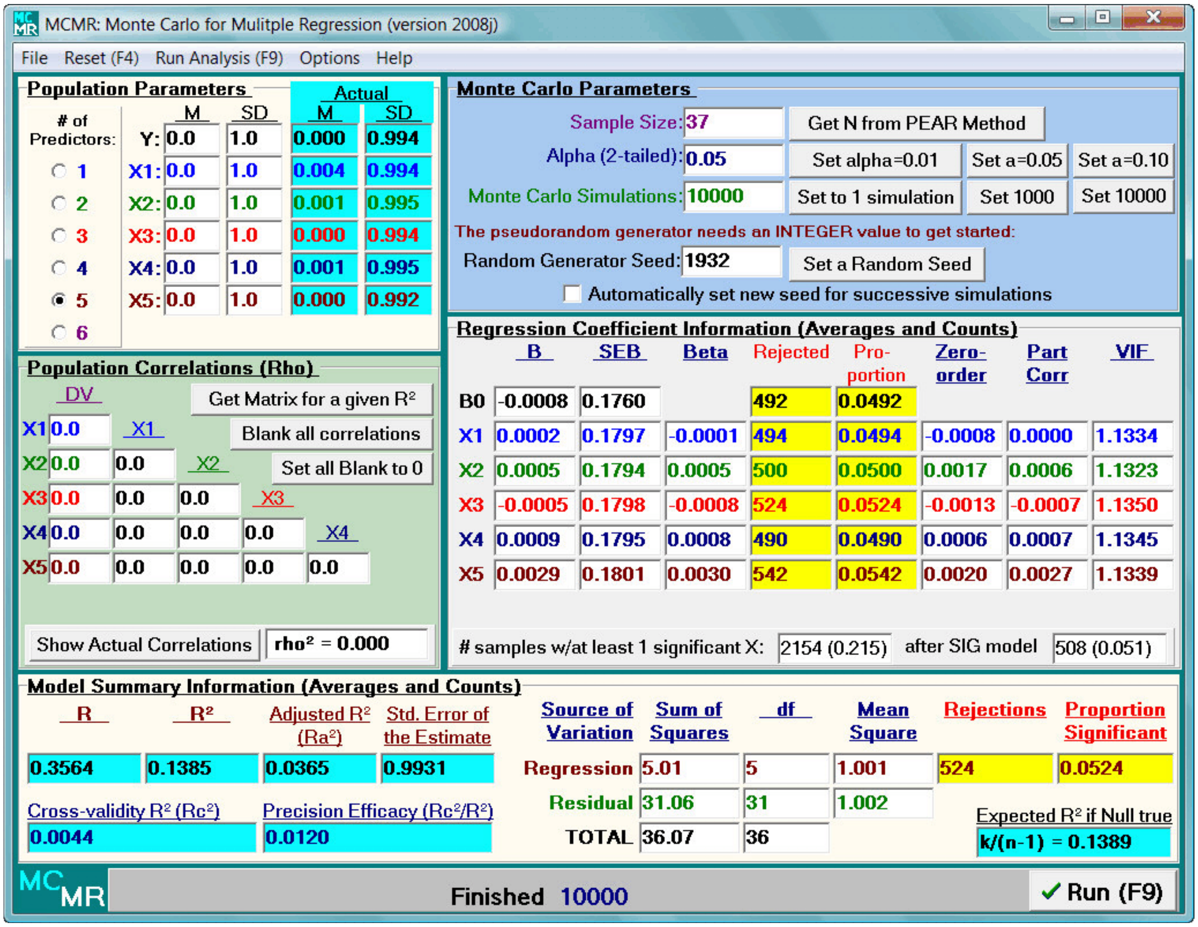
	M	SD	Actual M	Actual SD
Y:	0.0	1.0		
X1:	0.0	1.0		
X2:	0.0	1.0		
X3:	0.0	1.0		

Choose the number of predictors and Set the population means and standard deviations (Y is the dependent variable, X1 is predictor 1, etc.)



Set the population correlations (rho). You can get a random matrix that meets certain criteria (described later). Some matrices will not work as proper CORRELATION MATRICES. If one is entered, and error message will pop up, saying that the matrix is not Positive Definite (see Get Matrix section below).

After an analysis



Population Parameters

# of Predictors: ○ 1 ○ 2 ○ 3 ○ 4 ● 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	0.0	1.0	0.000	0.994
X1:	0.0	1.0	0.004	0.994
X2:	0.0	1.0	0.001	0.995
X3:	0.0	1.0	0.000	0.994
X4:	0.0	1.0	0.001	0.995
X5:	0.0	1.0	0.000	0.992

Population Correlations (Rho)

	DV	X1	X2	X3	X4
X1	0.0				
X2	0.0	0.0			
X3	0.0	0.0	0.0		
X4	0.0	0.0	0.0	0.0	
X5	0.0	0.0	0.0	0.0	0.0

rho² = 0.000

Monte Carlo Parameters

Sample Size: 37

Alpha (2-tailed): 0.05

Monte Carlo Simulations: 10000

The pseudorandom generator needs an INTEGER value to get started:

Random Generator Seed: 1932

☐ Automatically set new seed for successive simulations

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0008	0.1760		492	0.0492			
X1	0.0002	0.1797	-0.0001	494	0.0494	-0.0008	0.0000	1.1334
X2	0.0005	0.1794	0.0005	500	0.0500	0.0017	0.0006	1.1323
X3	-0.0005	0.1798	-0.0008	524	0.0524	-0.0013	-0.0007	1.1350
X4	0.0009	0.1795	0.0008	490	0.0490	0.0006	0.0007	1.1345
X5	0.0029	0.1801	0.0030	542	0.0542	0.0020	0.0027	1.1339

samples w/at least 1 significant X: 2154 (0.215) after SIG model 508 (0.051)

Model Summary Information (Averages and Counts)

R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.3564	0.1385	0.0365	0.9931

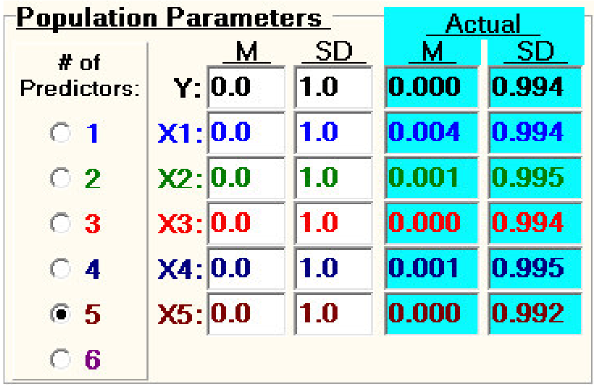
Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2/R^2)
0.0044	0.0120

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	5.01	5	1.001	524	0.0524
Residual	31.06	31	1.002		
TOTAL	36.07	36			

Expected R^2 if Null true $k/(n-1) = 0.1389$

Finished 10000

4 boxes contain results after an analysis, but not all are immediately obvious. Each section is described in greater detail below. This analysis was done with a seed of 1932. All population correlations were 0.0.

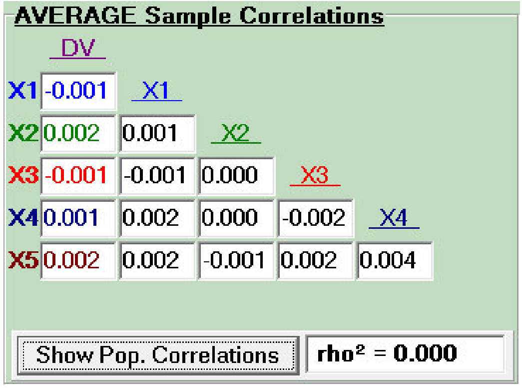


Population Parameters

# of Predictors: ○ 1 ○ 2 ○ 3 ○ 4 ● 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	0.0	1.0	0.000	0.994
X1:	0.0	1.0	0.004	0.994
X2:	0.0	1.0	0.001	0.995
X3:	0.0	1.0	0.000	0.994
X4:	0.0	1.0	0.001	0.995
X5:	0.0	1.0	0.000	0.992

The average ACTUAL means and standard deviations are reported in aqua.

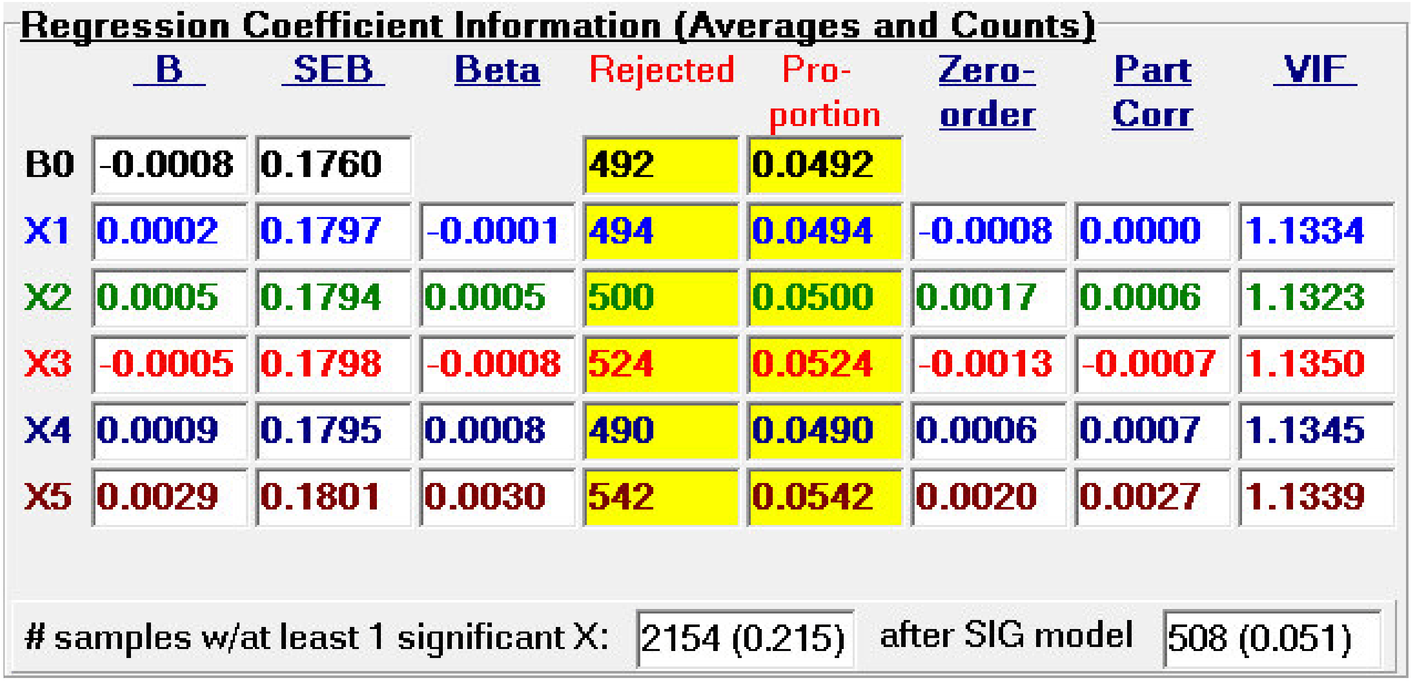


AVERAGE Sample Correlations

	DV	X1	X2	X3	X4
X1	-0.001				
X2	0.002	0.001			
X3	-0.001	-0.001	0.000		
X4	0.001	0.002	0.000	-0.002	
X5	0.002	0.002	-0.001	0.002	0.004

rho² = 0.000

If you hit the “Show Actual Correlations” button, you can see the average ACTUAL correlations. (*You must hit “Show Pop. Correlations” to run another analysis*)



Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0008	0.1760		492	0.0492			
X1	0.0002	0.1797	-0.0001	494	0.0494	-0.0008	0.0000	1.1334
X2	0.0005	0.1794	0.0005	500	0.0500	0.0017	0.0006	1.1323
X3	-0.0005	0.1798	-0.0008	524	0.0524	-0.0013	-0.0007	1.1350
X4	0.0009	0.1795	0.0008	490	0.0490	0.0006	0.0007	1.1345
X5	0.0029	0.1801	0.0030	542	0.0542	0.0020	0.0027	1.1339

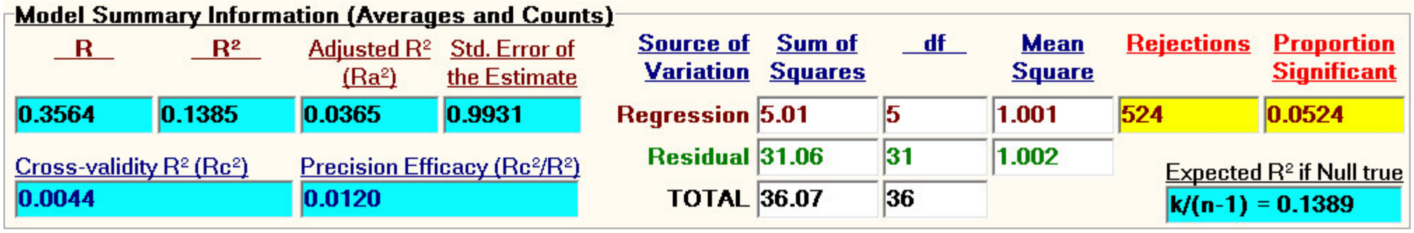
samples w/at least 1 significant X: 2154 (0.215) after SIG model 508 (0.051)

The average ACTUAL regression coefficient information is reported in this box — except for the “Rejected” and “Proportion” columns, which report the number (and proportion) of samples in which the particular regression coefficient (represented by X1, X2, etc.) was statistically significant.

“# samples w/at least 1 significant X” reports how many samples had at least one significant predictor.

“after SIG model” reports how many samples had at least one significant predictor following a significant overall regression model (the idea being that we don’t usually examine the statistical significance of regression coefficients unless the model was first significant—but that doesn’t mean that some predictors weren’t significant anyway).

B0 represents the CONSTANT in the regression equation. By default, B0 is not included in the 2 counts (above), but there is a menu option that will allow it to be included.



Model Summary Information (Averages and Counts)

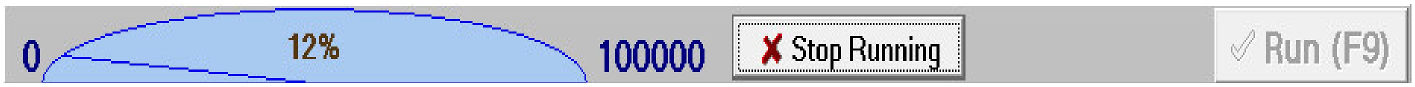
R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.3564	0.1385	0.0365	0.9931

Cross-validity R² (Rc²)	Precision Efficacy (Rc²/R²)
0.0044	0.0120

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	5.01	5	1.001	524	0.0524
Residual	31.06	31	1.002		
TOTAL	36.07	36			

Expected R² if Null true: k/(n-1) = 0.1389

Model summary information is provided here. Again, these are AVERAGE results except for the “Rejections” and “Proportion Significant” columns, which report how many (and the proportion of) samples that had statistically significant overall regression models.



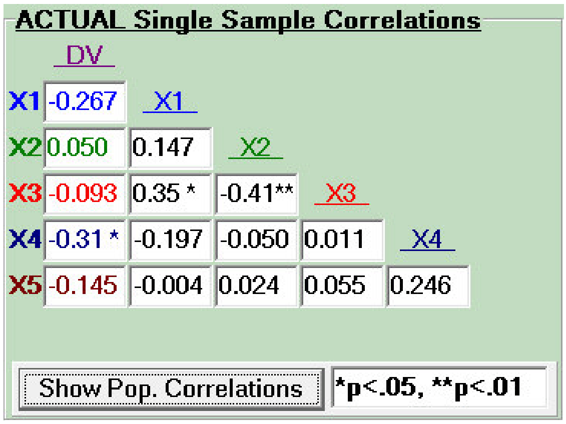
While the Monte Carlo simulations are running, the bottom panel (progress bar) looks like this. You can stop the Monte Carlo analyses if you need to by clicking the “Stop Running” button.

After the analysis is finished, the bottom panel will look like this. If you have aborted the process by pressing the “Stop Running” button, the number actually finished will appear in the panel.

Click "Show Pop. Correlations" button to RUN another analysis

If you review the ACTUAL correlations by clicking on the “Show Actual Correlations” button, you will not be able to continue with additional Monte Carlo analyses until you press the “Show Pop. Correlations” button (which is actually the same button as the “Show Actual Correlations” button.

Although not done in this example, when you run multiple SINGLE SAMPLE analyses, you will have the option of going backwards by one sample. Often, you get to clicking the “Run” button too quickly and you aren’t able to stop on a sample with interesting results. This “Back Up” button will allow you to go back 1 sample (but only 1).

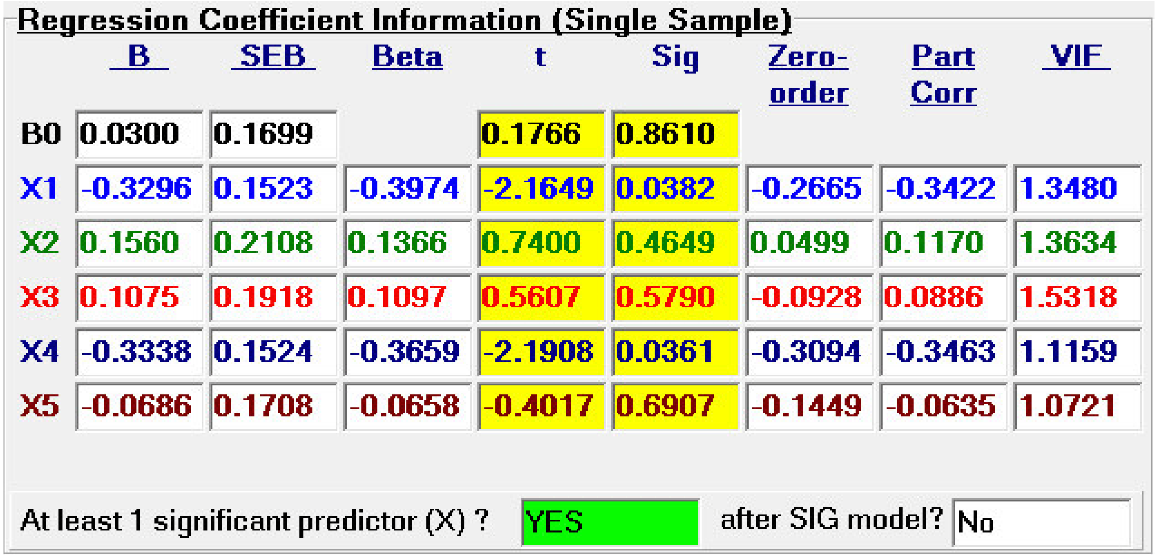


ACTUAL Single Sample Correlations

	DV	X1	X2	X3	X4
X1	-0.267				
X2	0.050	0.147			
X3	-0.093	0.35 *	-0.41**		
X4	-0.31 *	-0.197	-0.050	0.011	
X5	-0.145	-0.004	0.024	0.055	0.246

*p<.05, **p<.01

Another difference for SINGLE SAMPLE analyses is that statistically significant pairwise correlations are marked with asterisks when you click “Show Actual Correlations.”



Regression Coefficient Information (Single Sample)

	B	SEB	Beta	t	Sig	Zero-order	Part Corr	VIF
B0	0.0300	0.1699		0.1766	0.8610			
X1	-0.3296	0.1523	-0.3974	-2.1649	0.0382	-0.2665	-0.3422	1.3480
X2	0.1560	0.2108	0.1366	0.7400	0.4649	0.0499	0.1170	1.3634
X3	0.1075	0.1918	0.1097	0.5607	0.5790	-0.0928	0.0886	1.5318
X4	-0.3338	0.1524	-0.3659	-2.1908	0.0361	-0.3094	-0.3463	1.1159
X5	-0.0686	0.1708	-0.0658	-0.4017	0.6907	-0.1449	-0.0635	1.0721

At least 1 significant predictor (X) ? YES

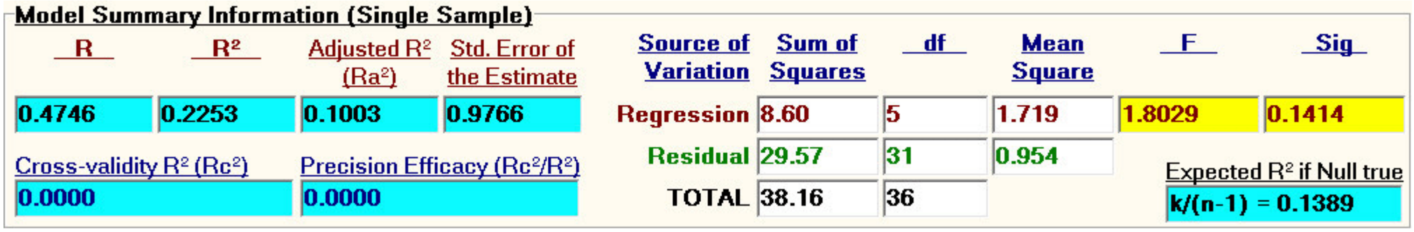
after SIG model? No

For SINGLE SAMPLE analyses, the “Rejected” and “Proportion” columns change to the actual *t* statistics and *p* values (“Sig”) for each regression coefficient.

By the way, “B” is the unstandardized regression coefficient, “SEB” is the standard error for the unstandardized regression coefficient, “Beta” is the standardized regression coefficient, “Zero-order” is the Pearson correlation between each predictor and Y, “Part Corr” is the part (or semi-partial) correlation between each predictor and Y GIVEN the other predictors in the model, and “VIF” is the variance inflation factor (1/Tolerance) used for diagnosing multicollinearity.

The “At least 1 significant predictor (X) ?” box shows whether any of the regression coefficients was statistically significant (but not which one).

Both bottom boxes turn from white to GREEN if “YES”



Model Summary Information (Single Sample)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.4746	0.2253	0.1003	0.9766

Cross-validity R² (Rc²)	Precision Efficacy (Rc²/R²)
0.0000	0.0000

Source of Variation	Sum of Squares	df	Mean Square	F	Sig
Regression	8.60	5	1.719	1.8029	0.1414
Residual	29.57	31	0.954		
TOTAL	38.16	36			

Expected R² if Null true

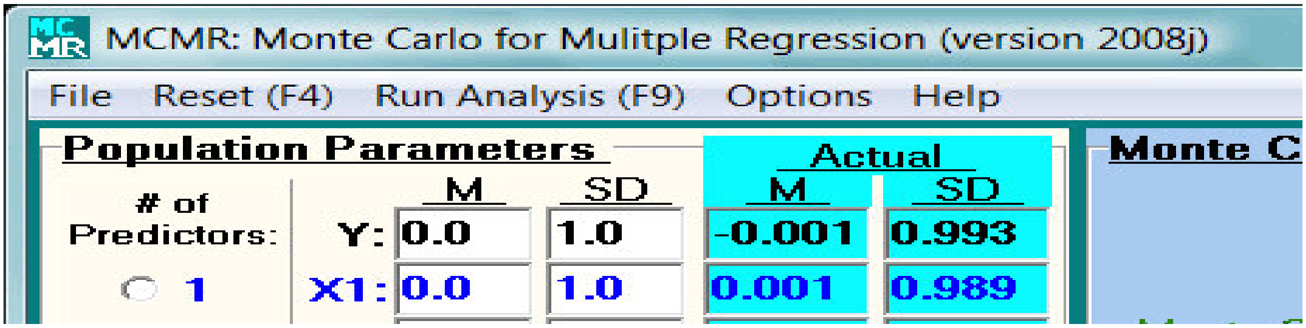
k/(n-1) = 0.1389

For SINGLE SAMPLE analyses, the “Rejections” and “Proportion Significant” columns change to the actual *F* statistic and the actual *p* value significance of the regression model (“Sig”).

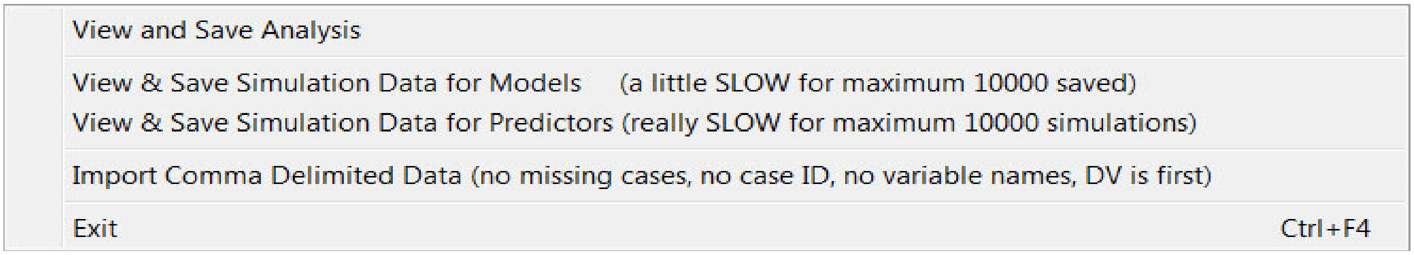
If the model is statistically significant, the “F” and “Sig” boxes turn from yellow to GREEN. If Adjusted R^2 or Cross-validity R^2 are negative they are set to 0.0 (theoretically, neither they nor R^2 can be negative).

By the way, the “Expected R^2 if Null True” box uses the calculation presented by Herzberg (1969), *k*/(*n*-1), to show the bias of the R^2 statistic. The “Options” menu allows you to change the information reported here to a few other things.

Menus



"File," "Options," and "Help" show sub-menus (below), but "Reset (F4)" and "Run Analysis (F9)" just perform the given action. "Reset (F4)" will return the program to the main opening screen and "Run Analysis (F9)" will run the analysis, just like clicking the "Run (F9)" button or pressing the F9 key.



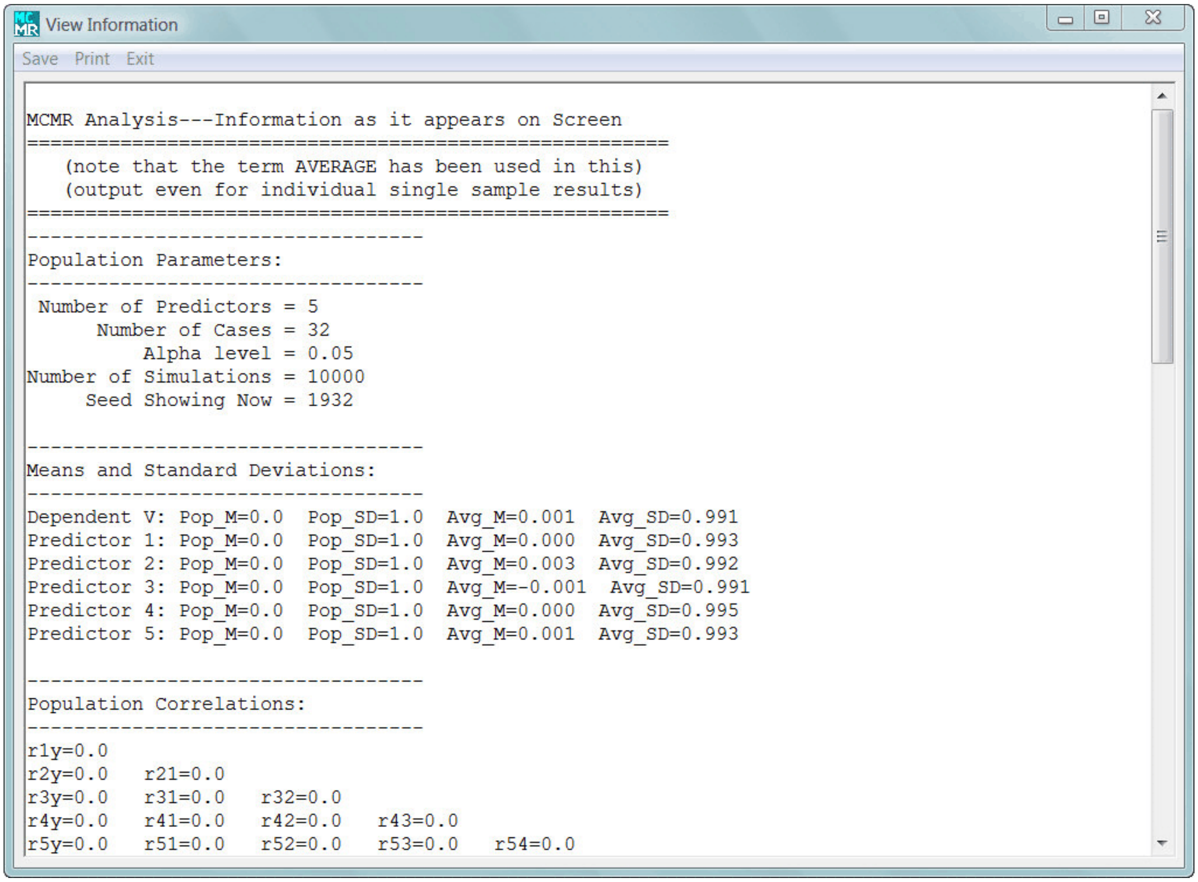
"View and Save Analysis" will show a text version of the results in another window (below), which will also allow you to save and print the results of the analysis.

"View & Save Simulation Data for Models" will save the Model Summary statistics (e.g., R^2, Standard Error of the Estimate) from all the Monte Carlo simulated samples (up to a maximum of 10,000) for analysis in any program that accepts Comma-Delimited text files. Variable names ARE included on the first line of the file.

"View & Save Simulation Data for Predictors" will save the Regression Coefficient statistics (e.g., B, SEB, Beta) from all the Monte Carlo simulated samples (up to a maximum of 10,000) for analysis in any program that accepts Comma-Delimited text files. Variable names ARE included on the first line of the file.

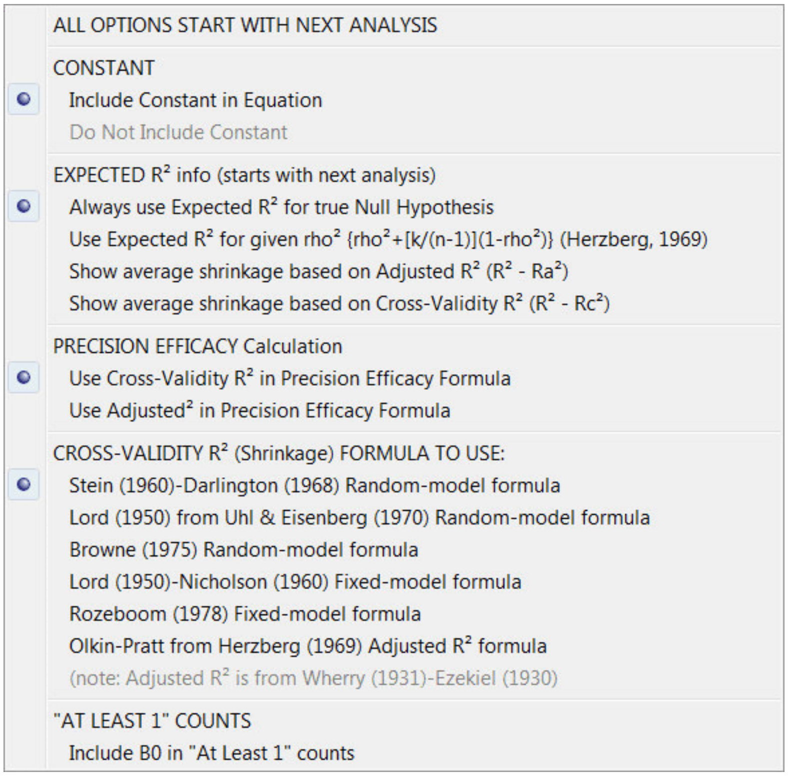
If you are running a SINGLE SAMPLE analysis, there is also an option to save SINGLE SAMPLE data. The data from the current single sample analysis is saved WITHOUT variable names on the first line.

"Import Comma Delimited Data" will allow you to read in data that you have previously saved with MCMR, or will allow you to import data saved in appropriate format from any other program (e.g., a spreadsheet or statistics program). The MCMR program assumes that NO variable names are listed on the first line—that is, the data begin on line 1.



```
MCMR Analysis---Information as it appears on Screen
=======================================================
   (note that the term AVERAGE has been used in this)
   (output even for individual single sample results)
=======================================================
----------------------------------
Population Parameters:
----------------------------------
 Number of Predictors = 5
      Number of Cases = 32
          Alpha level = 0.05
Number of Simulations = 10000
     Seed Showing Now = 1932

----------------------------------
Means and Standard Deviations:
----------------------------------
Dependent V: Pop_M=0.0  Pop_SD=1.0  Avg_M=0.001  Avg_SD=0.991
Predictor 1: Pop_M=0.0  Pop_SD=1.0  Avg_M=0.000  Avg_SD=0.993
Predictor 2: Pop_M=0.0  Pop_SD=1.0  Avg_M=0.003  Avg_SD=0.992
Predictor 3: Pop_M=0.0  Pop_SD=1.0  Avg_M=-0.001  Avg_SD=0.991
Predictor 4: Pop_M=0.0  Pop_SD=1.0  Avg_M=0.000  Avg_SD=0.995
Predictor 5: Pop_M=0.0  Pop_SD=1.0  Avg_M=0.001  Avg_SD=0.993

----------------------------------
Population Correlations:
----------------------------------
r1y=0.0
r2y=0.0   r21=0.0
r3y=0.0   r31=0.0   r32=0.0
r4y=0.0   r41=0.0   r42=0.0   r43=0.0
r5y=0.0   r51=0.0   r52=0.0   r53=0.0   r54=0.0
```

All “View and Save” options will open this window. From here, you can “Save” or “Print” the information in the window (using the appropriate menu option).



ALL OPTIONS START WITH NEXT ANALYSIS

CONSTANT
- ● Include Constant in Equation
- Do Not Include Constant

EXPECTED R^2 info (starts with next analysis)
- ● Always use Expected R^2 for true Null Hypothesis
- Use Expected R^2 for given rho^2 {rho^2+[k/(n-1)](1-rho^2)} (Herzberg, 1969)
- Show average shrinkage based on Adjusted R^2 (R^2 - Ra^2)
- Show average shrinkage based on Cross-Validity R^2 (R^2 - Rc^2)

PRECISION EFFICACY Calculation
- ● Use Cross-Validity R^2 in Precision Efficacy Formula
- Use Adjusted^2 in Precision Efficacy Formula

CROSS-VALIDITY R^2 (Shrinkage) FORMULA TO USE:
- ● Stein (1960)-Darlington (1968) Random-model formula
- Lord (1950) from Uhl & Eisenberg (1970) Random-model formula
- Browne (1975) Random-model formula
- Lord (1950)-Nicholson (1960) Fixed-model formula
- Rozeboom (1978) Fixed-model formula
- Olkin-Pratt from Herzberg (1969) Adjusted R^2 formula
- (note: Adjusted R^2 is from Wherry (1931)-Ezekiel (1930)

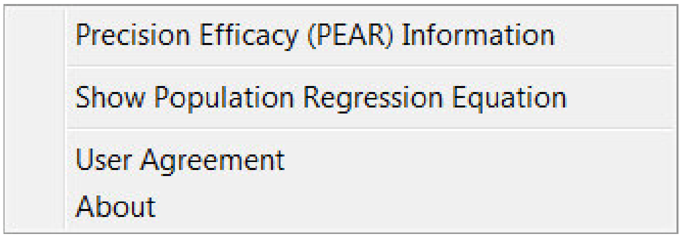
"AT LEAST 1" COUNTS
- Include B0 in "At Least 1" counts

Currently, only analyses with the Constant Included in the Equation are permitted.

There are 4 types of information that can be reported in the box that by default is labeled “Expected R^2 if Null true” — 2 for expected R^2 and 2 for shrinkage.

Precision Efficacy (Brooks, 1998) is calculated using Cross-Validity R^2 by default, but could be calculated using Adjusted R^2. (see help menu for additional information about Precision Efficacy) Different formulas can be used to calculate Cross-Validity R^2 — 6 are available here.

You can choose to have significant B0 included in the counts reported (by default it is not).

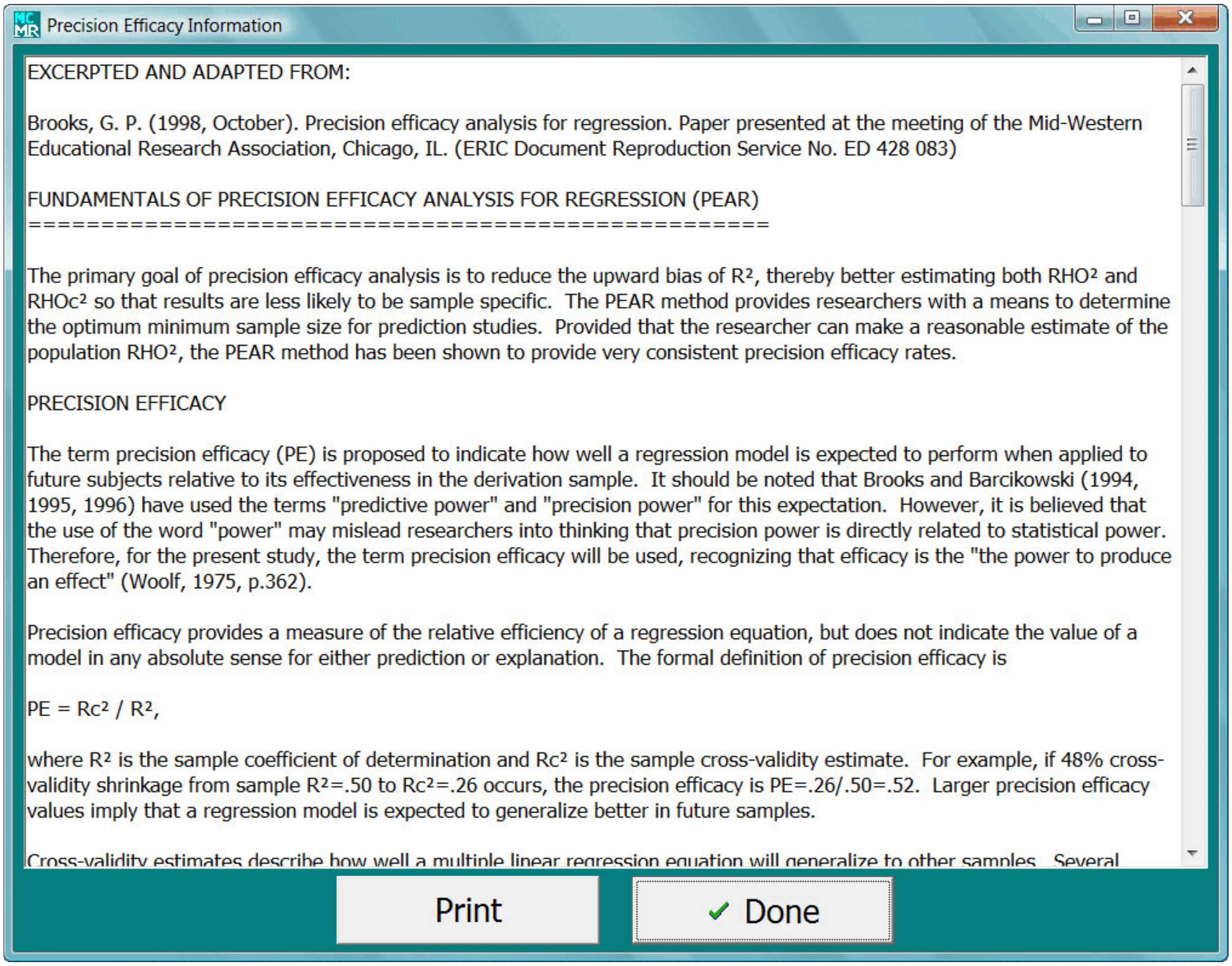


The “Precision Efficacy (PEAR) Information” option will open a window that contains an except from a paper written in 1998 (see below).

“Show Population Regression Equation” will show the STANDARDIZED regression model based on the Population Correlation matrix used to generate data for the analysis.

“User Agreement” opens a window with LICENSE information (important).

“About” provides some basic information about the MCMR program.



EXCERPTED AND ADAPTED FROM:

Brooks, G. P. (1998, October). Precision efficacy analysis for regression. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 428 083)

FUNDAMENTALS OF PRECISION EFFICACY ANALYSIS FOR REGRESSION (PEAR)
===

The primary goal of precision efficacy analysis is to reduce the upward bias of R^2, thereby better estimating both RHO^2 and $RHOc^2$ so that results are less likely to be sample specific. The PEAR method provides researchers with a means to determine the optimum minimum sample size for prediction studies. Provided that the researcher can make a reasonable estimate of the population RHO^2, the PEAR method has been shown to provide very consistent precision efficacy rates.

PRECISION EFFICACY

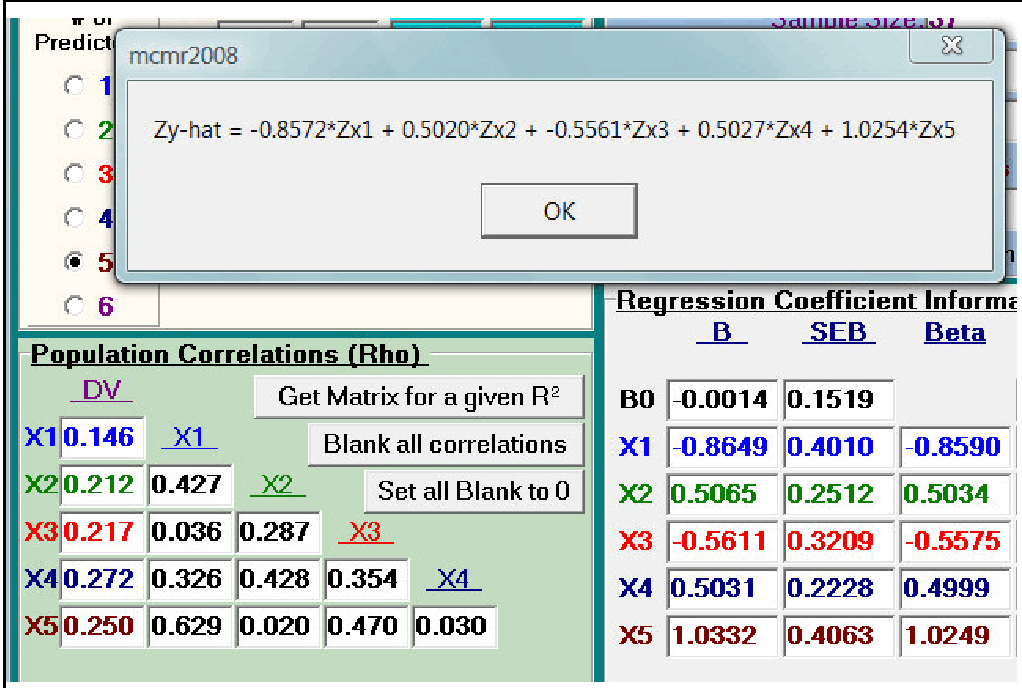
The term precision efficacy (PE) is proposed to indicate how well a regression model is expected to perform when applied to future subjects relative to its effectiveness in the derivation sample. It should be noted that Brooks and Barcikowski (1994, 1995, 1996) have used the terms "predictive power" and "precision power" for this expectation. However, it is believed that the use of the word "power" may mislead researchers into thinking that precision power is directly related to statistical power. Therefore, for the present study, the term precision efficacy will be used, recognizing that efficacy is the "the power to produce an effect" (Woolf, 1975, p.362).

Precision efficacy provides a measure of the relative efficiency of a regression equation, but does not indicate the value of a model in any absolute sense for either prediction or explanation. The formal definition of precision efficacy is

$PE = Rc^2 / R^2$,

where R^2 is the sample coefficient of determination and Rc^2 is the sample cross-validity estimate. For example, if 48% cross-validity shrinkage from sample $R^2=.50$ to $Rc^2=.26$ occurs, the precision efficacy is $PE=.26/.50=.52$. Larger precision efficacy values imply that a regression model is expected to generalize better in future samples.

Cross-validity estimates describe how well a multiple linear regression equation will generalize to other samples. Several



```
Zy-hat = -0.8572*Zx1 + 0.5020*Zx2 + -0.5561*Zx3 + 0.5027*Zx4 + 1.0254*Zx5
```

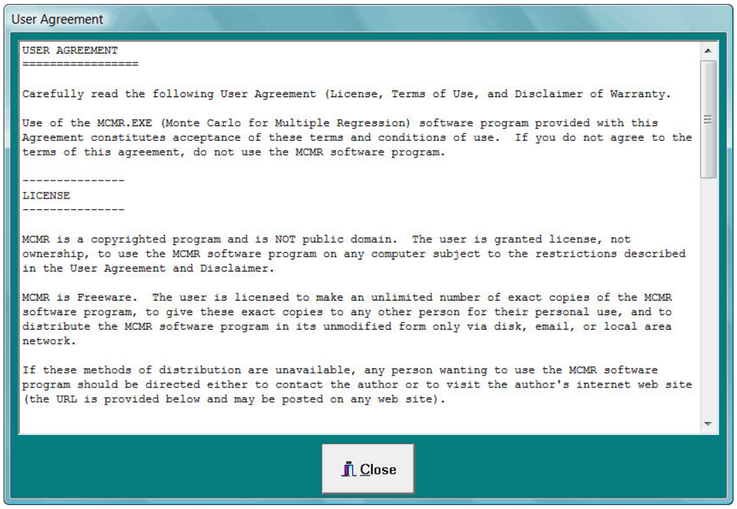
Population Correlations (Rho)

	DV	X1	X2	X3	X4
X1	0.146				
X2	0.212	0.427			
X3	0.217	0.036	0.287		
X4	0.272	0.326	0.428	0.354	
X5	0.250	0.629	0.020	0.470	0.030

Regression Coefficient Informa

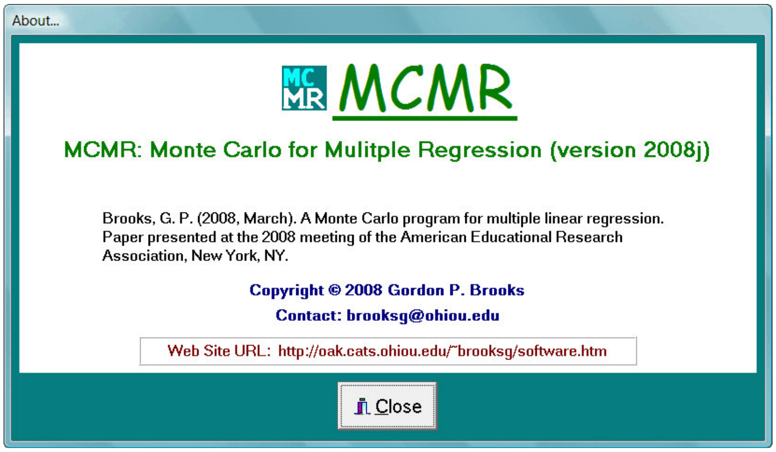
	B	SEB	Beta
B0	-0.0014	0.1519	
X1	-0.8649	0.4010	-0.8590
X2	0.5065	0.2512	0.5034
X3	-0.5611	0.3209	-0.5575
X4	0.5031	0.2228	0.4999
X5	1.0332	0.4063	1.0249

At any point, the user can request this pop-up window that shows the Population Standardized Regression model for comparison to current results.



```
USER AGREEMENT
=================

Carefully read the following User Agreement (License, Terms of Use, and Disclaimer of Warranty.

Use of the MCMR.EXE (Monte Carlo for Multiple Regression) software program provided with this
Agreement constitutes acceptance of these terms and conditions of use.  If you do not agree to the
terms of this agreement, do not use the MCMR software program.

---------------
LICENSE
---------------

MCMR is a copyrighted program and is NOT public domain.  The user is granted license, not
ownership, to use the MCMR software program on any computer subject to the restrictions described
in the User Agreement and Disclaimer.

MCMR is Freeware.  The user is licensed to make an unlimited number of exact copies of the MCMR
software program, to give these exact copies to any other person for their personal use, and to
distribute the MCMR software program in its unmodified form only via disk, email, or local area
network.

If these methods of distribution are unavailable, any person wanting to use the MCMR software
program should be directed either to contact the author or to visit the author's internet web site
(the URL is provided below and may be posted on any web site).
```

The User Agreement window, with important information about the legal use of the software.



MCMR: Monte Carlo for Mulitple Regression (version 2008j)

Brooks, G. P. (2008, March). A Monte Carlo program for multiple linear regression. Paper presented at the 2008 meeting of the American Educational Research Association, New York, NY.

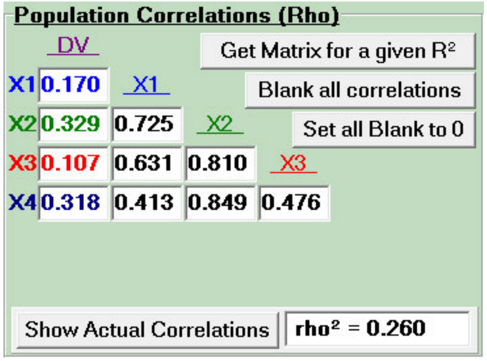
Copyright © 2008 Gordon P. Brooks

Contact: brooksg@ohiou.edu

Web Site URL: http://oak.cats.ohiou.edu/~brooksg/software.htm

The ABOUT window with the full name of the program, copyright and contact information, and the web site from which this and other software programs may be obtained.

Secondary Window: Get a Population Matrix with certain Given Characteristics

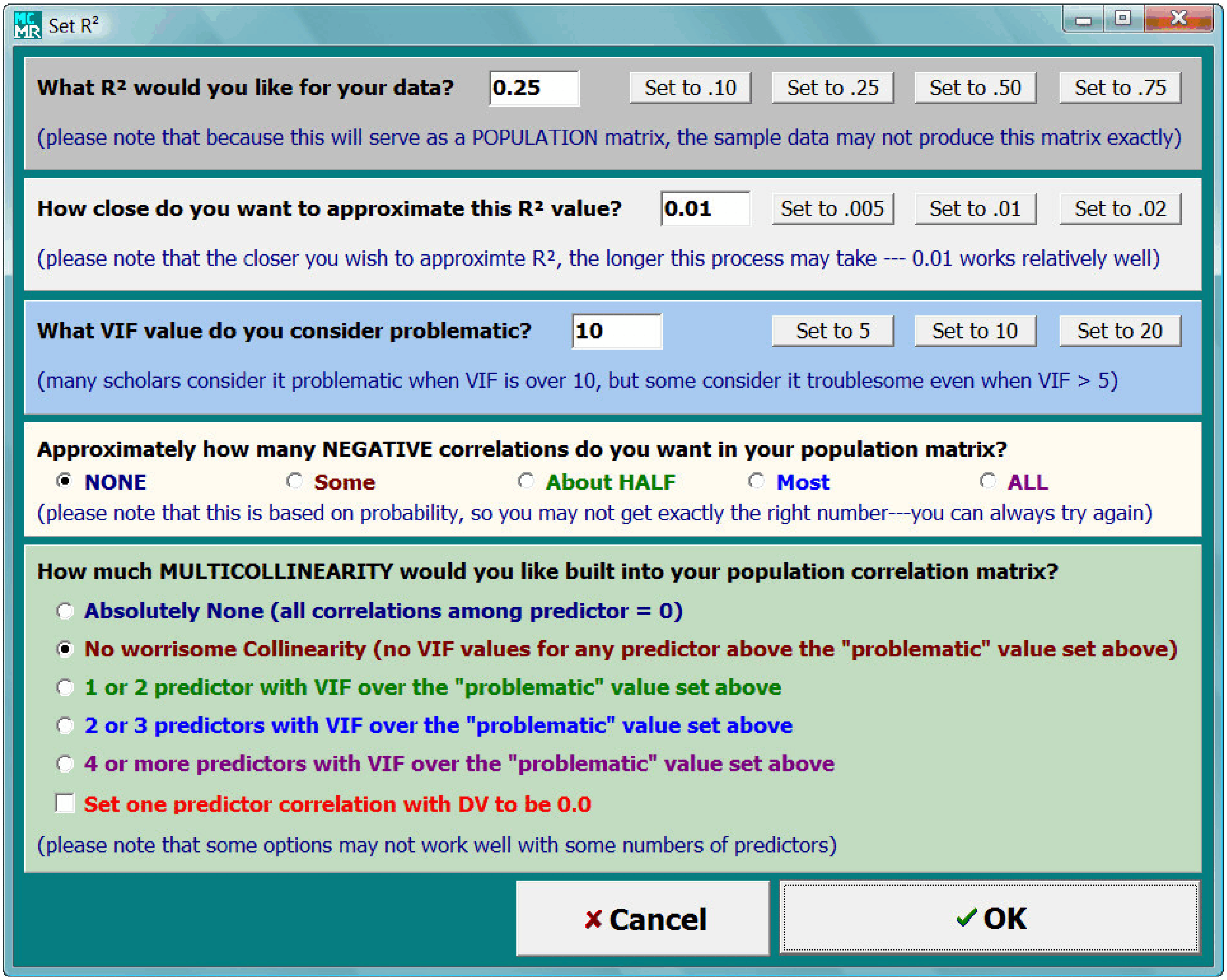


Population Correlations (Rho)

	DV	X1	X2	X3
X1	0.170			
X2	0.329	0.725		
X3	0.107	0.631	0.810	
X4	0.318	0.413	0.849	0.476

rho² = 0.260

If you click the “Get Matrix for a Given R2" button, the following window will open — allowing you to get a correlation matrix that meets certain criteria.



Set R²

What R² would you like for your data? 0.25

(please note that because this will serve as a POPULATION matrix, the sample data may not produce this matrix exactly)

How close do you want to approximate this R² value? 0.01

(please note that the closer you wish to approximte R², the longer this process may take --- 0.01 works relatively well)

What VIF value do you consider problematic? 10

(many scholars consider it problematic when VIF is over 10, but some consider it troublesome even when VIF > 5)

Approximately how many NEGATIVE correlations do you want in your population matrix?

● NONE ○ Some ○ About HALF ○ Most ○ ALL

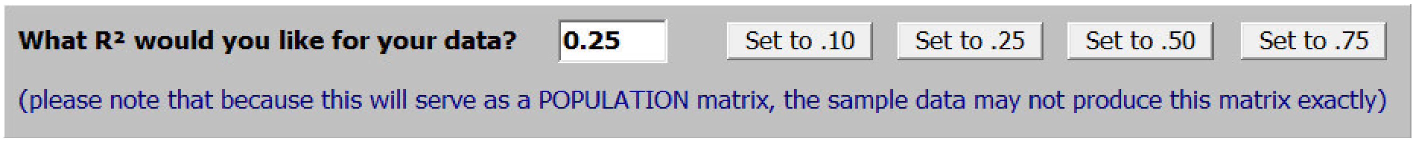
(please note that this is based on probability, so you may not get exactly the right number---you can always try again)

How much MULTICOLLINEARITY would you like built into your population correlation matrix?

- ○ Absolutely None (all correlations among predictor = 0)
- ● No worrisome Collinearity (no VIF values for any predictor above the "problematic" value set above)
- ○ 1 or 2 predictor with VIF over the "problematic" value set above
- ○ 2 or 3 predictors with VIF over the "problematic" value set above
- ○ 4 or more predictors with VIF over the "problematic" value set above
- ☐ Set one predictor correlation with DV to be 0.0

(please note that some options may not work well with some numbers of predictors)

Each section is described more below. When you click “OK” a correlation matrix will be found with the given criteria (if possible) AND that correlation matrix will be transferred to the main MCMR program screen into the “Population Correlations (rho)” section.

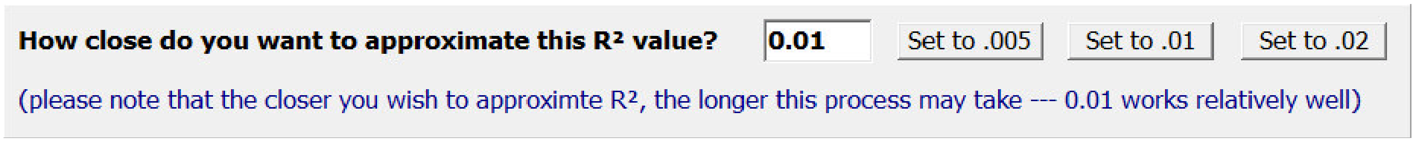


What R² would you like for your data? 0.25

(please note that because this will serve as a POPULATION matrix, the sample data may not produce this matrix exactly)

You can choose any R^2 for your POPULATION correlation matrix (so really this is a rho^2 or ρ^2 value), but buttons are provided for some common values (these are based on tables from Park & Dudycha, 1974).

Remember, however, that this will derive a POPULATION correlation matrix, from which samples will be drawn during the Monte Carlo process. This value says nothing specific about any of the R^2 values calculated in the samples (other than they should be from the population with the derived population correlation matrix).

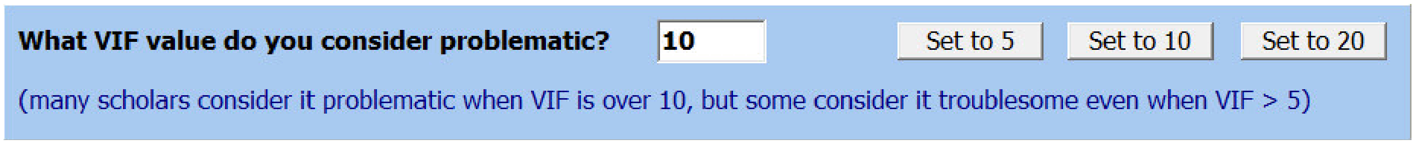


How close do you want to approximate this R² value? 0.01

(please note that the closer you wish to approximte R², the longer this process may take --- 0.01 works relatively well)

You can choose how close you want to approximate the population R^2 set in the previous box. While it is indeed possible to approximate some matrices very closely, anything smaller than 0.001 will likely take a good deal of time. The values 0.01, or even 0.005, seem to work pretty well if you really want to get exact.

Remember, however, that this is how closely you approximate the desired population R^2 in the POPULATION correlation matrix, and says nothing about the samples drawn during the Monte Carlo process.

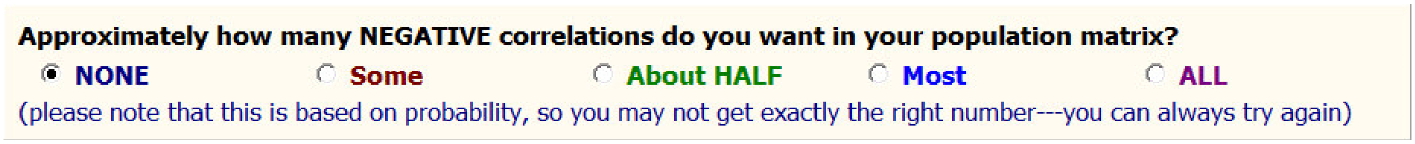


What VIF value do you consider problematic? 10

(many scholars consider it problematic when VIF is over 10, but some consider it troublesome even when VIF > 5)

You can set any value above 1.0 for the critical VIF threshold value. Most scholars choose 5.0 or 10.0, depending on how much MULTICOLLINEARITY(also called COLLINEARITY) you're willing to tolerate.

Recall that **VIF = 1/Tolerance**, where **Tolerance** $= 1 - \mathbf{R}_j^2$, where $\mathbf{R}_j^2$ is the squared correlation when the **j**th predictor acts as a temporary dependent variable being predicted by all the other predictors.

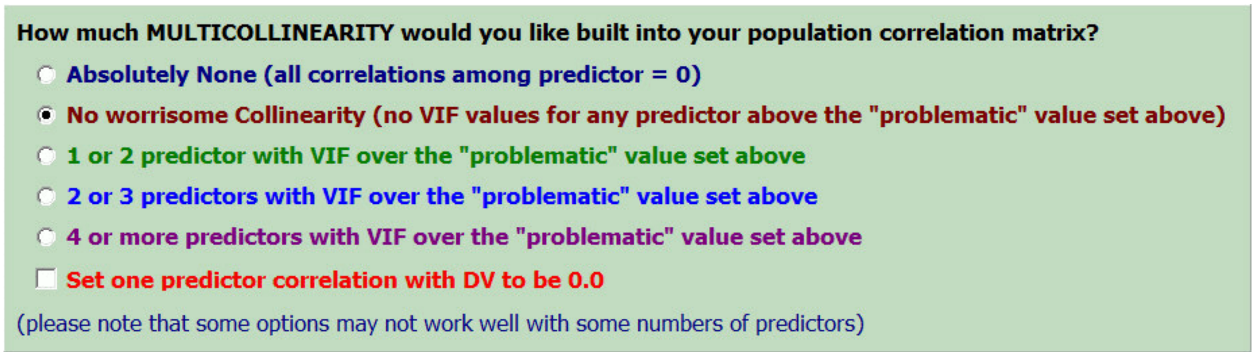


Approximately how many NEGATIVE correlations do you want in your population matrix?

● **NONE** ○ **Some** ○ **About HALF** ○ **Most** ○ **ALL**

(please note that this is based on probability, so you may not get exactly the right number---you can always try again)

This option will allow you to create a population correlation matrix with some (or many) negative correlations.



How much MULTICOLLINEARITY would you like built into your population correlation matrix?

○ Absolutely None (all correlations among predictor = 0)
● No worrisome Collinearity (no VIF values for any predictor above the "problematic" value set above)
○ 1 or 2 predictor with VIF over the "problematic" value set above
○ 2 or 3 predictors with VIF over the "problematic" value set above
○ 4 or more predictors with VIF over the "problematic" value set above
☐ Set one predictor correlation with DV to be 0.0

(please note that some options may not work well with some numbers of predictors)

This box will allow you to request a certain level of multicollinearity in your population correlation matrix.

$ “Absolutely None” requires that all intercorrelations among predictors are 0.0, but the correlations between the predictors and Y will be set randomly to provide the R^2 given above.

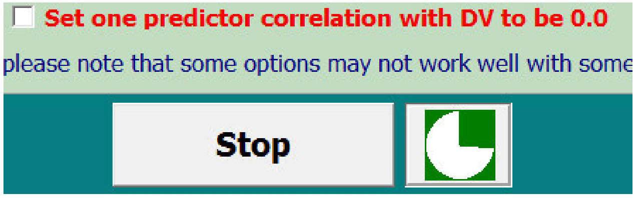
$ “No Worrisome Collinearity” will produce a population correlation matrix where all predictor intercorrelations will be non-zero, but will be probably smaller than the critical VIF set above.

$ “1 or 2 predictors with VIF” will produce a population correlation matrix such that predictor intercorrelations will probably result in at least 1, but not more than 2, VIF values over the critical value

$ “2 or 3 predictors with VIF” will produce a population correlation matrix such that predictor intercorrelations will probably result in at least 2, but not more than 3, VIF values over the critical value

$ “4 or more predictors with VIF” will produce a population correlation matrix such that predictor intercorrelations will probably result in at least 4 VIF values over the critical value

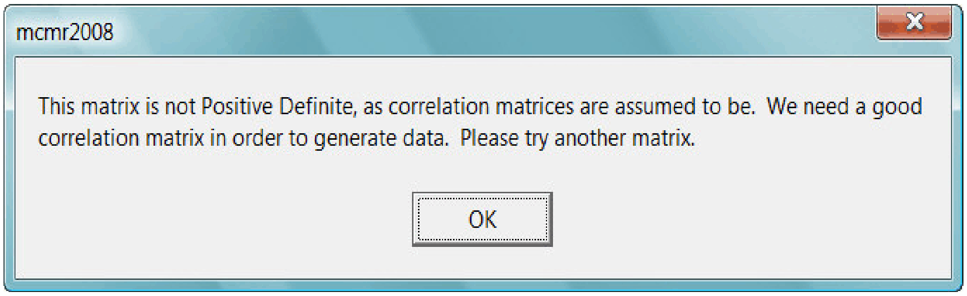
Note that “probably” was included in these descriptions. There are rare occasions, given certain starting correlations used in the algorithm, where the resulting correlation matrix does not match the criteria exactly. You can either go ahead and use the derived matrix, or simply try another. Different seeds used in each run of this sub-program result in different matrices being created.



☐ Set one predictor correlation with DV to be 0.0

please note that some options may not work well with some

If the little pie ever fills in all GREEN during this process, you probably have a matrix that cannot be created. You can try a few more times, if you’d like, because sometimes different seeds do produce workable results. You can also allow the program to continue running for a while, which sometimes will produce a workable result (the algorithm continues to adjust itself a little as it runs, which sometimes allows results to work).

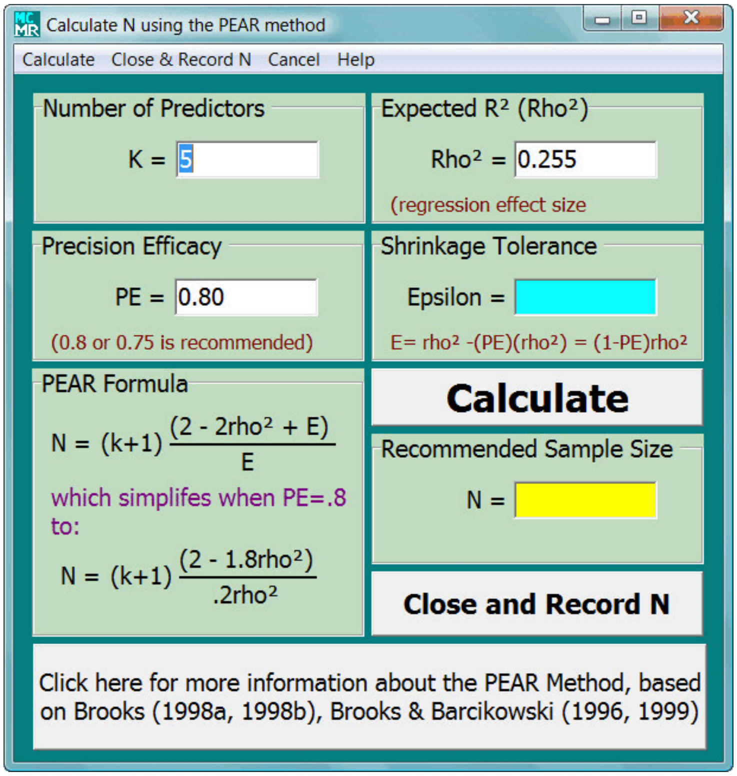


mcmr2008

This matrix is not Positive Definite, as correlation matrices are assumed to be. We need a good correlation matrix in order to generate data. Please try another matrix.

This error message will be shown whenever the “Stop” button is pushed (above), whenever the user has entered an inappropriate matrix, or on very rare occasions where rounding the derived correlations to 3 decimal places impacts the matrix enough to make it unusable.

Secondary Window: Get a Sample Size using the PEAR Method



K = 5

Rho² = 0.255

(regression effect size

PE = 0.80

(0.8 or 0.75 is recommended)

E= rho² -(PE)(rho²) = (1-PE)rho²

$$N = (k+1)\frac{(2 - 2rho^2 + E)}{E}$$

which simplifes when PE=.8 to:

$$N = (k+1)\frac{(2 - 1.8rho^2)}{.2rho^2}$$

Click here for more information about the PEAR Method, based on Brooks (1998a, 1998b), Brooks & Barcikowski (1996, 1999)

The user can change the parameters of the PEAR method (Brooks, 1998). By default, this window will provide the information for the analysis in the main window, if possible. For example, once the number of predictors is determined, it will be filled in here. Note that any number of predictors can be inserted.

More information about Precision Efficacy (PE) and the Precision Efficacy Analysis for Regression (PEAR) sample size method can be found by clicking the "Click here for more information" button (see below).

Briefly, however, Precision Efficacy is a complement to Proportional Shrinkage based on an appropriate Cross-validity R^2 (R_C^2) formula. Shrinkage itself (ε, or epsilon) can be written as

$$\boldsymbol{\varepsilon = R^2 - R_C^2}$$

whereas Proportional Shrinkage (PS) might be written as $\mathbf{PS = (R^2 - R_C^2) / R^2}$

Precision Efficacy would therefore be PE = 1 - PS, or

$$\mathbf{PE = R_C^2 / R^2}$$

Solving $PE = 1 - \varepsilon/R^2$ for ε, and replacing R^2 with an expected, a priori Re^2, results in the formula

$$\boldsymbol{\varepsilon = R_E^2 - (PE)(R_E^2)}$$

where R_E^2 is often just set at the expected population ρ^2. Because Precision Efficacy (PE) is usually set at .75 or .80, shrinkage would usually be $.25\rho^2$ or $.2\rho^2$, respectively. Note that shrinkage may also be set absolutely as something like $\varepsilon = .05$ or $\varepsilon = .10$.

Once parameters are set, "Calculate" will determine the required sample size. The recommended sample size will appear in the YELLOW box underneath the "Calculate" button.

"Close and Record N" will move this sample size to the main screen.

"Cancel" (on the menu bar) will close the dialog window without making any changes to the main screen.

Although the PEAR method was derived using Cross-Validity R^2 (Brooks, 1998), it is theoretically reasonable to apply the same idea to Precision Efficacy calculated using Adjusted R^2 instead. Algina and Olejnik (2000) have discussed a similar idea, but different approach, to sample sizes for Adjusted R^2.

In this case, sample sizes would be determined such that the SHRINKAGE from R^2 to Adjusted R^2 would be maintained within a certain range. For example, if R^2 is .25, then Adjusted R^2 would be at least .20 when Precision Efficacy of .80 was used as the criterion. The formula for sample sizes to be used with such an approach would be

$$\mathbf{N = (k + 1)(1 - R_E^2 + \varepsilon) / \varepsilon}$$

Where

R_E^2 = expected population ρ^2
k = number of predictors
$\varepsilon = (R^2 - R_A^2)$

as compared to

$$\mathbf{N = (k + 1)(2 - 2R_E^2 + \varepsilon) / \varepsilon}$$

where **$\varepsilon = (R^2 - R_C^2)$**

for Cross-Validity (see Brooks, 1998). Shrinkage tolerance can also be calculated as

$$\mathbf{\varepsilon = (1 - PE)\, R^2}$$

where, for PE = .80, it would simplify to (just like it would also for the Cross-Validity approach)

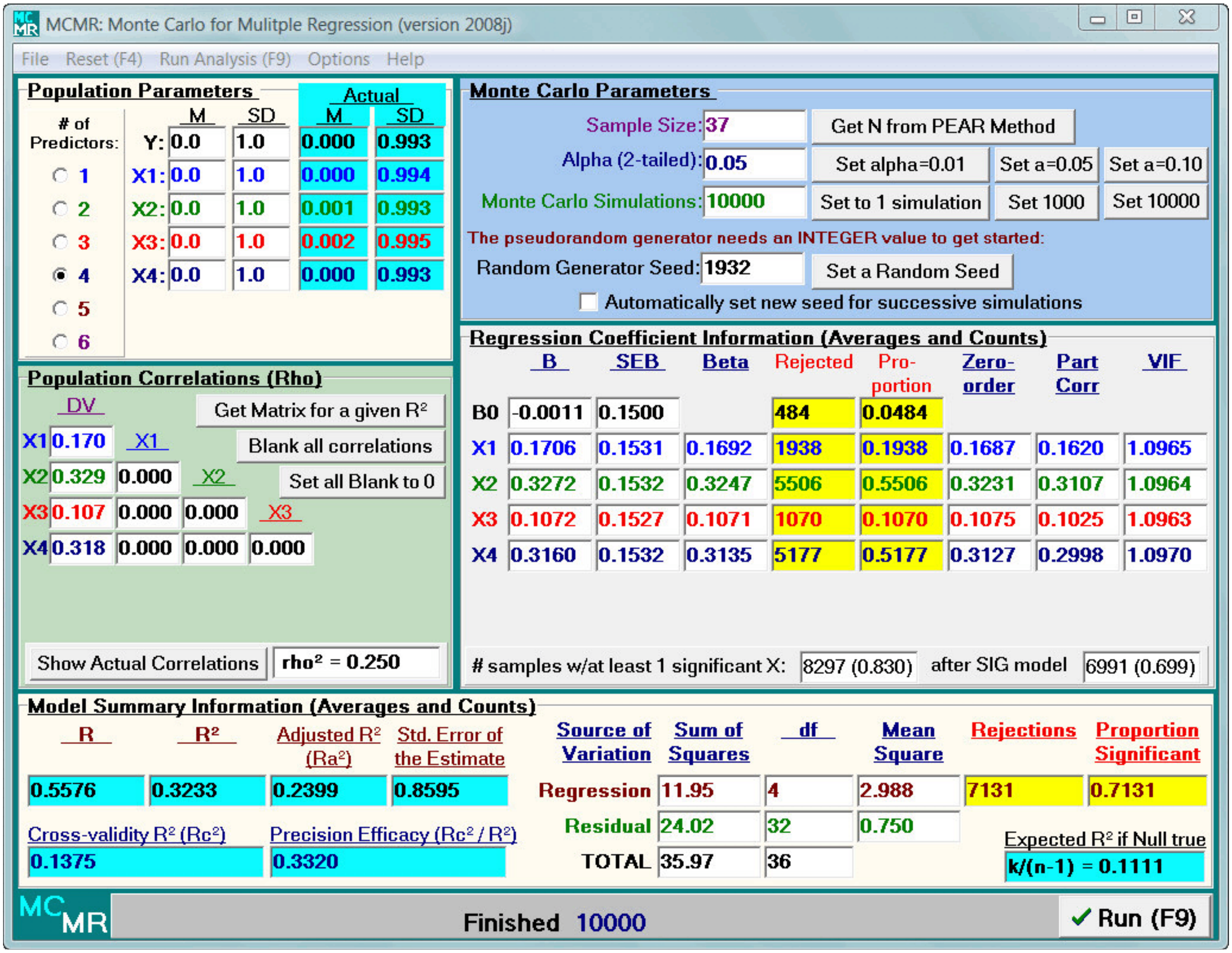
$$\mathbf{\varepsilon = .2R^2}$$

Recall that one of the options on the “Options” menu is to use Adjusted R^2 in the Precision Efficacy formula instead of Cross-Validity R^2.

The key difference is that for Cross-validity Precision Efficacy, the idea is to INCREASE Cross-validity R^2; however, for Adjusted R^2, the idea is more to DECREASE R^2, making it closer to the true population parameter (since Adjusted R^2 is usually a good estimate of rho^2).

Either method helps make the model more generalizable by decreasing the standard errors for the regression coefficients. The Corss-validity approach is more stringent because it accounts for error not only in the regression model derivation sample, but also for the error in future samples to which the regression model is applied.

An Example: Multicollinearity and Inflation of Standard Errors



Population Parameters

# of Predictors:		M	SD	Actual M	Actual SD
	Y:	0.0	1.0	0.000	0.993
○ 1	X1:	0.0	1.0	0.000	0.994
○ 2	X2:	0.0	1.0	0.001	0.993
○ 3	X3:	0.0	1.0	0.002	0.995
● 4	X4:	0.0	1.0	0.000	0.993
○ 5					
○ 6					

Monte Carlo Parameters

Sample Size: 37
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 10000
The pseudorandom generator needs an INTEGER value to get started:
Random Generator Seed: 1932
☐ Automatically set new seed for successive simulations

Population Correlations (Rho)

	DV	X1	X2	X3
X1	0.170			
X2	0.329	0.000		
X3	0.107	0.000	0.000	
X4	0.318	0.000	0.000	0.000

rho^2 = 0.250

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Pro-portion	Zero-order	Part Corr	VIF
B0	-0.0011	0.1500		484	0.0484			
X1	0.1706	0.1531	0.1692	1938	0.1938	0.1687	0.1620	1.0965
X2	0.3272	0.1532	0.3247	5506	0.5506	0.3231	0.3107	1.0964
X3	0.1072	0.1527	0.1071	1070	0.1070	0.1075	0.1025	1.0963
X4	0.3160	0.1532	0.3135	5177	0.5177	0.3127	0.2998	1.0970

samples w/at least 1 significant X: 8297 (0.830) after SIG model 6991 (0.699)

Model Summary Information (Averages and Counts)

R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.5576	0.3233	0.2399	0.8595

Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2 / R^2)
0.1375	0.3320

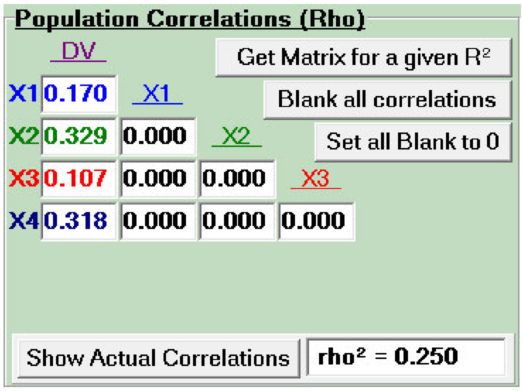
Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	11.95	4	2.988	7131	0.7131
Residual	24.02	32	0.750		
TOTAL	35.97	36			

Expected R^2 if Null true
k/(n-1) = 0.1111

Finished 10000

Let's assume that all predictor intercorrelations are 0.0, while predictor correlations with Y are non-zero such that rho^2 = .25.

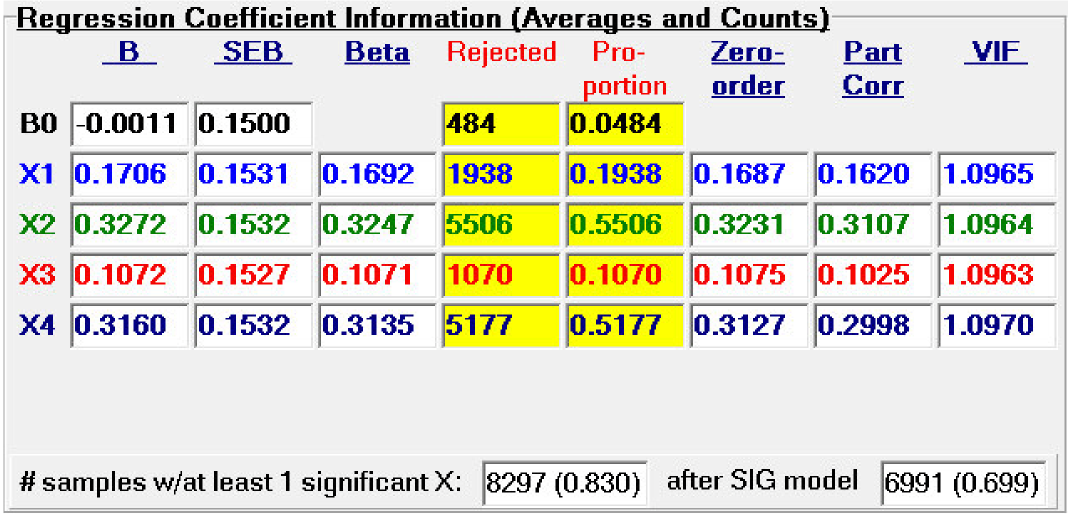
The seed is set to 1932, with N = 37, alpha = .05, and 10,000 simulated samples are drawn.



Population Correlations (Rho)

	DV	X1	X2	X3
X1	0.170			
X2	0.329	0.000		
X3	0.107	0.000	0.000	
X4	0.318	0.000	0.000	0.000

rho^2 = 0.250



Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Pro-portion	Zero-order	Part Corr	VIF
B0	-0.0011	0.1500		484	0.0484			
X1	0.1706	0.1531	0.1692	1938	0.1938	0.1687	0.1620	1.0965
X2	0.3272	0.1532	0.3247	5506	0.5506	0.3231	0.3107	1.0964
X3	0.1072	0.1527	0.1071	1070	0.1070	0.1075	0.1025	1.0963
X4	0.3160	0.1532	0.3135	5177	0.5177	0.3127	0.2998	1.0970

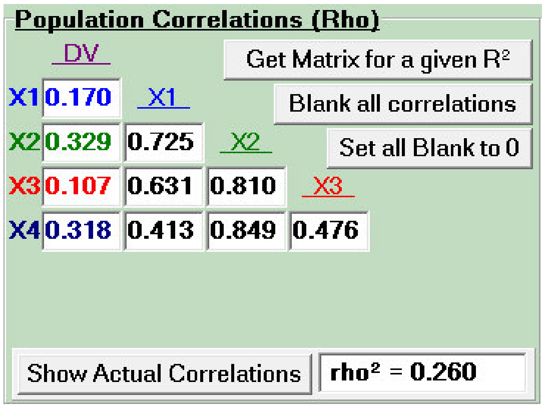
samples w/at least 1 significant X: 8297 (0.830) after SIG model 6991 (0.699)

In this case, the standard errors for the regression coefficients ("SEB") are each approximately 0.153. Note that the Variance Inflation Factors ("VIF") are all roughly 1.096—since there is no correlation among the predictors we would expect this to be near 1.0, but since each of the 10,000 samples drawn probably ad some minor correlation among the predictors, it will not be exactly 1.0.

If we arbitrarily add some correlation among the predictors, BUT LEAVE THE CORRELATIONS BETWEEN THE PREDICTORS AND Y THE SAME, we introduce multicollinearity.

Note that in this matrix, the rho^2 is not exactly .250 any more. This is arbitrary, but will have some minor impact on our results.

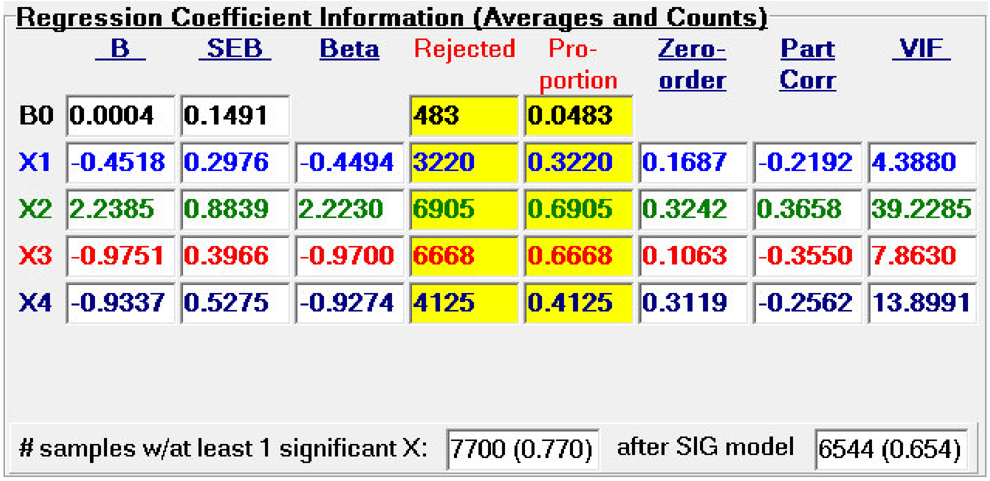
In particular, if you examine the model summary results (we won't here), you would see some minor differences — especially in R^2 and the Sum of Squares due to the regression (which impacts other things as well). This is not a REAL difference, but rather due to the different population conditions set by the slightly larger rho^2.



Population Correlations (Rho)

	DV	X1	X2	X3
X1	0.170			
X2	0.329	0.725		
X3	0.107	0.631	0.810	
X4	0.318	0.413	0.849	0.476

rho² = 0.260



Regression Coefficient Information (Averages and Counts)

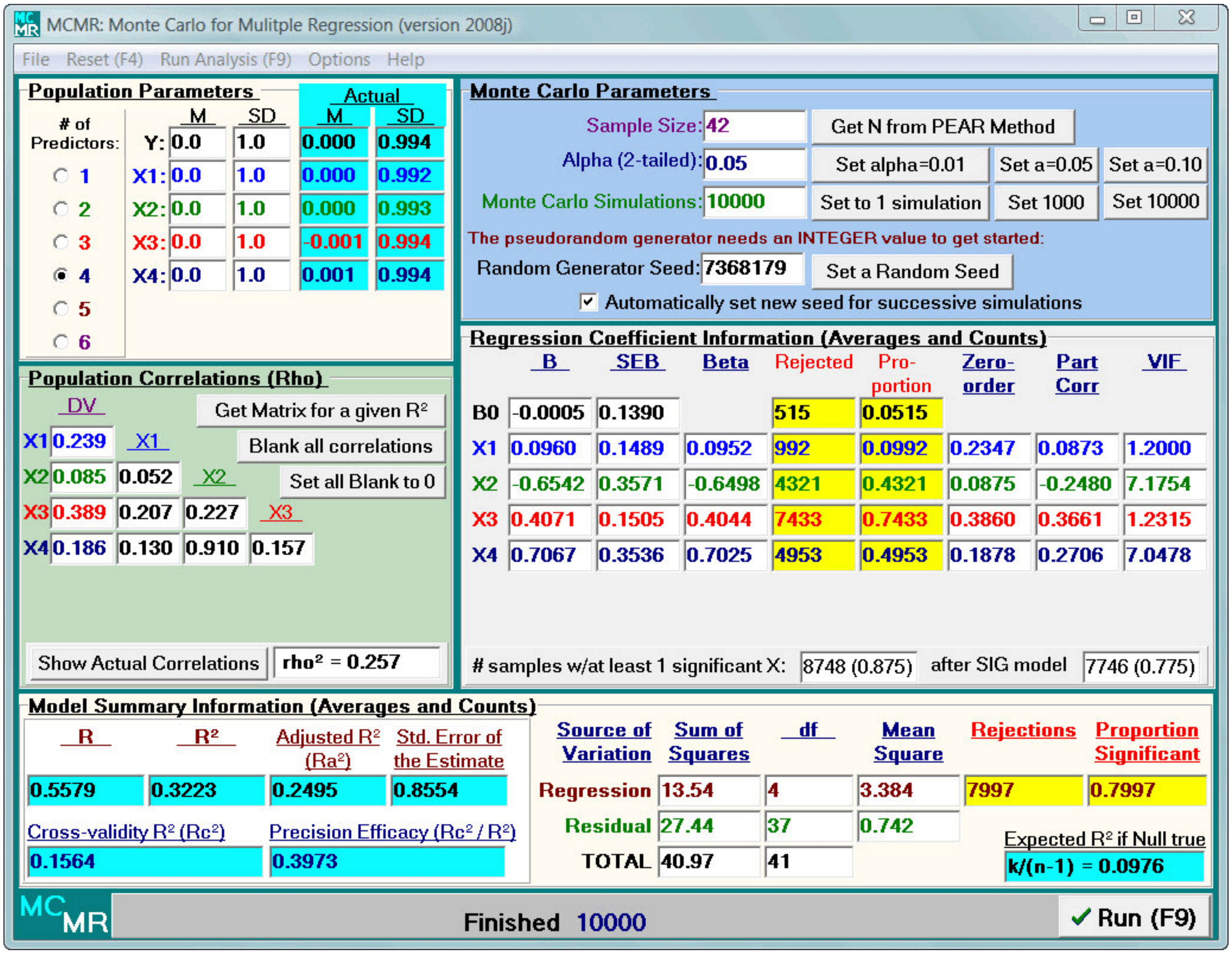
	B	SEB	Beta	Rejected	Pro-portion	Zero-order	Part Corr	VIF
B0	0.0004	0.1491		483	0.0483			
X1	-0.4518	0.2976	-0.4494	3220	0.3220	0.1687	-0.2192	4.3880
X2	2.2385	0.8839	2.2230	6905	0.6905	0.3242	0.3658	39.2285
X3	-0.9751	0.3966	-0.9700	6668	0.6668	0.1063	-0.3550	7.8630
X4	-0.9337	0.5275	-0.9274	4125	0.4125	0.3119	-0.2562	13.8991

samples w/at least 1 significant X: 7700 (0.770) after SIG model 6544 (0.654)

The most important differences in the results FOR THIS EXAMPLE are the "SEB" and "VIF" results. Note that all SEB values (except for B0) have increased due to the multicollinearity, as have the VIF values.

Other important results, of course, include the regression coefficients ("B" and "Beta") themselves, along with the number of times they were significant. Indeed, different predictors are significant more frequently before (X2 and X4) and after (X2 and X3) due to the multicollinearity introduced into the population, even though the pairwise relationships (zero-order correlations) between the predictors and the dependent variable have not changed.

An Example: Shrinkage and Sample Size



Population Parameters

# of Predictors: ○ 1 ○ 2 ○ 3 ● 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	0.0	1.0	0.000	0.994
X1:	0.0	1.0	0.000	0.992
X2:	0.0	1.0	0.000	0.993
X3:	0.0	1.0	-0.001	0.994
X4:	0.0	1.0	0.001	0.994

Monte Carlo Parameters

Sample Size: 42
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 10000
The pseudorandom generator needs an INTEGER value to get started:
Random Generator Seed: 7368179
☑ Automatically set new seed for successive simulations

Population Correlations (Rho)

	DV	X1	X2	X3
X1	0.239			
X2	0.085	0.052		
X3	0.389	0.207	0.227	
X4	0.186	0.130	0.910	0.157

rho² = 0.257

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0005	0.1390		515	0.0515			
X1	0.0960	0.1489	0.0952	992	0.0992	0.2347	0.0873	1.2000
X2	-0.6542	0.3571	-0.6498	4321	0.4321	0.0875	-0.2480	7.1754
X3	0.4071	0.1505	0.4044	7433	0.7433	0.3860	0.3661	1.2315
X4	0.7067	0.3536	0.7025	4953	0.4953	0.1878	0.2706	7.0478

samples w/at least 1 significant X: 8748 (0.875) after SIG model 7746 (0.775)

Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.5579	0.3223	0.2495	0.8554

Cross-validity R² (Rc²)	Precision Efficacy (Rc² / R²)
0.1564	0.3973

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	13.54	4	3.384	7997	0.7997
Residual	27.44	37	0.742		
TOTAL	40.97	41			

Expected R² if Null true
k/(n-1) = 0.0976

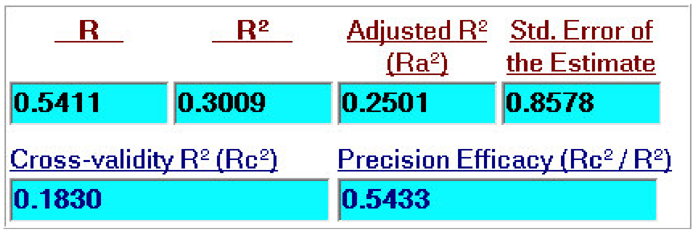
Finished 10000

Note that in this example, with a sample size of N = 42 (which provided statistical power for the model of approximately .80), shrinkage occurs from R^2 = .32 down to Adjusted R^2 = .25
or down to Cross-Validity R^2 = .16.

Recall that Adjusted R^2 represents the proportion of variance expected to be accounted for (explained) in the population if this particular regression model is used to predict scores in the population. It is generally considered a better SHRINKAGE estimate when **explanation** is the key purpose for the regression analysis.

Cross-validity R^2 represents the proportion of variance expected to be accounted for if this particular regression model is used in another sample of cases from the same population. It is generally considered a better SHRINKAGE estimate when **prediction** is the key purpose for the regression analysis.

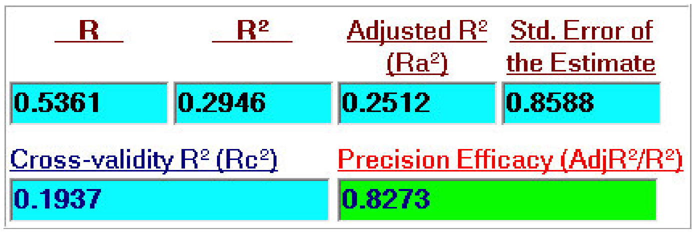
If we use N = 60 (based on 15 cases per predictor), shrinkage is less, but perhaps still too much.



R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.5411	0.3009	0.2501	0.8578

Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2 / R^2)
0.1830	0.5433

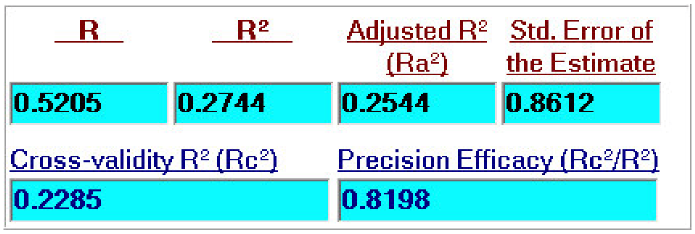
If we use, N = 70, which gives us some comfort that Precision Efficacy (using Adjusted R2) will be at least .80, shrinkage is even less.



R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.5361	0.2946	0.2512	0.8588

Cross-validity R^2 (Rc^2)	Precision Efficacy ($AdjR^2/R^2$)
0.1937	0.8273

If we use N = 150, which gives us comfort that Precision Efficacy (using Cross-validity R2) will be at least .80, reduces shrinkage even further.



R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.5205	0.2744	0.2544	0.8612

Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2/R^2)
0.2285	0.8198

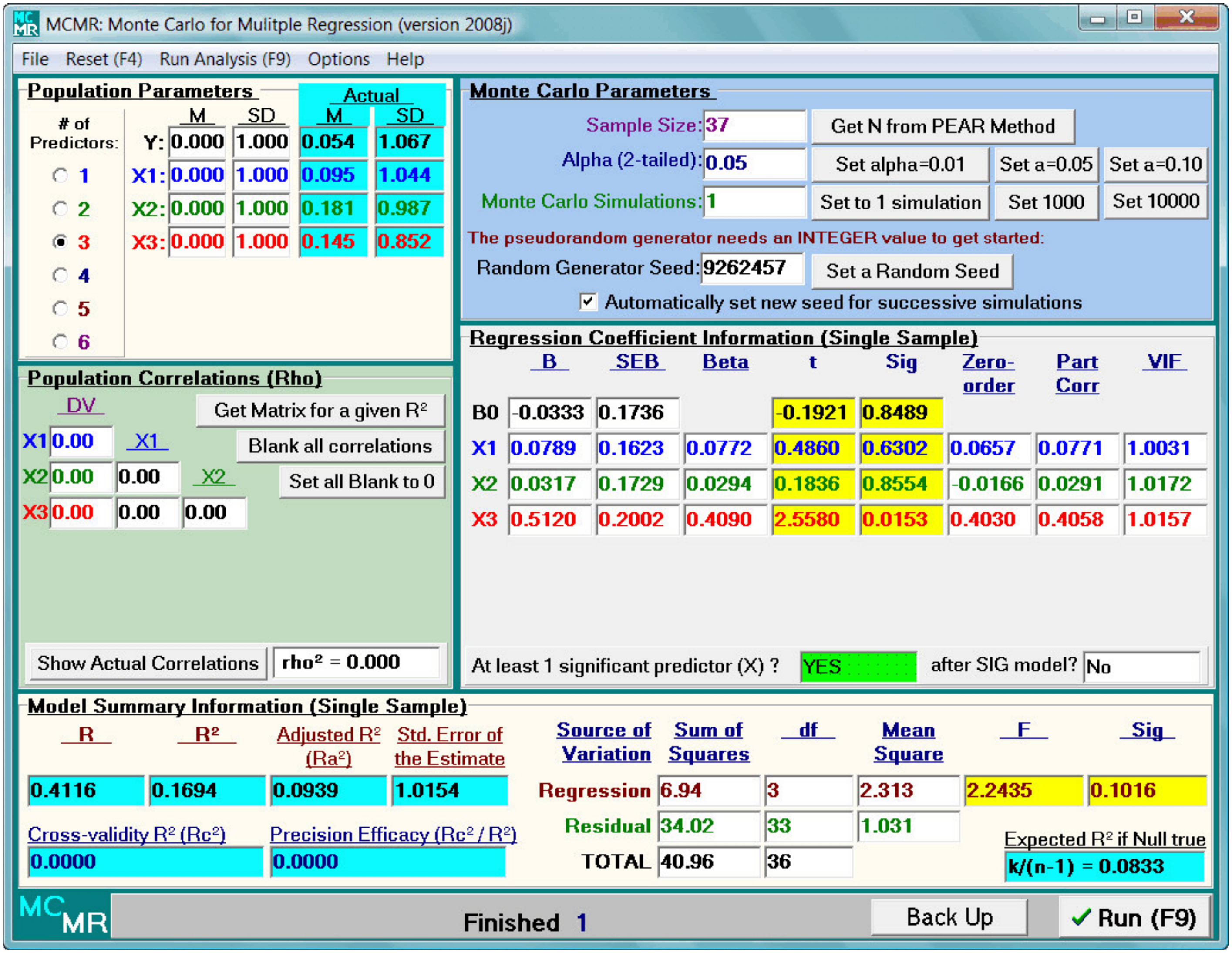
While there is no agreed-upon criterion for SHRINKAGE, several authors have recommended CROSS-VALIDATION as more appropriate methods for determining sample sizes than using statistical power (e.g., Algina & Keselman, 2000; Brooks & Barcikowski, 1999; Park & Dudycha, 1974; Stevens, 1996).

Note that there are also other methods that exist for calculating sample sizes in regression, including statistical power for the *t* tests of the regression coefficients and size of the confidence intervals for the regression coefficients (and therefore size of the standard errors of the regression coefficients).

There are many conventional rules (“rules of thumb”) that scholars have recommended over the years as well. These can all be tested and compared using the Monte Carlo method with the MCMR program.

Much more on the topic can be found in Brooks (1998).

An Example: Type I errors (and/or Statistical Power analyses)



Population Parameters

# of Predictors:		M	SD	Actual M	Actual SD
	Y:	0.000	1.000	0.054	1.067
○ 1	X1:	0.000	1.000	0.095	1.044
○ 2	X2:	0.000	1.000	0.181	0.987
● 3	X3:	0.000	1.000	0.145	0.852
○ 4					
○ 5					
○ 6					

Monte Carlo Parameters

Sample Size: 37

Alpha (2-tailed): 0.05

Monte Carlo Simulations: 1

The pseudorandom generator needs an INTEGER value to get started:

Random Generator Seed: 9262457

☑ Automatically set new seed for successive simulations

Population Correlations (Rho)

	DV	X1	X2
X1	0.00		
X2	0.00	0.00	
X3	0.00	0.00	0.00

$\rho^2 = 0.000$

Regression Coefficient Information (Single Sample)

	B	SEB	Beta	t	Sig	Zero-order	Part Corr	VIF
B0	-0.0333	0.1736		-0.1921	0.8489			
X1	0.0789	0.1623	0.0772	0.4860	0.6302	0.0657	0.0771	1.0031
X2	0.0317	0.1729	0.0294	0.1836	0.8554	-0.0166	0.0291	1.0172
X3	0.5120	0.2002	0.4090	2.5580	0.0153	0.4030	0.4058	1.0157

At least 1 significant predictor (X) ? YES after SIG model? No

Model Summary Information (Single Sample)

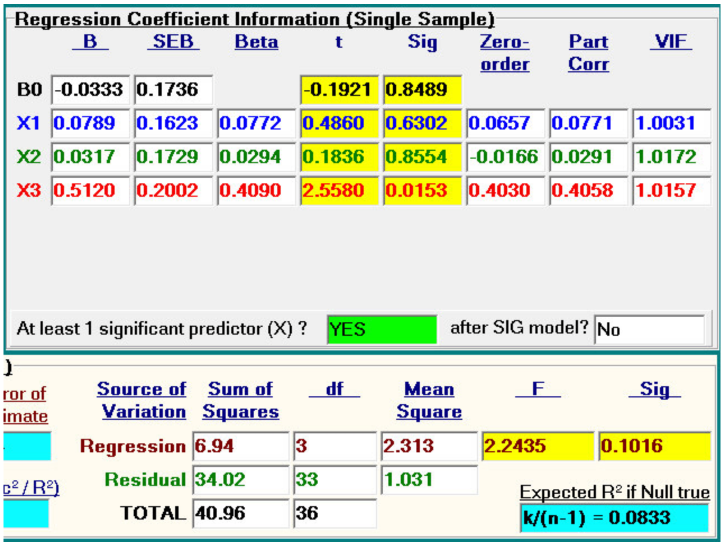
R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.4116	0.1694	0.0939	1.0154

Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2 / R^2)
0.0000	0.0000

Source of Variation	Sum of Squares	df	Mean Square	F	Sig
Regression	6.94	3	2.313	2.2435	0.1016
Residual	34.02	33	1.031		
TOTAL	40.96	36			

Expected R^2 if Null true $k/(n-1) = 0.0833$

Finished 1


Regression Coefficient Information (Single Sample)

	B	SEB	Beta	t	Sig	Zero-order	Part Corr	VIF
B0	-0.0333	0.1736		-0.1921	0.8489			
X1	0.0789	0.1623	0.0772	0.4860	0.6302	0.0657	0.0771	1.0031
X2	0.0317	0.1729	0.0294	0.1836	0.8554	-0.0166	0.0291	1.0172
X3	0.5120	0.2002	0.4090	2.5580	0.0153	0.4030	0.4058	1.0157

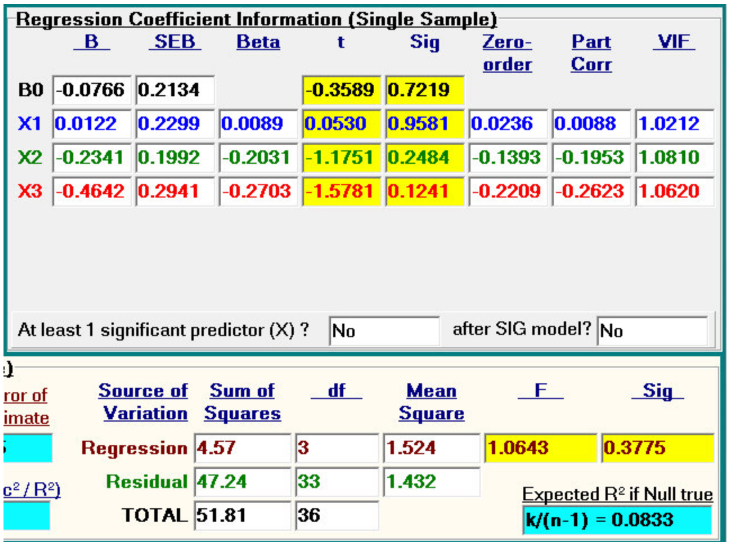
At least 1 significant predictor (X) ? YES after SIG model? No

Source of Variation	Sum of Squares	df	Mean Square	F	Sig
Regression	6.94	3	2.313	2.2435	0.1016
Residual	34.02	33	1.031		
TOTAL	40.96	36			

Expected R^2 if Null true $k/(n-1) = 0.0833$

We can run SINGLE SAMPLE analyses to show all the possible combinations of Type I errors that occur in multiple regression.

In this first example where all correlations are 0.0, one predictor (X3) is statistically significant, but the model is NOT statistically significant. Therefore, the count boxes show a GREEN YES for “At least 1 significant predictor (X)?” but a white NO for “after SIG model?”



Regression Coefficient Information (Single Sample)

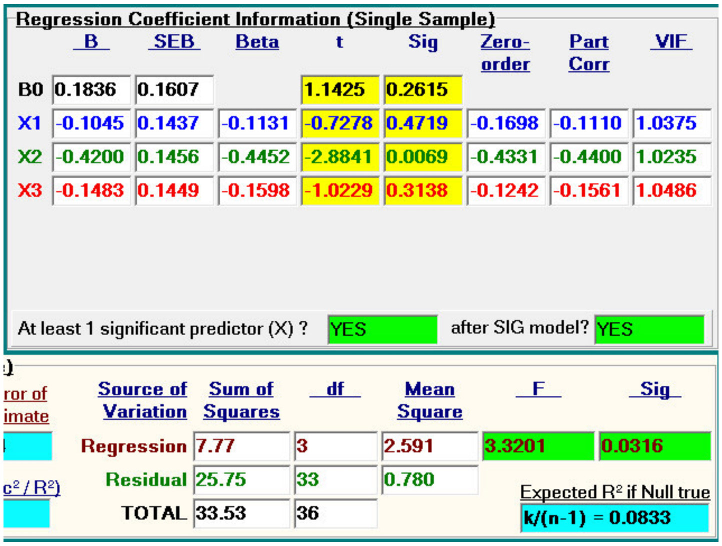
	B	SEB	Beta	t	Sig	Zero-order	Part Corr	VIF
B0	-0.0766	0.2134		-0.3589	0.7219			
X1	0.0122	0.2299	0.0089	0.0530	0.9581	0.0236	0.0088	1.0212
X2	-0.2341	0.1992	-0.2031	-1.1751	0.2484	-0.1393	-0.1953	1.0810
X3	-0.4642	0.2941	-0.2703	-1.5781	0.1241	-0.2209	-0.2623	1.0620

At least 1 significant predictor (X) ? No after SIG model? No

Source of Variation	Sum of Squares	df	Mean Square	F	Sig
Regression	4.57	3	1.524	1.0643	0.3775
Residual	47.24	33	1.432		
TOTAL	51.81	36			

Expected R² if Null true
k/(n-1) = 0.0833

In this second example where all correlations are 0.0, nothing was statistically significant. This is what we would expect most frequently when the Null Hypothesis is true.



Regression Coefficient Information (Single Sample)

	B	SEB	Beta	t	Sig	Zero-order	Part Corr	VIF
B0	0.1836	0.1607		1.1425	0.2615			
X1	-0.1045	0.1437	-0.1131	-0.7278	0.4719	-0.1698	-0.1110	1.0375
X2	-0.4200	0.1456	-0.4452	-2.8841	0.0069	-0.4331	-0.4400	1.0235
X3	-0.1483	0.1449	-0.1598	-1.0229	0.3138	-0.1242	-0.1561	1.0486

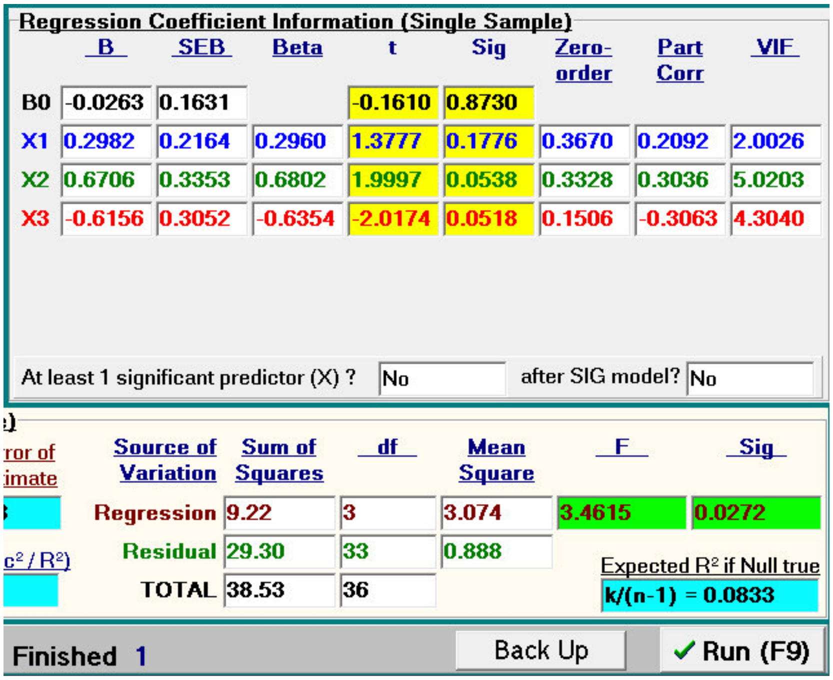
At least 1 significant predictor (X) ? YES after SIG model? YES

Source of Variation	Sum of Squares	df	Mean Square	F	Sig
Regression	7.77	3	2.591	3.3201	0.0316
Residual	25.75	33	0.780		
TOTAL	33.53	36			

Expected R² if Null true
k/(n-1) = 0.0833

In this third example where all correlations are 0.0, the overall regression model was statistically significant and at least one (here, exactly one, X2) predictor was statistically significant.

Note that different predictors are usually significant in different samples for Robustness (Type I error rate) analyses.



Regression Coefficient Information (Single Sample)

	B	SEB	Beta	t	Sig	Zero-order	Part Corr	VIF
B0	-0.0263	0.1631		-0.1610	0.8730			
X1	0.2982	0.2164	0.2960	1.3777	0.1776	0.3670	0.2092	2.0026
X2	0.6706	0.3353	0.6802	1.9997	0.0538	0.3328	0.3036	5.0203
X3	-0.6156	0.3052	-0.6354	-2.0174	0.0518	0.1506	-0.3063	4.3040

At least 1 significant predictor (X) ? No after SIG model? No

ror of imate

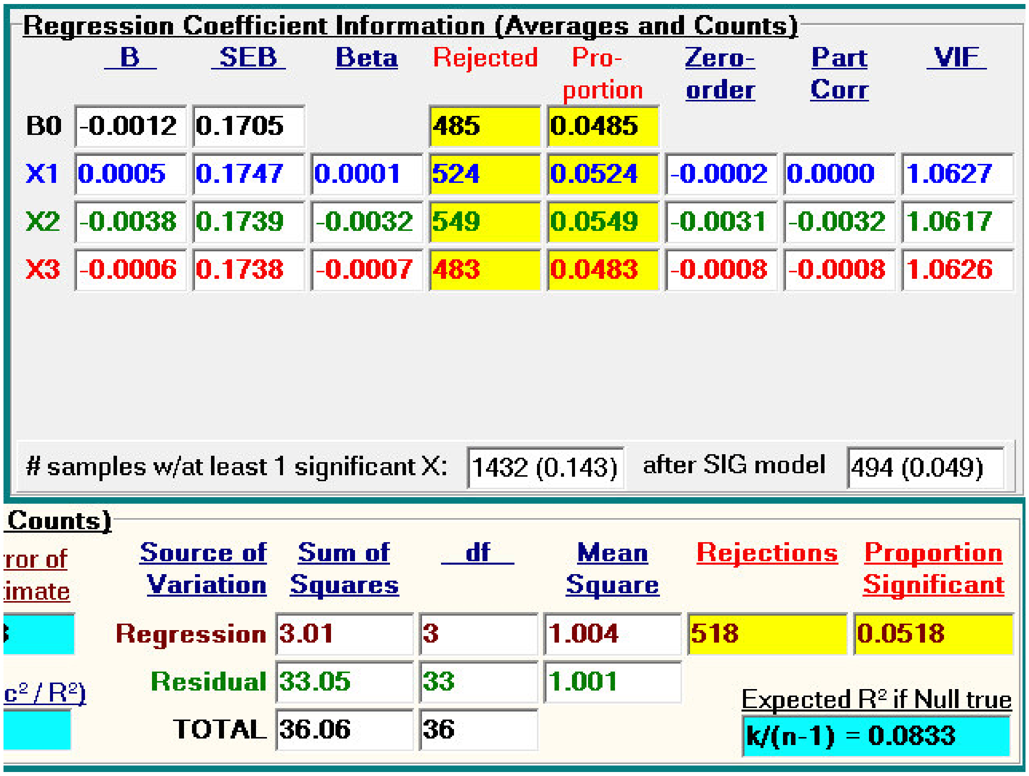
c²/R²)

Source of Variation	Sum of Squares	df	Mean Square	F	Sig
Regression	9.22	3	3.074	3.4615	0.0272
Residual	29.30	33	0.888		
TOTAL	38.53	36			

Expected R^2 if Null true $k/(n-1) = 0.0833$

NOTE: *This screen comes from an analysis with non-zero correlations, and therefore not a Type I error rate analysis.*

In this fourth example, the overall regression model was statistically significant, but NONE of the predictors was statistically significant. While this appears to be very rare when all correlations are 0.0 (a Type I error rate analysis), it occurs occasionally when the null hypothesis is not true.



Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Pro-portion	Zero-order	Part Corr	VIF
B0	-0.0012	0.1705		485	0.0485			
X1	0.0005	0.1747	0.0001	524	0.0524	-0.0002	0.0000	1.0627
X2	-0.0038	0.1739	-0.0032	549	0.0549	-0.0031	-0.0032	1.0617
X3	-0.0006	0.1738	-0.0007	483	0.0483	-0.0008	-0.0008	1.0626

samples w/at least 1 significant X: 1432 (0.143) after SIG model 494 (0.049)

Counts)

ror of imate

c²/R²)

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	3.01	3	1.004	518	0.0518
Residual	33.05	33	1.001		
TOTAL	36.06	36			

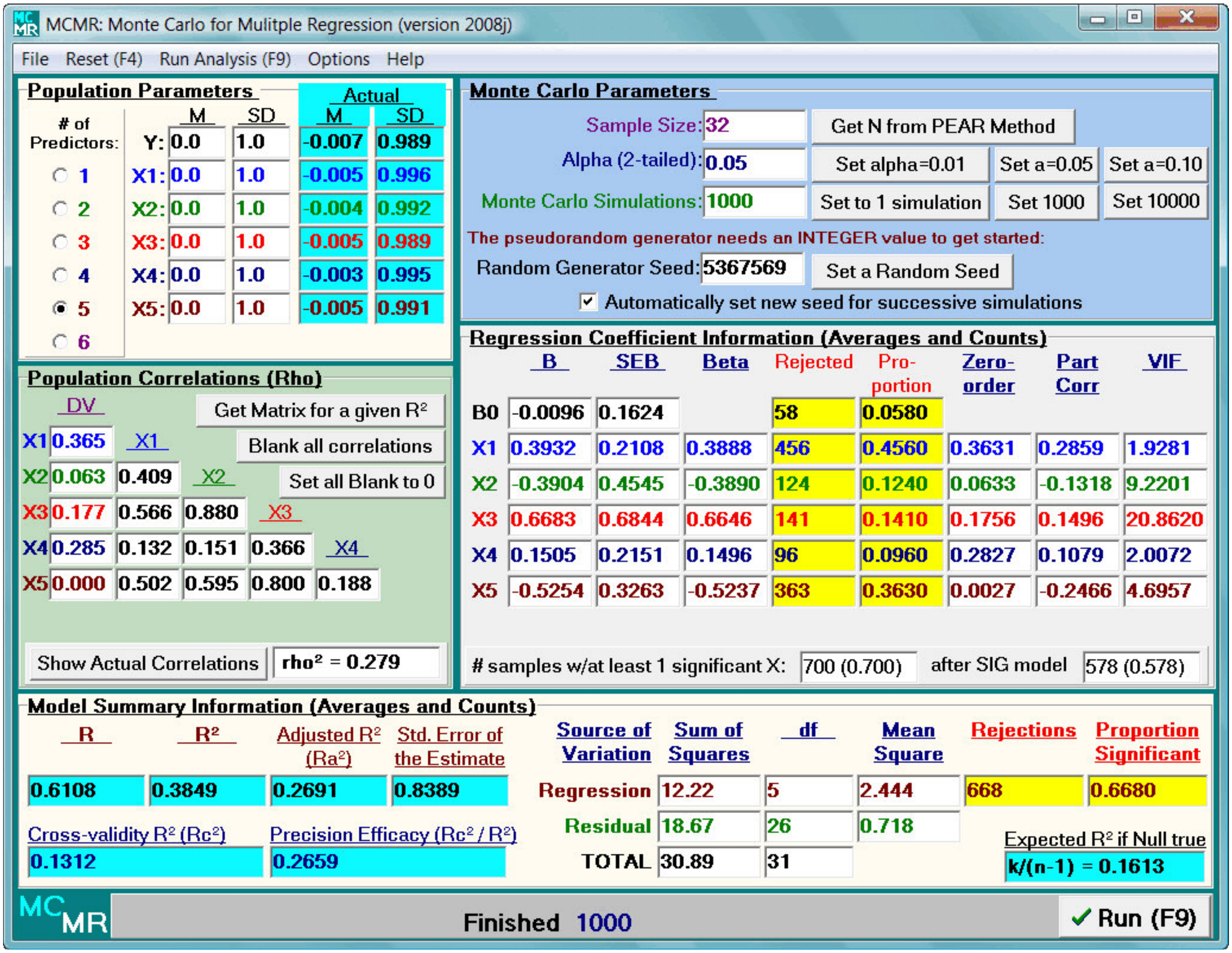
Expected R^2 if Null true $k/(n-1) = 0.0833$

Finally, after running through several samples to show students what a Type I error analysis is like, we can tell them that instead of us going one-by-one through these single samples and keeping track, we can just have the computer do it for us and run 10,000 samples all at once.

This screen shows the Monte Carlo results for 10,000 simulated samples. One can easily see the approximately .05 Type I error rate expected for all tests.

We can also discuss the idea of a "Protected F" test by reviewing the count boxes. Here, the proportion of simulated samples that had at least one statistically significant predictor FOLLOWING a statistically significant overall regression model is about .049 (5%). However, the proportion of samples that had any number of predictors that were statistically significant was about .14 (14%).

An Example: Suppressor Variables



Population Parameters

# of Predictors:		M	SD	Actual M	Actual SD
	Y:	0.0	1.0	-0.007	0.989
○ 1	X1:	0.0	1.0	-0.005	0.996
○ 2	X2:	0.0	1.0	-0.004	0.992
○ 3	X3:	0.0	1.0	-0.005	0.989
○ 4	X4:	0.0	1.0	-0.003	0.995
● 5	X5:	0.0	1.0	-0.005	0.991
○ 6					

Monte Carlo Parameters

Sample Size: 32
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 1000
The pseudorandom generator needs an INTEGER value to get started:
Random Generator Seed: 5367569
☑ Automatically set new seed for successive simulations

Population Correlations (Rho)

	DV	X1	X2	X3	X4
X1	0.365				
X2	0.063	0.409			
X3	0.177	0.566	0.880		
X4	0.285	0.132	0.151	0.366	
X5	0.000	0.502	0.595	0.800	0.188

rho² = 0.279

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0096	0.1624		58	0.0580			
X1	0.3932	0.2108	0.3888	456	0.4560	0.3631	0.2859	1.9281
X2	-0.3904	0.4545	-0.3890	124	0.1240	0.0633	-0.1318	9.2201
X3	0.6683	0.6844	0.6646	141	0.1410	0.1756	0.1496	20.8620
X4	0.1505	0.2151	0.1496	96	0.0960	0.2827	0.1079	2.0072
X5	-0.5254	0.3263	-0.5237	363	0.3630	0.0027	-0.2466	4.6957

samples w/at least 1 significant X: 700 (0.700) after SIG model 578 (0.578)

Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.6108	0.3849	0.2691	0.8389

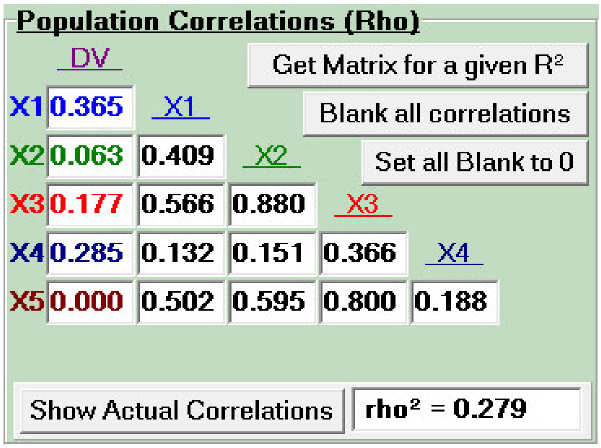
Cross-validity R² (Rc²)	Precision Efficacy (Rc² / R²)
0.1312	0.2659

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	12.22	5	2.444	668	0.6680
Residual	18.67	26	0.718		
TOTAL	30.89	31			

Expected R² if Null true
k/(n-1) = 0.1613

Finished 1000

If we arbitrarily set a population correlation matrix in which one predictor has zero (0.0) correlation with the dependent variable (DV) but has non-zero correlation with the other predictors, we can examine suppressor relationships.



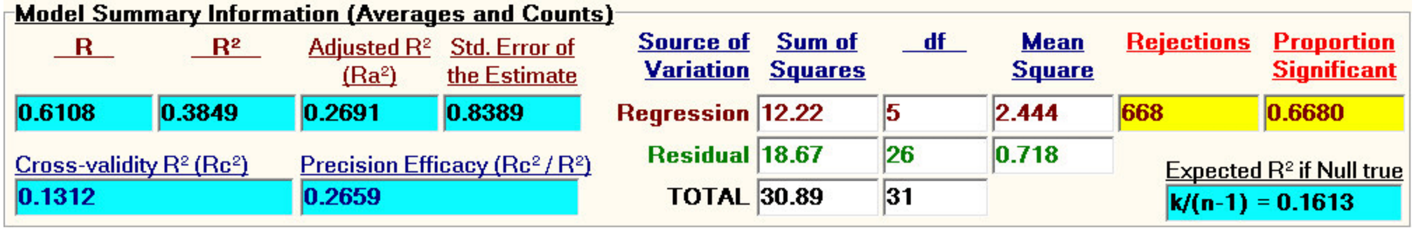
Population Correlations (Rho)

	DV	X1	X2	X3	X4
X1	0.365				
X2	0.063	0.409			
X3	0.177	0.566	0.880		
X4	0.285	0.132	0.151	0.366	
X5	0.000	0.502	0.595	0.800	0.188

rho² = 0.279

You can see a little better the correlations here.

Note the population multiple rho^2 for this correlation matrix is .279



Model Summary Information (Averages and Counts)

R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.6108	0.3849	0.2691	0.8389

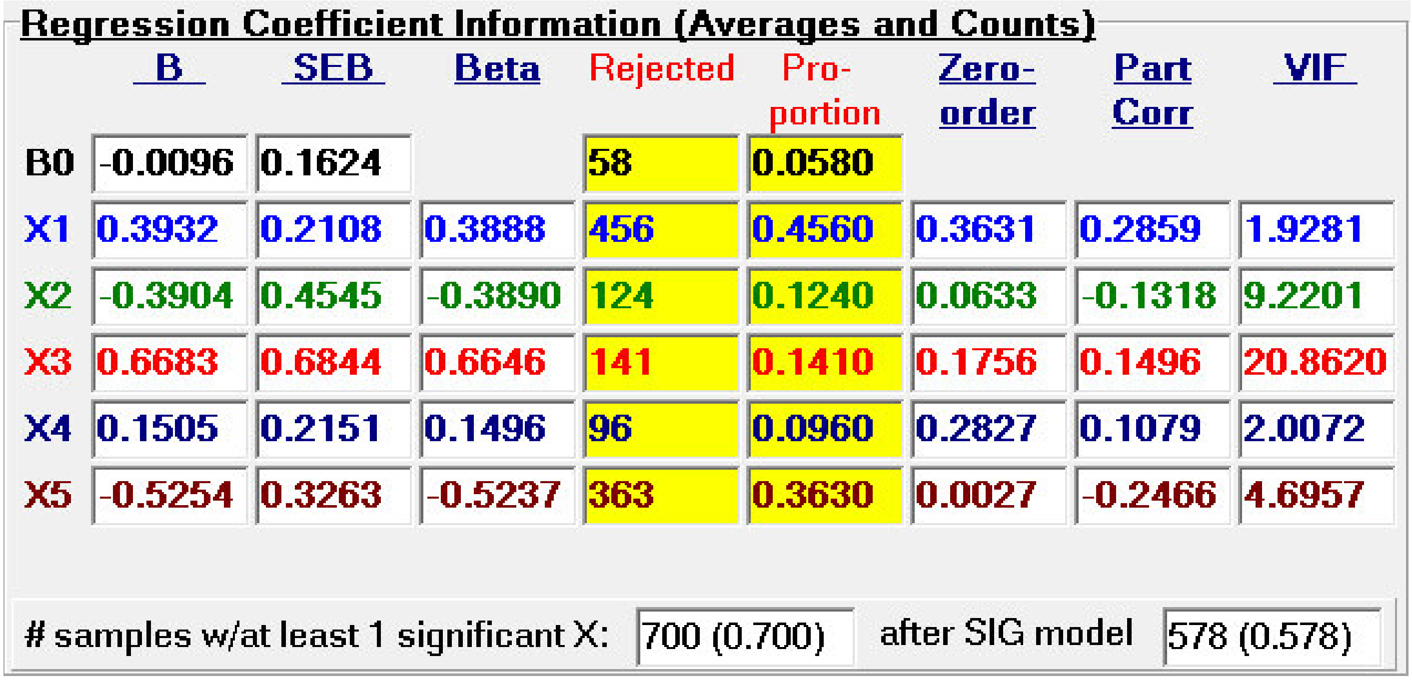
Cross-validity R^2 (Rc^2): 0.1312

Precision Efficacy (Rc^2 / R^2): 0.2659

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	12.22	5	2.444	668	0.6680
Residual	18.67	26	0.718		
TOTAL	30.89	31			

Expected R^2 if Null true: $k/(n-1) = 0.1613$

We have an R2 value of .38 for this analysis.

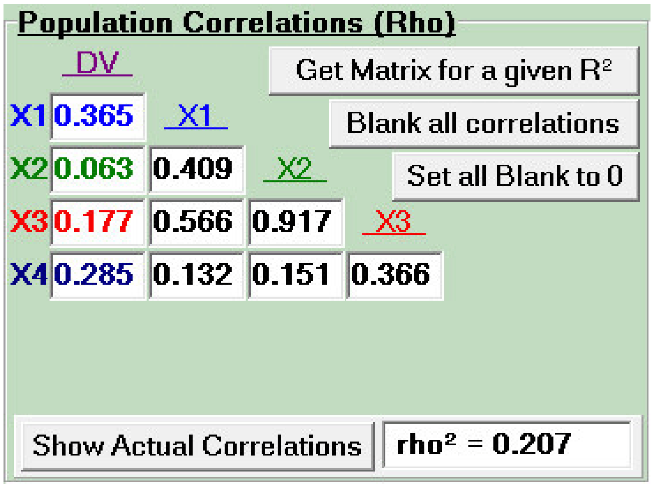


Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0096	0.1624		58	0.0580			
X1	0.3932	0.2108	0.3888	456	0.4560	0.3631	0.2859	1.9281
X2	-0.3904	0.4545	-0.3890	124	0.1240	0.0633	-0.1318	9.2201
X3	0.6683	0.6844	0.6646	141	0.1410	0.1756	0.1496	20.8620
X4	0.1505	0.2151	0.1496	96	0.0960	0.2827	0.1079	2.0072
X5	-0.5254	0.3263	-0.5237	363	0.3630	0.0027	-0.2466	4.6957

samples w/at least 1 significant X: 700 (0.700) after SIG model 578 (0.578)

Note the VIF is high for X3, not the variable with 0.0 correlation with the dependent variable (which is X5). However, there is a strong correlation between X3 and X5.



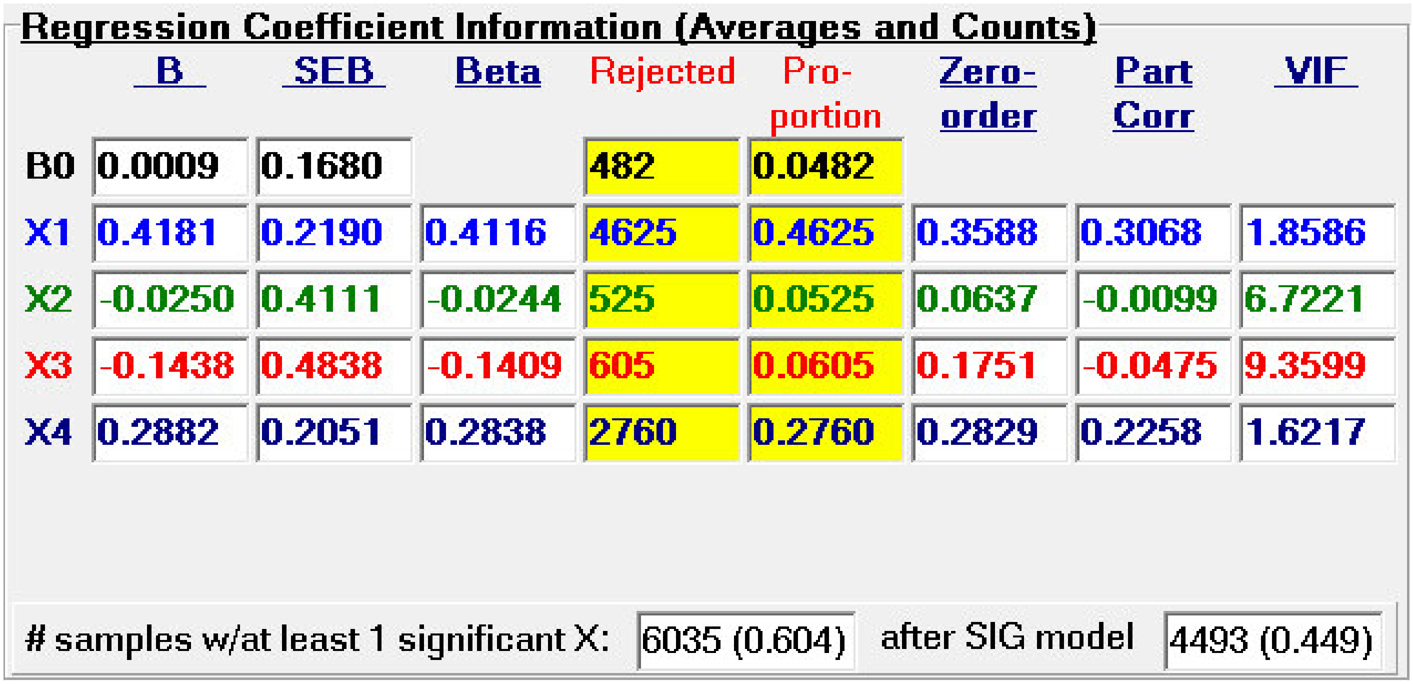
Population Correlations (Rho)

	DV	X1	X2	X3
X1	0.365			
X2	0.063	0.409		
X3	0.177	0.566	0.917	
X4	0.285	0.132	0.151	0.366

$rho^2 = 0.207$

If we remove X5 from the analysis in an effort to remove the multicollinearity (because among the predictors, it has very little correlation with Y), we would have this correlation matrix.

Note that rho^2 is lower without X5 EVEN THOUGH it had no correlation with the Dependent Variable !!

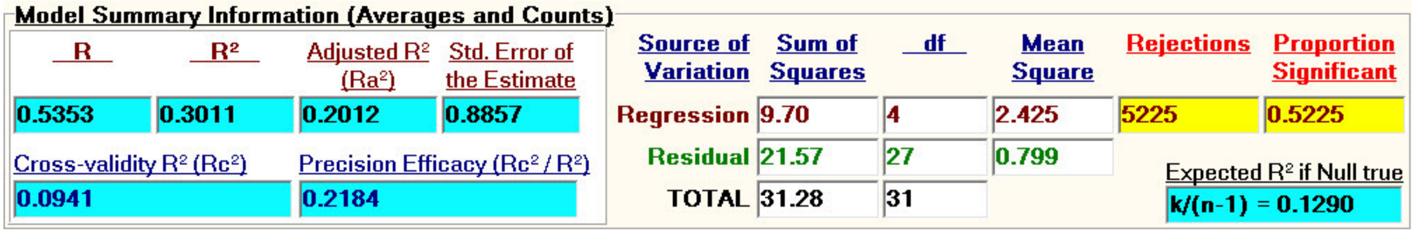


Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Pro-portion	Zero-order	Part Corr	VIF
B0	0.0009	0.1680		482	0.0482			
X1	0.4181	0.2190	0.4116	4625	0.4625	0.3588	0.3068	1.8586
X2	-0.0250	0.4111	-0.0244	525	0.0525	0.0637	-0.0099	6.7221
X3	-0.1438	0.4838	-0.1409	605	0.0605	0.1751	-0.0475	9.3599
X4	0.2882	0.2051	0.2838	2760	0.2760	0.2829	0.2258	1.6217

samples w/at least 1 significant X: 6035 (0.604) after SIG model 4493 (0.449)

Note that multicollinearity has been removed (as evidenced by all VIF < 10).



Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.5353	0.3011	0.2012	0.8857

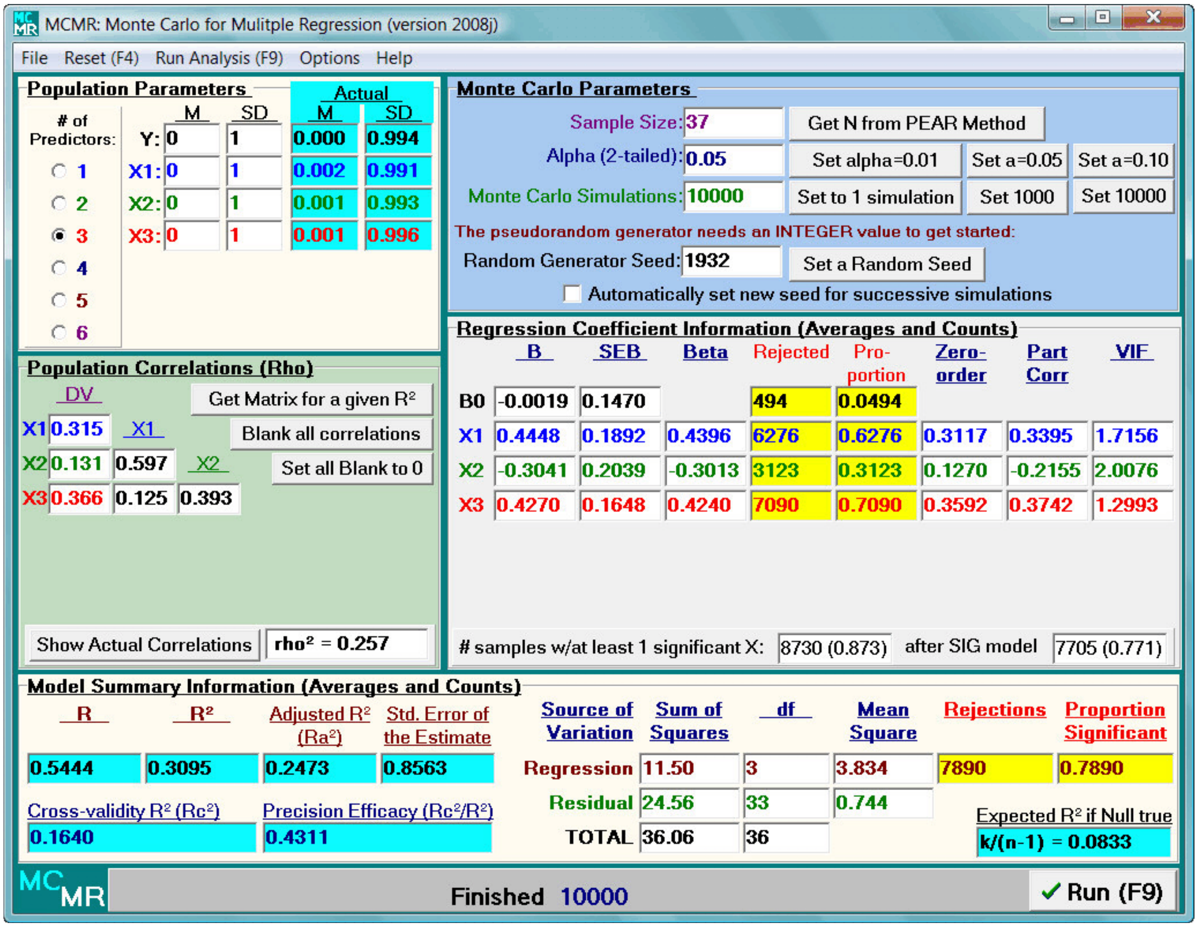
Cross-validity R² (Rc²)	Precision Efficacy (Rc² / R²)
0.0941	0.2184

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	9.70	4	2.425	5225	0.5225
Residual	21.57	27	0.799		
TOTAL	31.28	31			

Expected R² if Null true
k/(n-1) = 0.1290

However, the relationship in terms of R^2 is not as high as it was with the apparently useless predictor (.38 then versus .30 now).

An Example: Impact of Means and Standard Deviations on Regression Results



Population Parameters

# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	0	1	0.000	0.994
X1:	0	1	0.002	0.991
X2:	0	1	0.001	0.993
X3:	0	1	0.001	0.996

Population Correlations (Rho)

	DV	X1	X2
X1	0.315		
X2	0.131	0.597	
X3	0.366	0.125	0.393

rho² = 0.257

Monte Carlo Parameters

Sample Size: 37

Alpha (2-tailed): 0.05

Monte Carlo Simulations: 10000

The pseudorandom generator needs an INTEGER value to get started:

Random Generator Seed: 1932

☐ Automatically set new seed for successive simulations

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0019	0.1470		494	0.0494			
X1	0.4448	0.1892	0.4396	6276	0.6276	0.3117	0.3395	1.7156
X2	-0.3041	0.2039	-0.3013	3123	0.3123	0.1270	-0.2155	2.0076
X3	0.4270	0.1648	0.4240	7090	0.7090	0.3592	0.3742	1.2993

samples w/at least 1 significant X: 8730 (0.873) after SIG model 7705 (0.771)

Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.5444	0.3095	0.2473	0.8563

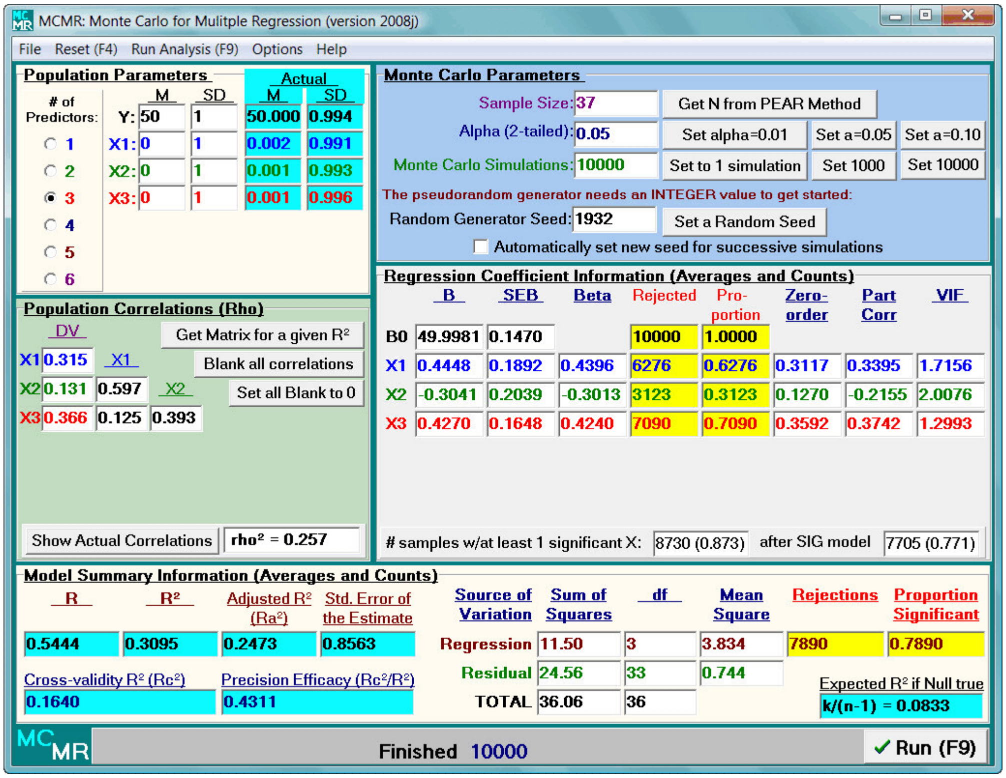
Cross-validity R² (Rc²): 0.1640

Precision Efficacy (Rc²/R²): 0.4311

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	11.50	3	3.834	7890	0.7890
Residual	24.56	33	0.744		
TOTAL	36.06	36			

Expected R² if Null true: k/(n-1) = 0.0833

Finished 10000



Population Parameters

# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	50	1	50.000	0.994
X1:	0	1	0.002	0.991
X2:	0	1	0.001	0.993
X3:	0	1	0.001	0.996

Population Correlations (Rho)

	DV	X1	X2
X1	0.315		
X2	0.131	0.597	
X3	0.366	0.125	0.393

rho² = 0.257

Monte Carlo Parameters

Sample Size: 37

Alpha (2-tailed): 0.05

Monte Carlo Simulations: 10000

The pseudorandom generator needs an INTEGER value to get started:

Random Generator Seed: 1932

☐ Automatically set new seed for successive simulations

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	49.9981	0.1470		10000	1.0000			
X1	0.4448	0.1892	0.4396	6276	0.6276	0.3117	0.3395	1.7156
X2	-0.3041	0.2039	-0.3013	3123	0.3123	0.1270	-0.2155	2.0076
X3	0.4270	0.1648	0.4240	7090	0.7090	0.3592	0.3742	1.2993

samples w/at least 1 significant X: 8730 (0.873) after SIG model 7705 (0.771)

Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.5444	0.3095	0.2473	0.8563

Cross-validity R² (Rc²): 0.1640

Precision Efficacy (Rc²/R²): 0.4311

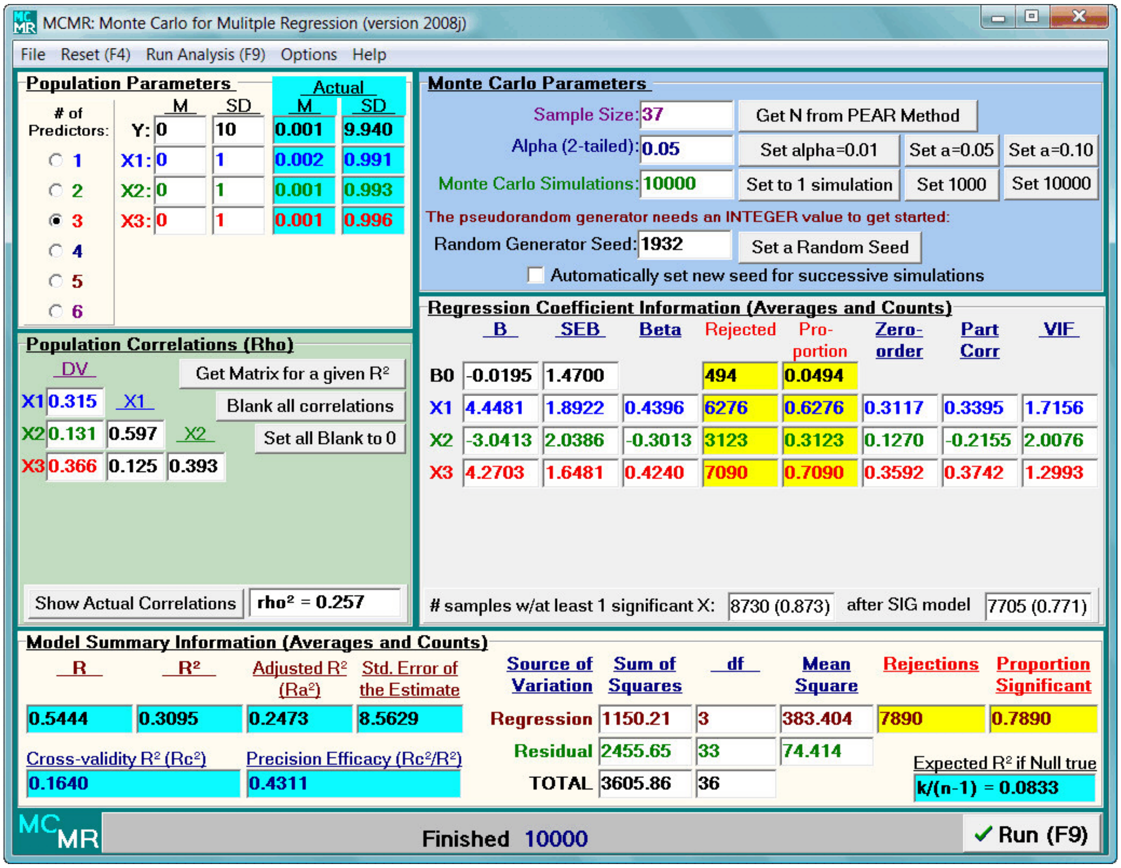
Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	11.50	3	3.834	7890	0.7890
Residual	24.56	33	0.744		
TOTAL	36.06	36			

Expected R² if Null true: k/(n-1) = 0.0833

Finished 10000

The important thing to notice as we change from all standardized data (above), to a Dependent Variable Mean of 50 (while standard deviation remains 1.0) is that only the CONSTANT B0 and its statistical significance changed.

NOTHING ELSE changed !!



Population Parameters

\# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	0	10	0.001	9.940
X1:	0	1	0.002	0.991
X2:	0	1	0.001	0.993
X3:	0	1	0.001	0.996

Monte Carlo Parameters

Sample Size: 37

Alpha (2-tailed): 0.05

Monte Carlo Simulations: 10000

The pseudorandom generator needs an INTEGER value to get started:

Random Generator Seed: 1932

☐ Automatically set new seed for successive simulations

Population Correlations (Rho)

	DV	X1	X2
X1	0.315		
X2	0.131	0.597	
X3	0.366	0.125	0.393

$rho^2 = 0.257$

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-0.0195	1.4700		494	0.0494			
X1	4.4481	1.8922	0.4396	6276	0.6276	0.3117	0.3395	1.7156
X2	-3.0413	2.0386	-0.3013	3123	0.3123	0.1270	-0.2155	2.0076
X3	4.2703	1.6481	0.4240	7090	0.7090	0.3592	0.3742	1.2993

\# samples w/at least 1 significant X: 8730 (0.873) after SIG model 7705 (0.771)

Model Summary Information (Averages and Counts)

R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.5444	0.3095	0.2473	8.5629

Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2/R^2)
0.1640	0.4311

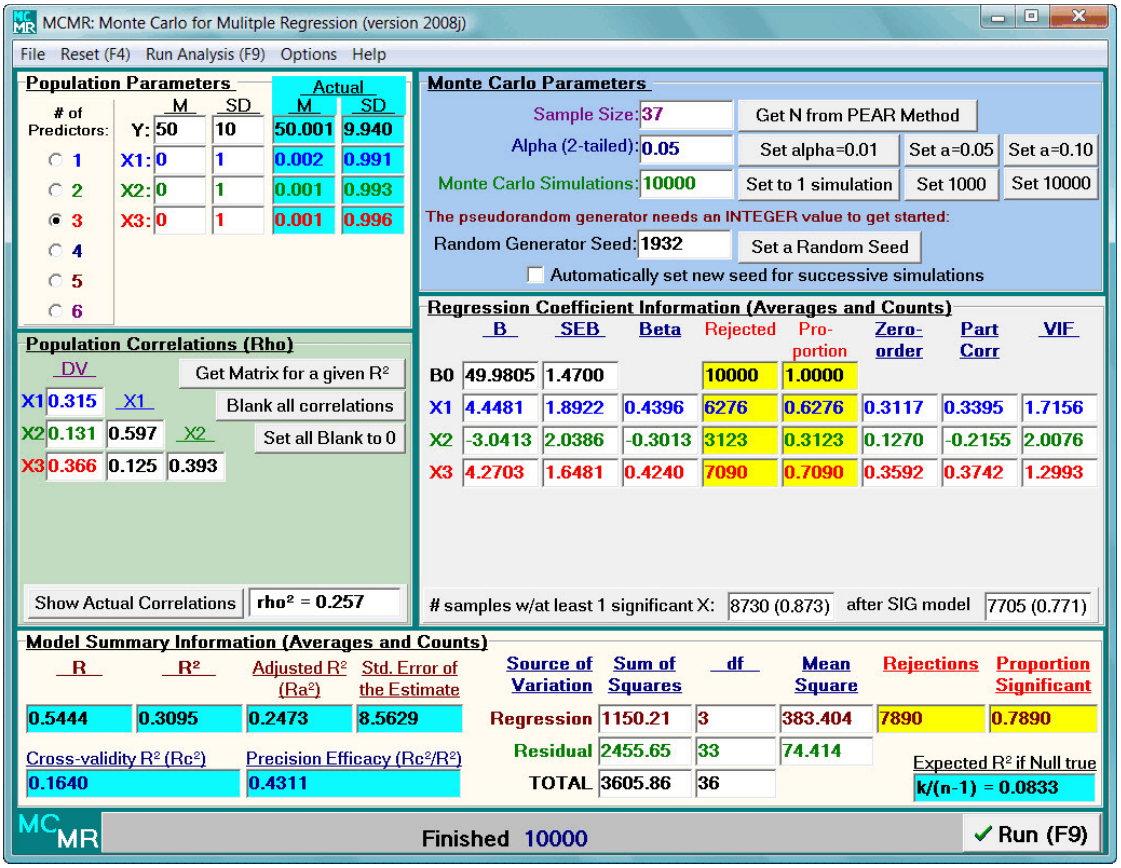
Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	1150.21	3	383.404	7890	0.7890
Residual	2455.65	33	74.414		
TOTAL	3605.86	36			

Expected R^2 if Null true $k/(n-1) = 0.0833$

Finished 10000

However, when the Dependent Variable Mean is 0.0, but the Standard Deviation changes to 10.0, several things change, most notably the regression coefficients and their significance and the SUMS OF SQUARES.

But none of the other important model information changed (e.g., R^2, F, rejections, Beta, VIF).



Population Parameters

\# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	50	10	50.001	9.940
X1:	0	1	0.002	0.991
X2:	0	1	0.001	0.993
X3:	0	1	0.001	0.996

Monte Carlo Parameters

Sample Size: 37

Alpha (2-tailed): 0.05

Monte Carlo Simulations: 10000

The pseudorandom generator needs an INTEGER value to get started:

Random Generator Seed: 1932

☐ Automatically set new seed for successive simulations

Population Correlations (Rho)

	DV	X1	X2
X1	0.315		
X2	0.131	0.597	
X3	0.366	0.125	0.393

$rho^2 = 0.257$

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	49.9805	1.4700		10000	1.0000			
X1	4.4481	1.8922	0.4396	6276	0.6276	0.3117	0.3395	1.7156
X2	-3.0413	2.0386	-0.3013	3123	0.3123	0.1270	-0.2155	2.0076
X3	4.2703	1.6481	0.4240	7090	0.7090	0.3592	0.3742	1.2993

\# samples w/at least 1 significant X: 8730 (0.873) after SIG model 7705 (0.771)

Model Summary Information (Averages and Counts)

R	R^2	Adjusted R^2 (Ra^2)	Std. Error of the Estimate
0.5444	0.3095	0.2473	8.5629

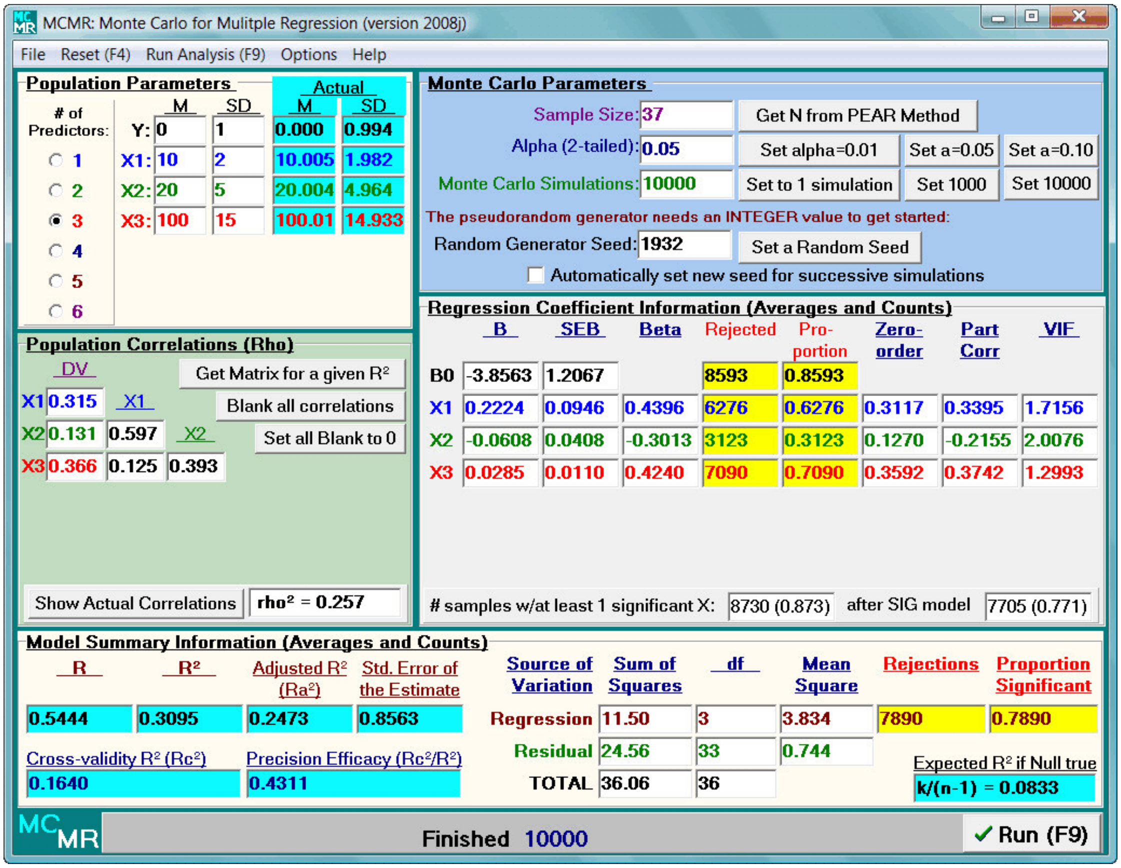
Cross-validity R^2 (Rc^2)	Precision Efficacy (Rc^2/R^2)
0.1640	0.4311

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	1150.21	3	383.404	7890	0.7890
Residual	2455.65	33	74.414		
TOTAL	3605.86	36			

Expected R^2 if Null true $k/(n-1) = 0.0833$

Finished 10000

Changing both the Mean and the Standard Deviation combines these previous two results. That is, all the information EXCEPT B0 remains the same as the previous example. But now with the Y mean at 50, B0 changed to match (and is significant more often).



Population Parameters

# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	0	1	0.000	0.994
X1:	10	2	10.005	1.982
X2:	20	5	20.004	4.964
X3:	100	15	100.01	14.933

Population Correlations (Rho)

	DV	X1	X2
X1	0.315		
X2	0.131	0.597	
X3	0.366	0.125	0.393

rho² = 0.257

Monte Carlo Parameters

Sample Size: 37
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 10000
The pseudorandom generator needs an INTEGER value to get started:
Random Generator Seed: 1932
☐ Automatically set new seed for successive simulations

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	-3.8563	1.2067		8593	0.8593			
X1	0.2224	0.0946	0.4396	6276	0.6276	0.3117	0.3395	1.7156
X2	-0.0608	0.0408	-0.3013	3123	0.3123	0.1270	-0.2155	2.0076
X3	0.0285	0.0110	0.4240	7090	0.7090	0.3592	0.3742	1.2993

samples w/at least 1 significant X: 8730 (0.873) after SIG model 7705 (0.771)

Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.5444	0.3095	0.2473	0.8563

Cross-validity R² (Rc²): 0.1640
Precision Efficacy (Rc²/R²): 0.4311

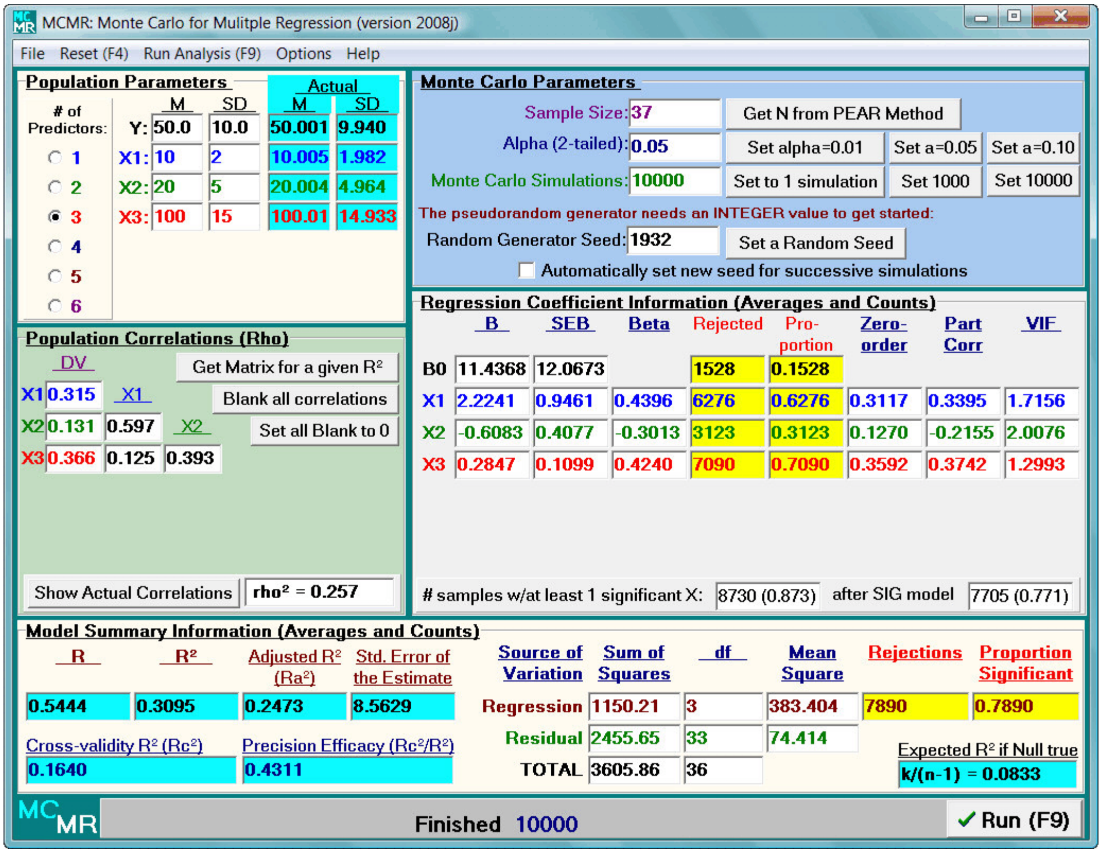
Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	11.50	3	3.834	7890	0.7890
Residual	24.56	33	0.744		
TOTAL	36.06	36			

Expected R² if Null true: k/(n-1) = 0.0833

Finished 10000

If we change the predictor Means and Standard Deviations, but leave the Dependent Variable Y standardized, you can see several differences — most notably in the regression coefficients.

The "Sum of Squares" values have returned to what they were in the first example.



Population Parameters

# of Predictors: ○ 1 ○ 2 ● 3 ○ 4 ○ 5 ○ 6

	M	SD	Actual M	Actual SD
Y:	50.0	10.0	50.001	9.940
X1:	10	2	10.005	1.982
X2:	20	5	20.004	4.964
X3:	100	15	100.01	14.933

Population Correlations (Rho)

	DV	X1	X2
X1	0.315		
X2	0.131	0.597	
X3	0.366	0.125	0.393

rho² = 0.257

Monte Carlo Parameters

Sample Size: 37
Alpha (2-tailed): 0.05
Monte Carlo Simulations: 10000
The pseudorandom generator needs an INTEGER value to get started:
Random Generator Seed: 1932
☐ Automatically set new seed for successive simulations

Regression Coefficient Information (Averages and Counts)

	B	SEB	Beta	Rejected	Proportion	Zero-order	Part Corr	VIF
B0	11.4368	12.0673		1528	0.1528			
X1	2.2241	0.9461	0.4396	6276	0.6276	0.3117	0.3395	1.7156
X2	-0.6083	0.4077	-0.3013	3123	0.3123	0.1270	-0.2155	2.0076
X3	0.2847	0.1099	0.4240	7090	0.7090	0.3592	0.3742	1.2993

samples w/at least 1 significant X: 8730 (0.873) after SIG model 7705 (0.771)

Model Summary Information (Averages and Counts)

R	R²	Adjusted R² (Ra²)	Std. Error of the Estimate
0.5444	0.3095	0.2473	8.5629

Cross-validity R² (Rc²): 0.1640
Precision Efficacy (Rc²/R²): 0.4311

Source of Variation	Sum of Squares	df	Mean Square	Rejections	Proportion Significant
Regression	1150.21	3	383.404	7890	0.7890
Residual	2455.65	33	74.414		
TOTAL	3605.86	36			

Expected R² if Null true: k/(n-1) = 0.0833

Finished 10000

Finally, if everything changes, the regression coefficients all change, but note that all the MODEL summary information and the CORRELATION information remains the same.

Means and Standard Deviations have not impact on the decisions regarding the Null Hypotheses for either coefficients or the model, nor on the interpretations of the value of the predictors or the model.

References

Algina, J., & Keselman, H. J. (2000). Cross-validation sample sizes. *Applied Psychological Measurement*, *24*, 173–179.

Algina, J., & Olejnik, S. (2000). Determining sample size for accurate estimation of the squared multiple correlation coefficient. *Multivariate Behavioral Research*, *35* (*1*), 119-137.

Brooks, G. P. (1998a, October). *Precision Efficacy Analysis for Regression*. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL.

Brooks, G. P. (1998b). *Precision efficacy analysis for regression: Development and justification of a new sample size method for multiple linear regression*. Unpublished dissertation, Ohio University, Athens.

Brooks, G. P., & Barcikowski, R. S. (1996). Precision power and its application to the selection of regression sample sizes. *Mid-Western Educational Researcher*, *9*(*4*), 10-17.

Browne, M. W. (1975). Predictive validity of a linear regression equation. *British Journal of Mathematical and Statistical Psychology*, *28*, 79-87.

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

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

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

Herzberg, P. A. (1969). The parameters of cross-validation. *Psychometrika Monograph Supplement*, *34* (2, Pt. 2).

Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.

L'Ecuyer, P. (1988). Efficient and portable combined random number generators. *Communications of the ACM*, *31*, 742-749, 774.

Lord, F. M. (1950). *Efficiency of prediction when a regression equation from one sample is used in a new sample* (Research Bulletin No. 50-40). Princeton, NJ: Educational Testing Service.

Nash, J. C. (1990). *Compact numerical methods for computers: Linear algebra and function minimisation* (2nd ed.). New York: Adam Hilger.

Nicholson, G. E. (1960). Prediction in future samples. In I. Olkin et al. (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 322-330). Palo Alto, CA: Stanford.

Park, C. N., & Dudycha, A. L. (1974). A cross-validation approach to sample size determination for regression models. *Journal of the American Statistical Association*, *69*, 214-218.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical recipes in Pascal: The art of scientific computing*. New York: Cambridge University.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.). New York: Cambridge University.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlations: A clarification. *Psychological Bulletin*, *85*, 1348-1351.

Rozeboom, W. W. (1981). The cross-validational accuracy of sample regressions. *Journal of Educational Statistics*, *6*, 179-198.

Sawyer, R. (1982). Sample size and the accuracy of predictions made from multiple regression equations. *Journal of Educational Statistics*, *7*, 91-104.

Stein, C. (1960). Multiple regression. In I. Olkin et al. (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp.425-443). Palo Alto, CA: Stanford University.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Uhl, N., & Eisenberg, T. (1970). Predicting shrinkage in the multiple correlation coefficient. *Educational and Psychological Measurement*, *30*, 487-489.

Send correspondence to: Gordon P. Brooks
Ohio University
Email: brooksg@ohio.edu

Multiple Linear Regression Viewpoints
Information for Contributors

Multiple Linear Regression Viewpoints (*MLRV*) is a publication sponsored by the American Educational Research Association's Special Interest Group on Multiple Linear Regression: The General Linear Model (MLR/GLM SIG). It is published twice a year to facilitate communication among professionals who focus their research on the theory, application, or teaching of multiple linear regression models and/or the general linear model. Manuscripts submitted to *MLRV* should conform to the language, style, and format of the Publication Manual of the American Psychological Association (5th ed., 2001). Three copies (two blind) of a doubled spaced manuscript (including equations, footnotes, quotes, and references) of approximately 25 pages in length, a 100 word abstract, and an IBM formatted diskette with the manuscript formatted in WordPerfect or Word should be submitted to one of the editors listed below.

Mathematical and Greek symbols should be clear and concise. All figures and diagrams must be photocopy-ready for publication. Manuscripts will be anonymously peer reviewed by two editorial board members. Author identifying information should appear on the title page of only one submitted manuscript. The review process will take approximately 3 months. A letter acknowledging receipt of the manuscript will be sent to the first author, and upon review, a letter indicating the peer review decision will be sent to the first author. Potential authors are encouraged to contact the editors to discuss ideas for contributions or determine if their manuscript is suitable for publication in *MLRV*.

EDITOR INFORMATION

Randall E. Schumacker, Editor *MLRV*
College of Education
P.O. Box 870231
Carmichael Hall 316
The University of Alabama
Tuscaloosa, AL 35487-0231
(205) 348-6062
rschumacker@ua.edu

T. Mark Beasley, Associate Editor
Department of Biostatistics
School of Public Health
309E Ryals Public Health Bldg.
University of Alabama, Birmingham
Birmingham, AL 35294
(205) 975-4957
mbeasley@uab.edu

ORDER INFORMATION

Cynthia Campbell, MLR/GLM SIG Executive Secretary
Department of Educational Technology, Research & Assessment
Northern Illinois University
DeKalb, IL 60115-2854
ccampbell@niu.edu

Check out our website at: http://mlrv.ua.edu/

POSTMASTER: Send address changes to:
Cynthia Campbell, MLR/GLM SIG Executive Secretary
Department of Educational Technology, Research & Assessment
Northern Illinois University
DeKalb, IL 60115-2854

Multiple Linear Regression Viewpoints (ISSN 0195-7171) is published by the AERA Special Interest Group on Multiple Linear Regression: General Linear Model through the **University of Alabama at Birmingham** and **Northern Illinois University**.