DigitalCommons@WayneState

Journal of Modern Applied Statistical Methods

Volume 1 | Issue 2 Article 44

11-1-2002

Determining Predictor Importance In Multiple Regression Under Varied Correlational And Distributional Conditions

Tiffany A. Whittaker
The University of Texas at Austin, tiffany.whittaker@mail.utexas.edu

Rachel T. Fouladi
University of Texas M.D. Anderson Cancer Center at Houston

Natasha J. Williams
University of Texas at Austin

Journal of Modern Applied Statistical Methods
Fall 2002, Vol. 1, No 2, 354-366

1538 – 9472/02/$30.00

Determining Predictor Importance In Multiple Regression Under Varied Correlational And Distributional Conditions

Tiffany A. Whittaker
University of Texas at Austin

Rachel T. Fouladi
University of Texas
M.D. Anderson Cancer Center
at Houston

Natasha J. Williams
University of Texas at Austin

This study examines the performance of eight methods of predictor importance under varied correlational and distributional conditions. The proportion of times a method correctly identified the dominant predictor was recorded. Results indicated that the new methods of importance proposed by Budescu (1993) and Johnson (2000) outperformed commonly used importance methods.

Key words: Multiple Regression; Predictor Importance; Relative Importance; Multicollinearity.

Introduction

One of the most common statistical techniques used today is Multiple Regression (MR) Analysis (Neter, Kutner, Nachtsheim, & Wasserman, 1996). Once the predictors are selected for the MR model, researchers typically wish to establish the relative importance of the predictors when predicting the dependent variable. According to Healy (1990), the most typical request of statistical consultants when conducting MR analyses is to determine the relative importance of the predictor variables in the model, with the key focus on the question: Of all the predictors in the MR model, which one influences the criterion variable the most?

According to Kruskal (1984), there are two motives as to why relative importance is so meaningful to researchers: 1) technological motives and 2) scientific motives. The technological motive is produced from the hopes of implementing change that is effective and economical. For example, "what should we attend to first in trying to reduce cancer deaths, improve education, maintain our systems of highways, increase productivity growth, etc." (Kruskal, 1984, p. 39). The scientific motive is produced from the attempt to increase one's basic understanding of some phenomenon with no concern of implementing immediate change. For example, "which variables should we examine in our next experiment or survey…since we never have the resources to examine all?" (Kruskal, 1984, p. 39). Regardless of the motive, predictor importance is of great concern when conducting MR analyses.

Consider p predictors, $x_1 \ldots x_p$, of the criterion variable y. When the predictor variables in the MR model are perfectly uncorrelated, relative importance can simply be determined from the squared value of the zero-order correlations between the criterion and each of the predictors ($\rho_{yx_j}{}^2$, $j = 1 \ldots p$) which, in that case, sum to the model's squared multiple correlation (Budescu, 1993):

$$\rho_{y.x_1 \ldots x_p}{}^2 = \sum_{j=1}^{p} \rho_{yx_j}{}^2 . \qquad (1)$$

Tiffany A. Whittaker and Natasha J. Williams, Department of Educational Psychology, University of Texas at Austin. Rachel T. Fouladi, Department of Behavioral Science, University of Texas MD Anderson Cancer Center in Houston. Tiffany A. Whittaker's email address: Twhittaker@mail.utexas.edu. Natasha J. Williams' email address: Tashwill@aol.com. Rachel T. Fouladi's email: Rfouladi@mail.mdanderson.org.

Thus, the relative contribution of each predictor may be expressed in terms of percentages, as can be seen from the following equation (Lindeman, Merenda, & Gold, 1980, p. 119):

$$\text{Percentage Contribution} = 100\,\frac{\rho_{yx_j}^{\;2}}{\rho_{y.x_1\ldots x_p}^{\;2}}, \qquad (2)$$

and this can be interpreted as the percentage of total variance in the criterion accounted for by a predictor. However, when the predictors are correlated with each other, which is normally the case, the above relationship is no longer viable. This is because part of a predictor's contribution becomes a shared contribution with one or more of the other predictor variables with which it happens to be correlated (Lindeman et al., 1980).

Many techniques have been proposed to assess the relative importance of predictors in ordinary least squares (OLS) MR models, with little consensus on which method is best employed (for reviews, see Budescu, 1993; Darlington, 1968). Proposed methods to determine the importance of the *j*th predictor of *y* include: 1) the squared zero-order correlation between the criterion variable and the predictor, $\rho_{yx_j}^{\;2}$; 2) the standardized regression coefficient for the predictor in the *p*-predictor MR model, $\beta_j^{\,*}$; 3) the *t*-statistic for the test of the regression coefficient in the *p*-predictor MR model, t_j; 4) the product of the standardized regression coefficient for a predictor and its zero-order correlation with the criterion (Pratt, 1987), $\beta_j{*}\rho_{yx_j}$; 5) the squared partial correlation of the criterion variable and the predictor, $\rho_{yx_j.x_1\ldots x_{j-1}x_{j+1}\ldots x_p}^{\;2}$; and 6) the squared semi-partial correlation of the criterion variable and the predictor, $\rho_{y(x_j.x_1\ldots x_{j-1}x_{j+1}\ldots x_p)}^{\;2}$ (c.f., Darlington, 1968; Budescu, 1993; Johnson, 2000). All of these methods of determining predictor importance provide the same information when the predictors are not intercorrelated. However, the information they provide is not equivalent when the predictors are correlated (Darlington, 1968).

The lack of consensus as to which importance method to use is understandable when one considers the differences between these methods, the most visible difference being the definition of importance adopted when using these various methods (Budescu, 1993). For instance, the squared value of the zero-order correlation between the criterion and the predictor, $\rho_{yx_j}^{\;2}$, is the proportion of variance in the criterion accounted for by the predictor (Cohen & Cohen, 1975). Thus, it only illustrates a predictor's direct effect on the criterion (Budescu, 1993). Standardized regression coefficients, $\beta_j^{\,*}$, are interpreted as the amount of change that occurs in the criterion variable for each standard deviation change in a predictor variable while holding all other predictors in the model constant (Bring, 1994).

Hence, a predictor's importance is dependent upon its own contribution to the model, which is contingent upon the other predictors' contributions (Budescu, 1993). The *t*-values associated with the estimates of the coefficients for the predictors are computed to test the null hypothesis that each population regression coefficient in the model is equal to zero ($\beta_j = 0$) (Lindeman et al., 1980). When computing a *t*-value for a predictor, it represents the increase in the model's squared multiple correlation when adding the predictor to the MR model after all the additional $p - 1$ predictors have already been included in the MR model (Bring, 1994). Hence, a predictor's importance is dependent upon its own contribution to the model, which is contingent upon the other predictors' contributions. The product of the standardized regression coefficient for a predictor and its zero-order correlation with the criterion (Pratt, 1987), $\beta_j{*}\rho_{yx_j}$, represents both a predictor's total effect ($\beta_j{*}$) and direct effect (ρ_{yx_j}). The squared partial correlation, $\rho_{yx_j.x_1\ldots x_{j-1}x_{j+1}\ldots x_p}^{\;2}$, and the predictor's "usefulness" (i.e., the squared semipartial correlation), $\rho_{y(x_j.x_1\ldots x_{j-1}x_{j+1}\ldots x_p)}^{\;2}$, (Darlington, 1968) can be perceived as the proportion of variance in the criterion that can be explained by each predictor variable contingent upon the other predictors' contributions (Budescu, 1993). Evidently, the definition of importance varies widely from method to method. Accordingly, these methods can often lead to different conclusions as to the relative importance of the same predictor variables (Budescu, 1993).

Dominance Analysis

Budescu (1993) recently suggested a new method, called Dominance Analysis, that identifies predictor importance while accounting for a predictor's direct, partial, and total effect. Where x_i and x_j are a pair of predictors in the original set of p predictors, and x_h is any subset of the remaining p–2 predictors, x_i "weakly dominates" x_j , if the following relationships among squared multiple correlations hold for all possible x_h:

$$\rho_{y.x_i x_h}^{\;2} \geq \rho_{y.x_j x_h}^{\;2} \qquad (3)$$

or

$$(\rho_{y.x_i x_h}^{\;2} - \rho_{y.x_h}^{\;2}) \quad \geq \quad (\rho_{y.x_j x_h}^{\;2} - \rho_{y.x_h}^{\;2}), \qquad (4)$$

where $\rho_{y.x_i x_h}^{\;2}$ is the squared multiple correlation of the model which includes predictor x_i and the remaining predictors, x_h, while excluding predictor x_j. After establishing pairwise "dominance or equality" for each $p(p–1)/2$ x_i x_j pairings, the next step is to compute

$$C_{x_i}^{(k)} = \sum(\rho_{y.x_i x_h}^{\;2} - \rho_{y.x_h}^{\;2}) \,/\, m \qquad (5)$$

for each variable x_i across all m models with k + 1 predictors (x_i and $k = 0\ldots p - 1$ variables), where x_h is any possible subset of k predictors with x_i excluded and $m = \binom{p-1}{k}$. Lastly, Budescu advises the computation of

$$C_{x_i} = \sum_{k=0}^{p-1} C_{x_i}^{(k)} \,/\, p\,, \qquad (6)$$

which provides a meaningful decomposition of the p-predictor model's squared multiple correlation.

Johnson's Index

Johnson (2000) critiqued Budescu's method and noted that computations are tedious and require more time as the number of predictor variables in the model increases (Johnson, 2000). Johnson (2000) suggested an alternative method that yields similar results with less computation, extending the work of Gibson (1962), Johnson (1966), and Green Carroll, and DeSarbo (1978). Without loss of generality, let **X** be an $N \times p$ full-rank matrix of predictor scores in standard score form, and **y** be the $p \times 1$ criterion score vector also in standard score form. Singular value decomposition yields **X** = **PΔQ'**, where **P** consists of eigenvectors of **XX'**, **Q** consists of eigenvectors of **X'X**, and Δ is the diagonal matrix with the square roots of corresponding eigenvalues on the diagonal. Let **Z**= **PQ'**, which yields a best-fitting (minimum sum of squared residuals) set of orthogonal variables to **X**. Let the regression of **y** on **Z** yield the vector of regression weights $\boldsymbol{\beta}_Z^*$, and the regression of **X** on **Z** yield the matrix of regression weights **Λ***. Using the notation,

$$\Lambda^{*[2]} = \left\| \lambda_{jk}^2 \right\| \qquad (7)$$

and

$$\boldsymbol{\beta}^{*[2]} = \left\| \beta_{Z\,jk}^{*\,2} \right\|, \qquad (8)$$

Johnson's index for each predictor's relative importance is obtained from the elements of $\boldsymbol{\varepsilon} = \Lambda^{*[2]} \boldsymbol{\beta}^{*[2]}$, which when summed yield the original p-predictor model's squared multiple correlation (Johnson, 2000).

Using an actual data set, Johnson compared his method (ε) with seven other measures of importance. These seven measures included the following: 1) the squared zero-order correlation between the criterion and the predictor; 2) the squared value of the standardized regression coefficient; 3) the product of the standardized regression coefficient for a predictor and its zero-order correlation with the criterion, $\beta_j^* \rho_{yx}$; 4) the t-statistic associated with a predictor; 5) the squared value of the standardized partial regression coefficient from regressing the criterion on the orthogonal predictors (Gibson, 1962); 6) Green, Carroll, and DeSarbo's (1978) relative weight measure (δ_j^2); and 7) Budescu's (1993) Dominance Analysis method (C_{x_i}). Relative weights for various predictor variables were calculated using each of the different importance methods. Johnson concluded that his method (ε), Budescu's (1993) method (C_{x_i}), and Green et al.'s (1978) method (δ_j^2) were comparable in terms of the relative weights assigned to the predictors and that these methods are the most efficient in obtaining the indirect and direct effects of the predictors on the criterion variable.

Johnson further examined the efficiency of his method by comparing it to both Budescu's (1993) C_{x_i} and Green et. al.'s (1978) δ_j^2 across various regression models. Using 31 different sets of data (both authentic and simulated), Johnson calculated the relative importance weights assigned by each of the three different methods. The number of predictors in the MR model varied from 3 to 10, and the mean correlation among predictor variables varied from .10 to .70. Using Budescu's (1993) method as the standard, mean differences between the weights were calculated across the predictor variables. Johnson found that the mean difference between his method and Budescu's (1993) method was smaller than the mean difference between Budescu's method and Green et. al.'s (1978) method. The mean differences between the relative importance weights were not related to the number of predictors in the model, but were related to the mean correlation among predictors in the model. Thus, Johnson's and Budescu's methods demonstrated similar findings as to the relative weights assigned, but as the mean correlation between the predictor variables increased, so did the differences between Johnson's and Budescu's (1993) methods. Still, as the mean correlation among predictors increased, Green et al.'s (1978) method deviated more from Budescu's (1993) method than Johnson's method. Johnson attributed the deviation between his method and Budescu's (1993) method to the fact that regression coefficients become unstable under conditions of multicollinearity, suggesting that both measures may generate questionable results under these conditions. Nevertheless, Johnson (2000) did not report which method performed the best in terms of correctly identifying the known dominant or most important predictor. In addition, results were not reported with respect to the performance of the predictor importance methods under various distributional conditions, such as multivariate nonnormality.

Normality of predictor and criterion variables is not an assumption of MR, however, nonnormality of predictor and criterion variables may create nonnormality in the error (residual) distributions, which is an assumption of MR. A violation of this assumption affects the validity of significance tests, such as *t*-tests, and increases the sample to sample variance of the regression coefficients. These effects are both due to the increase in the standard errors for the regression coefficients which occurs when the errors are nonnormally distributed (Hamilton, 1992).

Therefore, this study seeks to compare the performance of the new importance methods (i.e., Johnson's and Budescu's methods) to the other proposed measures of predictor importance in terms of identifying the known, correct dominant predictor. In addition, the current study will investigate the performance of these methods under a range of sample and distributional conditions using simulated data as well as a sample data set.

Methodology

Monte Carlo Study

A Monte Carlo simulation experiment was first conducted to compare methods of predictor importance under conditions of normality and nonnormality in the predictors and criterion, homogenous correlations among predictors, and heterogeneous correlations between predictors and the criterion. Data were generated from multivariate normal and nonnormal populations using the Headrick and Sawilowsky (1999) approach, which has been proposed as an alternative to other methods used for generating skewed and kurtotic distributions (e.g., Vale & Maurelli, 1983).

The correct identification of the known dominant predictor was examined under the following conditions:

Methods of Importance. Eight methods of importance were investigated. These included: 1) the squared zero-order correlation between the criterion variable and the predictor, $\rho_{yx_j}^{\;2}$; 2) the standardized regression coefficient for the predictor in the *p*-predictor MR model, β^*_j; 3) the *t*-statistic for the test of the regression coefficient in the *p*-predictor MR model, t_j; 4) the product of the standardized regression coefficient for a predictor and its zero-order correlation with the criterion (Pratt, 1987), $\beta_j * \rho_{yx_j}$; 5) the squared partial correlation of the criterion variable and the predictor, $\rho_{yx_j \cdot x_1 \ldots x_{j-1} x_{j+1} \ldots x_p}{}^2$; 6) the squared semi-partial correlation of the criterion variable and the predictor, $\rho_{y(x_j \cdot x_1 \ldots x_{j-1} x_{j+1} \ldots x_p)}{}^2$; 7) Budescu's (1993) dominance measure, C_{x_j}, and 8) Johnson's (2000) Epsilon index, ε_j.

Correlations among predictors. To represent low, moderate, and high multicollinearity levels among the predictor variables, data were generated from populations where predictors were homogeneously intercorrelated where the magnitude of the correlations equaled .10, .40, or .70.

Correlations between dominant predictor and criterion. Data were from populations where the predictors were heterogeneously correlated with the criterion. To establish known dominance of a predictor, the most important predictor correlated .40 or .60 with the criterion while the correlation between the additional predictors and the criterion equaled .30.

Distribution type. Data were distributed from both multivariate normal and nonnormal distributions, where the levels of skew and kurtosis for the predictors and the criterion were (sk, ku): (0, 0) for a normal distribution, (0, 6) for a symmetric and heavy-tailed distribution, or (2, 6) for an asymmetric and heavy-tailed distribution. These levels of skew and kurtosis were selected to compare the performance of the importance methods under the normal distribution as well as under some commonly encountered nonnormal distributions (Micceri, 1989).

Number of predictors, p. To represent a low, moderate, and high number of predictors in the MR model, data were from p-variate multinormal and multi-nonnormal populations, where p equaled 4, 6, or 8.

Sample size, n. To represent a wide range of sample sizes similar to those that may be encountered in the health, behavioral, and social scienes where extremely small as well as large sample studies are conducted, data were generated at specific ratios of sample size to number of variables, where n was either $2p$, $4p$, $10p$, $20p$, or $40p$.

The six factors were fully crossed and each condition was replicated 1,000 times. Under each condition, the number of times that the correct predictor was identified as dominant was recorded.

Results

A six-way factorial ANOVA [8 (methods of importance) × 3 (correlations among predictors) × 2 (correlations between dominant predictor and criterion) × 3 (distribution type) x 3 (number of predictors) × 5 (sample size)], with repeated measures on the importance methods, was performed on the hit rates. However, only a maximum of three-way interactions was investigated.

Four-way and five-way interactions were not investigated because separate ANOVAs for each importance method indicated that the three-way ANOVA models accounted for more than 90% of the variance in the hit rates (R^2 ranged from .93 to .96). Because differential performance of the importance methods was the focus of the current research, only the interactions between the repeated measures factor (importance method) and the additional between-subjects factors were examined, as well as the main effect for importance method.

To control for Type I error, only those interactions with the repeated measures factor that obtained a significance level less than .001 were examined. These interactions consisted of the following and are discussed in this order: Importance Method × Correlation Between Dominant Predictor and Criterion × Sample Size; Importance Method × Correlation Among Predictors × Sample Size; Importance Method × Correlation Among Predictors; Importance Method × Sample Size. The Least Significant Difference (LSD) test was used for post hoc multiple comparisons. Again, to control for Type I error, only the pairwise differences that obtained a significance level less than .001 were examined.

Importance Method × Correlation Between Dominant Predictor and Criterion × Sample Size. The ANOVA indicated a significant interaction between importance method, correlation between dominant predictor and criterion, and sample size, $F(28, 840) = 2.20$, $p <$

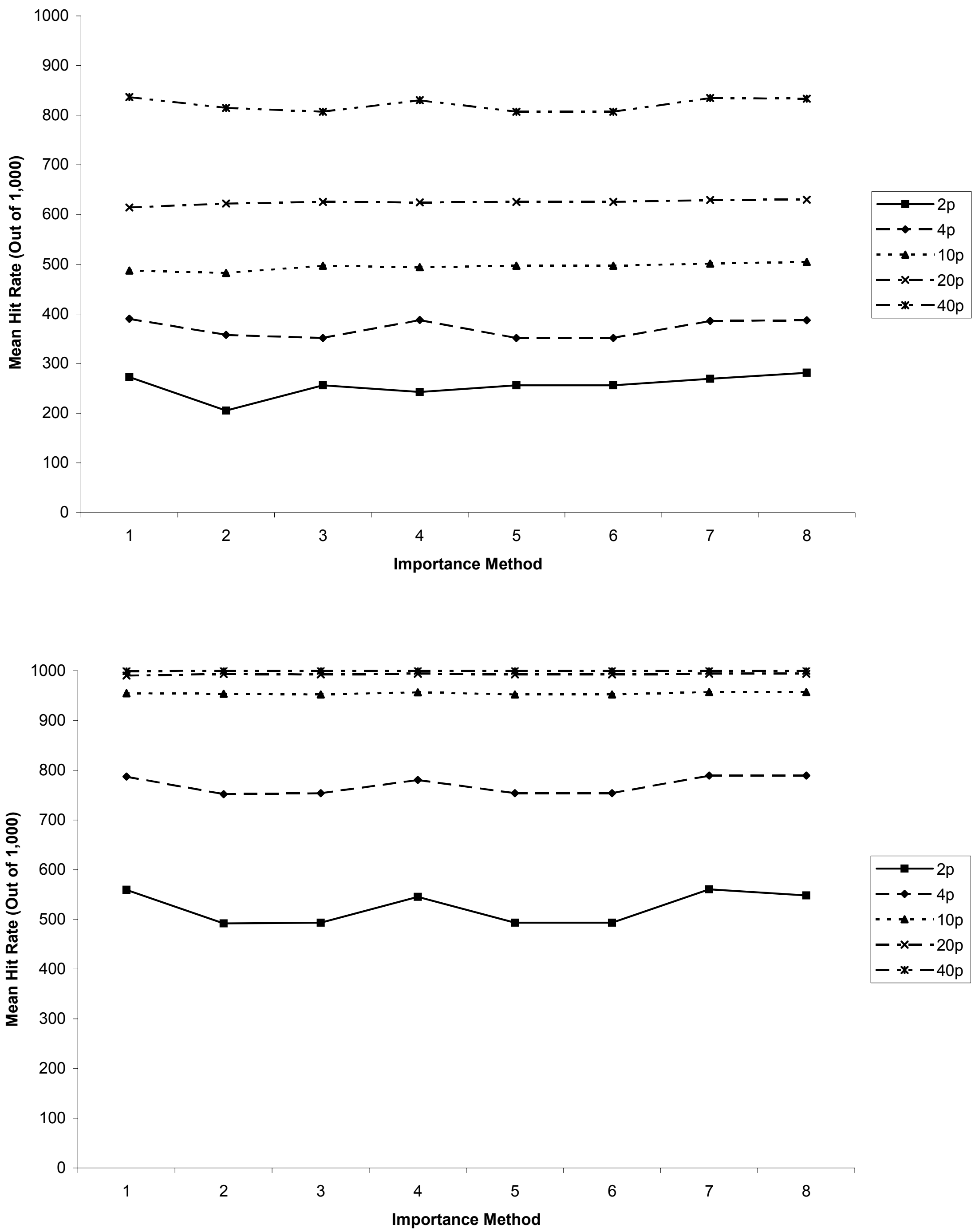
.001 (η^2 = .07). Post-hoc tests indicated that when the correlation between dominant predictor and criterion was low (.40) and sample size was small (2*p*), Budescu's method and Johnson's ε_j method performed comparably, outperforming the standardized regression coefficient and the method endorsed by Pratt (1987) (the product of the standardized regression coefficient for a predictor and its zero-order correlation with the criterion) in terms of identifying the dominant predictor (see Figure 1a); the standardized regression coefficient was outperformed by all of the other seven methods.

When the correlation between dominant predictor and criterion was low (.40) and sample size was at 4*p*, Budescu's and Johnson's methods again performed comparably, outperforming the *t*-statistic, the squared partial correlation, and the squared semi-partial correlation; Pratt's method significantly outperformed the standardized regression coefficient while the squared zero-order correlation did not significantly differ from any of the other importance methods. There were no significant differences between the importance methods when sample sizes ranged from 10*p* to 40*p*.

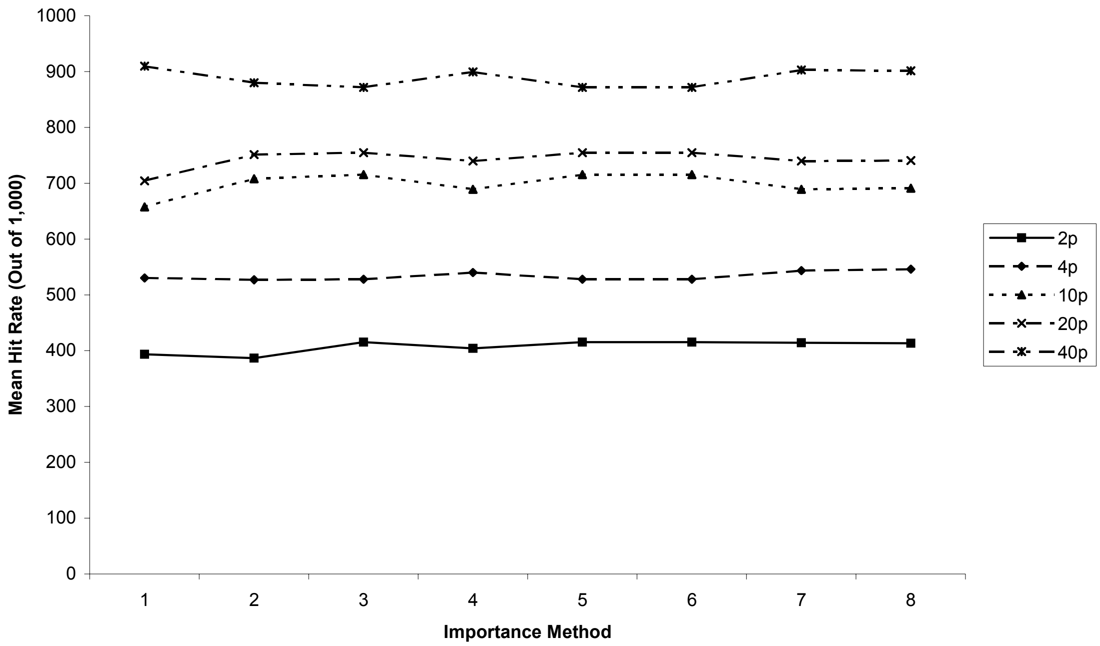
When the correlation between the dominant predictor and criterion was high (.60) and sample size was low (2*p*), the squared zero-order correlation, Pratt's method, Budescu's method, and Johnson's method all performed comparably and outperformed the standardized regression coefficient, the *t*-statistic, the squared partial correlation, and the squared semi-partial correlation (see Figure 1b). When the correlation between dominant predictor and criterion was high (.60) and sample size was at 4*p*, Budescu's method and Johnson's method again performed comparably, outperforming the *t*-statistic, the squared partial correlation, and the squared semi-partial correlation while Budescu's and Pratt's methods outperformed the standardized regression coefficient; the squared zero-order correlation did not significantly differ from any of the importance methods in terms of identifying the dominant predictor. There were no other significant differences between importance methods for sample sizes ranging from 10*p* to 40*p*.

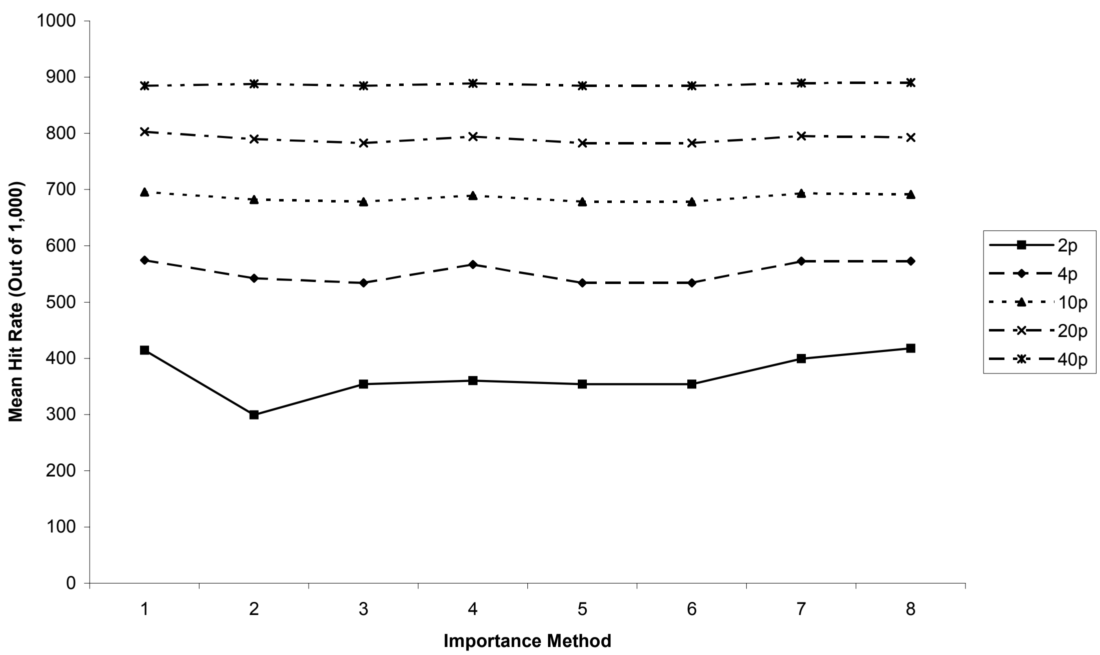
Importance Method × Sample Size. The ANOVA also indicated a significant interaction between importance method and sample size, $F(28, 840) = 4.84$, $p < .001$ (η^2 = .14). Post hoc tests indicated that when sample size was small (2*p*), the squared zero-order correlation, Budescu's method, and Johnson's method performed comparably, significantly outperforming the standardized regression coefficient, the *t*-statistic, Pratt's method, the squared partial correlation, and the squared semi-partial correlation (see Table 2); Pratt's method significantly outperformed the standardized regression coefficient.

When the sample size was 4*p*, Pratt's method, Budescu's method, and Johnson's method performed comparably, significantly outperforming the standardized regression coefficient, the *t*-statistic, the squared partial correlation, and the squared semi-partial correlation; the squared zero-order correlation did not significantly differ from any of the other importance methods. No other significant differences were detected at other sample sizes (10*p*-40*p*).



Figures 1a-b. Mean hit rates (out of 1,000 replications) as a function of importance method and sample size at a) low (.40), and b) high (.60) correlation between dominant predictor and criterion. Importance methods are: 1 = squared zero-order correlation; 2 = standardized regression coefficient; 3 = *t*-statistic; 4 = Pratt's method; 5 = squared partial correlation; 6 = squared semi-partial correlation; 7 = Budescu's method; 8 = Johnson's method.





Figures 2a-b. Mean hit rates (out of 1,000 replications) as a function of importance method and sample size at a) low (.10), and b) moderate (.40) correlation among predictors. Importance methods are: 1 = squared zero-order correlation; 2 = standardized regression coefficient; 3 = *t*-statistic; 4 = Pratt's method; 5 = squared partial correlation; 6 = squared semi-partial correlation; 7 = Budescu's method; 8 = Johnson's method.

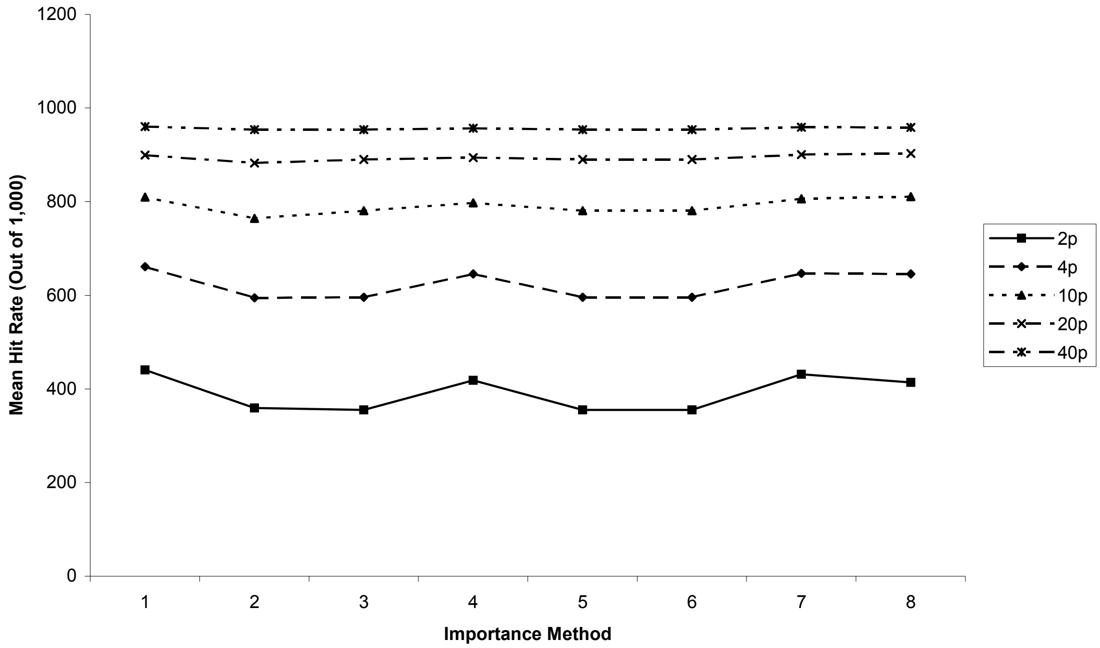


Figure 2c. Mean hit rates (out of 1,000 replications) as a function of importance method and sample size at high (.70) correlation among predictors. Importance methods are: 1 = squared zero-order correlation; 2 = standardized regression coefficient; 3 = t-statistic; 4 = Pratt's method; 5 = squared partial correlation; 6 = squared semi-partial correlation; 7 = Budescu's method; 8 = Johnson's method.

Table 1 Mean Number of Hits (Standard Deviations) out of 1,000 as a Function of Correlation Among Predictor

	$\rho_{yx_j}^2$	β^*_j	t_j	$\beta_j^* \rho_{yx_j}$	$\rho_{yx_j.x_1...x_p}^2$	$\rho_{y(x_j.x_1...x_p)}^2$	C_{x_j}	ε_j
Correlation Among Predictors								
.10	639.04	650.57	657.09	654.34	657.09	657.09	657.84	658.39
	(308.56)	(280.38)	(276.92)	(288.04)	(276.92)	(276.92)	(289.53)	(287.38)
.40	674.47	640.32	646.70	659.93	646.70	646.70	669.93	672.99
	(258.34)	(285.23)	(272.93)	(269.14)	(272.93)	(272.93)	(260.16)	(257.07)
.70	754.11	710.88	715.03	742.28	715.03	715.03	748.61	746.11
	(258.55)	(290.35)	(286.93)	(266.94)	(286.93)	(286.93)	(264.19)	(268.64)

Main Effect of Importance Method. The ANOVA also indicated a significant main effect of importance method, $F(7, 840) = 20.01$, $p < .001$ ($\eta^2 = .14$). The mean number of hits out of 1,000 for each importance method is reported in Table 3. Post hoc tests indicated that Budescu's method (C_{x_j}), and Johnson's index (ε_j) performed similarly by outperforming the remaining measures when identifying the dominant predictor, with the exception of the squared zero-order correlation. The squared zero-order correlation and Pratt's method significantly outperformed the standardized regression coefficient, the t-statistic, the squared partial correlation, and the squared semi-partial correlation, which all performed comparably.

Table 2: Mean Number of Hits (Standard Deviations) out of 1,000 as a Function of Sample Size

	$\rho_{yx_j}^2$	β^*_j	t_j	$\beta_j{*}\,\rho_{yx_j}$	$\rho^2_{yx_j.x_1...x_p}$	$\rho^2_{y(x_j.x_1...x_p)}$	C_{x_j}	ε_j
Sample Size								
2*p*	416.04	348.50	374.91	394.09	374.91	374.91	414.80	414.70
	(195.50)	(182.51)	(177.38)	(186.74)	(177.38)	(177.38)	(188.25)	(186.70)
4*p*	588.69	554.56	552.59	584.02	552.59	552.59	587.48	588.11
	(254.93)	(250.73)	(254.55)	(249.80)	(254.55)	(254.55)	(255.49)	(254.33)
10*p*	720.98	718.04	724.56	725.30	724.56	724.56	729.39	731.04
	(269.44)	(252.66)	(246.10)	(256.93)	(246.10)	(246.10)	(254.59)	(253.97)
20*p*	802.15	807.96	809.19	809.24	809.19	809.19	811.85	812.06
	(234.72)	(208.31)	(208.01)	(215.61)	(208.01)	(208.01)	(215.50)	(215.47)
40*p*	918.19	907.22	903.46	914.94	903.46	903.64	917.13	916.57
	(106.50)	(117.46)	(124.86)	(105.94)	(124.86)	(124.86)	(103.89)	(104.46)

Table 3: Mean Number of Hits (Standard Deviations) out of 1,000

$\rho_{yx_j}^2$	β^*_j	t_j	$\beta_j{*}\,\rho_{yx_j}$	$\rho^2_{yx_j.x_1...x_p}$	$\rho^2_{y(x_j.x_1...x_p)}$	C_{x_j}	ε_j
689.21[ab]	667.26[c]	672.94[bc]	685.52[bd]	672.94[bc]	672.94[bc]	692.13[a]	692.50[ad]
(279.33)	(285.99)	(279.58)	(276.79)	(279.58)	(279.58)	(273.58)	(273.03)

Note. Means that share the same letter superscript do not significantly differ.

Conclusion

One of the primary reasons for conducting this study was to determine which importance measure performs better in terms of identifying the correct dominant predictor. Similar to Johnson's (2000) findings, this Monte Carlo study indicates that Budescu's method (C_{x_j}) and Johnson's index (ε_j) perform comparably in terms of identifying the dominant predictor. Overall, both Budescu's and Johnson's methods also outperform the additional importance methods, with the exception of the squared zero-order correlation.

Trends did appear in the interactions that further substantiate the use of either Budescu's method or Johnson's method when determining predictor importance, especially under very small sample size conditions (2*p*-4*p*). As sample size increased (at 10*p*), however, the differences between all the importance methods became negligible, regardless of multicollinearity or dominance level. Budescu's method did differ from Johnson's method under the various levels of multicollinearity, in that Johnson's method performed better than Budescu's under moderate multicollinearity with a very small sample size (2p), whereas Budescu's method performed better than Johnson's under high multicollinearity with a very small sample size (2p). Again, however, as sample size increased, the differences between these two methods became negligible under these multicollinearity conditions. The squared zero-order correlation did not appear to differentiate itself as a viable measure of importance as it did not significantly differ from additional importance methods under certain conditions.

Interestingly, two of the factors investigated in the current study did not interact with the various importance methods in either two-way or three-way interactions, such as the number of predictors in the MR model or distribution type. This indicates that no significant differences emerge between the importance methods as a function of the levels of either of these factors. Still, the levels of the factors used in the current study may not have been extreme enough to be able to examine differences between importance methods. Thus, future studies could examine the effect of MR models with a larger number of predictors under more extreme levels of multivariate nonnormality.

In the current study, the *t*-statistic, the squared partial correlation, and the squared semi-partial correlation all performed identically, identifying the dominant predictor the same number of times under each condition. This may have been due to the homogeneous correlations among the predictor variables. As a result, real and simulated data sets with heterogeneous correlations among predictors were used to determine if these methods would differ under such conditions. The results of these analyses indicated that these three methods still identified the dominant predictor identically, indicating that the similarities between these three methods must be due to their definitions. In other words, all three methods are related to the variance in the model's multiple squared correlation that is attributable to a predictor variable after consideration of the additional variables' contribution to the model's squared multiple correlation.

Nursing Facility Consumer Satisfaction Survey

In an effort to improve the quality of care provided in nursing facilities, the Nursing Facility Consumer Satisfaction Survey (NFCSS) was developed (c.f., Cortés, Montgomery, Morrow, & Monroe, 2000). The survey consists of 12 items that assess general and specific consumer satisfaction with nursing facility care in certain domains, such as incontinence, physical activity, and medication management. Two versions of the survey were developed, one for nursing home residents and the other for family respondents. Each item is scored using a 7-point Likert scale ranging from 1 (very dissatisfied) to 7 (very satisfied).

In the first phase of a statewide longitudinal study, the survey was administered to a total of 138 family respondents of residents across 100 nursing facilities (Fouladi, 2001). For the purposes of this paper, 3 items which assess different types of activity satisfaction were selected to predict general satisfaction with the goal of identifying which activity satisfaction item is most associated with general satisfaction. One predictor variable was represented by the item on the survey: "How satisfied are you with the facility's ability to provide activities that your family member enjoy(s)?", to which responses symbolized satisfaction with enjoyable or recreational activities.

The second predictor variable was represented by the item: "How satisfied are you with the facility's ability to provide activities that keep your family member as physically active as possible?", which symbolized satisfaction with physical activities. The third predictor was represented by the item: "How satisfied are you with the facility's ability to provide activities that keep your family member as mentally alert as possible?", which symbolized satisfaction with mental alertness activities. The criterion variable represented overall satisfaction with the nursing facility and corresponded to the item: "Overall, how satisfied are you with your family member's experience in this nursing facility?".

These four items on the survey are shown in the Appendix. This particular model was selected due to the high level of multicollinearity among the predictor variables and the moderate correlation between each predictor variable and the criterion. In addition, the distributional properties of the variables in the data set are comparable to the distributional properties of the variables from the simulation study. Intercorrelations among the predictor variables and the criterion variable and their descriptive statistics are shown in Table 4.

Results

Table 5 shows the predictor variables' relative weights assigned by each importance method. With the exception of the squared zero order correlation and Pratt's method, $\beta_j * \rho_{yx_j}$, all of the importance methods selected the enjoyable activities predictor (predictor 1) as the most important variable. In contrast, the squared zero order correlation selected the physical activities predictor as most important and Pratt's method, $\beta_j * \rho_{yx_j}$, assigned the same weights to both enjoyable and physical activities, producing a tie between these two variables in terms of importance.

Conclusion

This data set demonstrates how similar both Budescu's (C_{x_j}) and Johnson's (ε_j) methods are in that they assigned identical weights to each predictor variable. Excluding the squared zero order correlation and Pratt's method, $\beta_j * \rho_{yx_j}$, all of the importance methods performed similarly to these two new methods, selecting enjoyable activities as the most important of the three predictor variables. Nonetheless, these additional methods do not take into account a predictor's direct and indirect effects as do both Budescu's (C_{x_j}) and Johnson's (ε_j) methods.

Researchers typically wish to establish the relative importance of predictors in MR models. Many techniques are used to do this, however, no consensus exists as to which is best. This is due to the common problem of multicollinearity, which renders the typical methods ambiguous and dependent upon the measure's definition of importance.

Budescu (1993) and Johnson (2000) have both established methods of importance that attempt to control for multicollinearity problems. The results of the simulation study are consistent with Johnson's (2000) finding that Budescu's method and Johnson's index perform comparably.

However, Budescu's method requires one to perform all possible regressions, which becomes fatiguing as the number of predictors in the MR model increases. Because Budescu's measure and Johnson's index performed comparably, it appears that Johnson's index would be the most computationally efficient measure to use if one is interested in determining predictor importance while accounting for a predictor's direct and total effects.

Table 4: Nursing Facility Consumer Satisfaction Survey Variables' Intercorrelations and Descriptive Statistics (N = 138)

Variables	*1*	*2*	*3*	*4*
1. Enjoyable Activities	--	.63*	.59*	.49*
2. Physical Activites		--	.73*	.50*
3. Mental Alertness Activities			--	.45*
4. Overall Satisfaction				--
Mean	6.01	5.77	5.79	6.25
Standard Deviation	1.02	1.19	1.10	0.93
Skew	-1.93	-1.72	-1.42	-1.85
Kurtosis	5.55	3.57	2.64	4.29

Note. * $p < .001$.

Table 5
Comparison of Relative Weights Calculated by Each Importance Method for the NFCSS Data

Predictors	$\rho_{yx_j}^2$	β^*_j	t_j	$\beta_j * \rho_{yx_j}$	$\rho_{yx_j \cdot x_1 \ldots x_p}^2$	$\rho_{y(x_j \cdot x_1 \ldots x_p)}^2$	C_{x_j}	ε_j
Enjoyable Activities	.24	.27	2.80	.13	.06	.04	.12	.12
Physical Activities	.25	.25	2.26	.13	.04	.03	.11	.11
Mental Alertness Activities	.20	.11	.99	.05	.01	.01	.08	.08

Note. N = 138. Average intercorrelation (in absolute value) among predictors = .65.

Future research should examine how various importance methods perform with heterogeneous correlations among predictor variables, which is typically the case with MR models. The focus of the current study was to determine the correct known dominant predictor, which is a commonly asked question by researchers. Still, there are instances in which researchers wish to know the rank order of predictor importance. In other words, which is the most important, the next most important, etc. Thus, future research could be implemented to investigate the performance of importance methods in terms of identifying the correct ranking of predictor variable importance.

The effects of multicollinearity and multivariate nonnormality on the importance methods were of particular interest in the current study. Although multicollinearity did affect the performance of relative importance methods, multivariate nonnormality did not. This is encouraging because multivariate nonnormality is typically found in real world data sets (Micceri, 1989). Additional research could examine extreme levels of multivariate nonnormality to determine whether there is a threshold at which point nonnormality does affect importance methods.

References

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, *48*(3), 209-213.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*(3), 542-551.

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

Cortés, L. L., Montgomery, E. W., Morrow, K. A., & Monroe, D. M. (December, 2000). *A statewide assessment of quality of care, quality of life and consumer satisfaction in Texas Medicaid nursing facilities.* Austin, TX: Texas Department of Human Services, Long Term Care Office of Programs, Medical Quality Assurance.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69*(3), 161-182.

Fouladi, R. T. (January, 2001). *Texas Department of Human Services Rider 26 longitudinal study of nursing home quality of care.* Austin, TX: Texas Department of Human Services, Long Term Care Office of Programs, Medical Quality Assurance.

Gibson, W. A. (1962). Orthogonal predictors: A possible resolution of the Hoffman-Ward controversy. *Psychological Reports*, *11*, 32-34.

Green, P. E., Carroll, J. D., & DeSarbo, W. S. (1978). A new measure of predictor importance in multiple regression. *Journal of Marketing Research*, *15*, 356-360.

Hamilton, L. C. (1992). *Regression with graphics.* Belmont, CA: Duxbury Press.

Headrick, T. C. & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distribution: Extending the Fleishman power method. *Psychometrika*, *64*(1), 25-35.

Healy, M. J. R. (1990). Measuring importance. *Statistics in medicine*, *9*, 633-637.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*(1), 1-19.

Johnson, R. M. (1966). The minimal transformation to orthonormality. *Psychometrika*, *31*, 61-66.

Kruskal, W. (1984). Concepts of relative importance. *Qüestiió*, *8*(1), 39-45.

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis.* Glenview, IL: Scott, Foresman and Company.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156-166.

Neter, J., Kutner, M. H., Nachtsheim, C. J., &Wasserman, W. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental designs*. (3rd ed.). Chicago, IL: Richard D. Irwin, Inc.

Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukkila & S. Puntanen (Eds.), *Proceedings of Second International Tampere Conference in Statistics* (pp. 245-260). University of Tampere, Finland.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465-471.