Routledge
Taylor & Francis Group

Multivariate Behavioral Research

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

A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression

Jeff W. Johnson

To cite this article: Jeff W. Johnson (2000) A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression, Multivariate Behavioral Research, 35:1, 1-19, DOI: 10.1207/S15327906MBR3501_1

To link to this article: https://doi.org/10.1207/S15327906MBR3501_1

Published online: 10 Jun 2010.

Submit your article to this journal

Article views: 1553

Citing articles: 395 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=hmbr20

Multivariate Behavioral Research, *35* (1), 1-19
Copyright © 2000, Lawrence Erlbaum Associates, Inc.

A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression

Jeff W. Johnson
Personnel Decisions Research Institutes, Inc.

The relative weight of predictor variables in multiple regression is difficult to determine because of non-zero predictor intercorrelations. *Relative weight* (also called relative importance by some researchers) is defined here as the proportionate contribution each predictor makes to R^2, considering both its unique contribution and its contribution when combined with other variables. Although there are no unambiguous measures of relative weight when variables are correlated, some measures have been shown to provide meaningful results (Budescu, 1993; Lindeman, Merenda, & Gold, 1980). These measures are very difficult to implement, however, when the number of predictors is greater than about five. A method is proposed that is computationally efficient with any number of predictors, and is shown to produce results that are very similar to those produced by more complex methods. Recommendations are made for when this procedure may be applied and in what situations it is not appropriate.

One of the most common requests of statistical consultants and researchers is for a measure of the relative importance of each predictor variable in a multiple regression analysis (Green, Carroll, & DeSarbo, 1978; Healy, 1990; Johnson, 1966). For example, customer satisfaction researchers are often interested in determining how customers' perceptions of specific attributes measured on a survey contribute to their ratings of overall satisfaction (Grisaffe, 1993; Heeler, Okechuku, & Reid, 1979; Jaccard, Brinberg, & Ackerman, 1986). Standardized regression coefficients and zero-order correlations are often used to answer this question, but many authors have maintained that these indices are inadequate in this case because they fail to consider both the effect the variable has by itself and in combination with the other variables in the

An earlier version of this article was presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Inc., Dallas, TX, April, 1998. I extend special thanks to Robert Cudeck for his many helpful suggestions and for pointing out other relevant research. Thanks also to Fred Oswald for his comments on earlier drafts of this article and for independently testing the procedure.

Correspondence concerning this article should be addressed to Jeff W. Johnson, Personnel Decisions Research Institutes, 43 Main Street Southeast, Suite 405, Minneapolis, Minnesota 55414. Electronic mail may be sent to jeffj@pdi-corp.com.

model (e.g., Budescu, 1993; Green & Tull, 1975; Hoffman, 1960). It would be very informative to have a measure that considers the direct effect of each predictor and its joint effect with other variables when partitioning the predictable criterion variance among them. This is what many previous articles on this topic have referred to as predictor "importance." Although the search for a meaningful index of importance has been ongoing for years (Budescu, 1993; Gibson, 1962; Goldberger, 1964), recent articles on the topic have shown that there is no single solution to the problem, and no "best" solution is likely to exist (Bring, 1994; Budescu, 1993; Healy, 1990; Kruskal, 1987; Kruskal & Majors, 1989; Theil, 1987).

Although previously suggested approaches were presented as measures of relative importance, this terminology is confusing because of the many different definitions of "importance" (Bring, 1994). For example, sometimes clients who ask for the most important variables in the regression equation really want to know how to get nearly as high an R^2 by measuring a smaller number of variables. In this case, all-subsets regression would be appropriate (Neter, Wasserman, & Kutner, 1985). If the predictor variables have a relevant, known ordering, the increase in R^2 as each variable is entered into the model is appropriate (Lindeman, Merenda, & Gold, 1980; Williams, 1978). If a hierarchical model (in which variables have direct and indirect effects on others) is postulated, path analysis would be appropriate (e.g., Borman, White, & Dorsey, 1995).

The measure proposed in this article is intended to be used when there is no inherent ordering of the variables, and the researcher is interested in the relative contribution each variable makes to the prediction of a dependent variable, considering both its unique contribution and its contribution when combined with other variables. To avoid confusion, I use the term "relative weight" to refer to this contribution, rather than relative importance. This is consistent with how Hoffman (1960, 1962) used the term, defining it as the proportionate contribution each predictor makes to the squared multiple correlation coefficient when that coefficient is expressed as the sum of contributions from the separate predictors. Although some alternative measures discussed in this article do not sum to the squared multiple correlation, I still refer to them as measures of relative weight to remain consistent.

The objectives of this article are to suggest an efficient method of calculating the relative weight of predictor variables, and to demonstrate the potential usefulness of this method by applying it to different multiple regression problems and comparing the results to results obtained by other methods. The reason for introducing another approach to the problem is that the proposed index has properties that make it preferable to previously suggested approaches. The proposed method is offered as a more computationally efficient and potentially useful aid to the interpretation of regression models.

Measures of Relative Weight

In the special case when all predictor variables are uncorrelated, the determination of relative weight is straightforward. By dividing the squared zero-order correlations between the predictor variables and the dependent variable (i.e., the squared standardized regression coefficients) by the squared multiple correlation, the relative weight of each variable is easily expressed as the proportion of predictable variance for which it accounts. This simple relationship no longer exists, however, in the presence of correlated predictors. Not only do the squared standardized regression coefficients no longer sum to the squared multiple correlation, but two variables that are highly correlated with each other and with the dependent variable may have very different regression coefficients. As relative weight is defined here, however, variables such as these should have very similar relative weights, because they are very similar to each other and predict the dependent variable about equally. Although many authors have suggested using standardized regression coefficients as measures of relative weight (e.g., Afifi & Clarke, 1990; Walsh, 1990), the confounding influence of correlations between predictor variables makes standardized regression coefficients uninterpretable in terms of relative weight.

If standardized regression coefficients are used as a measure of relative weight, the following conditions are likely to arise as a result of correlated predictors: (a) An exaggeration of the relative weight of the predictor variable most highly correlated with the dependent variable, (b) a diminishing of the relative weight of other variables in the model, (c) a reversal of signs (e.g., positive zero-order correlation but negative regression weight) that could make a variable appear to have an effect the opposite of its true relationship, and (d) small differences in samples could cause large differences in regression weights. Thus, the search for a measure that addresses the problem of correlated predictors has considerable practical value.

Many measures of relative weight have been suggested. Examples are the zero-order correlation between a predictor and the dependent variable, the product of the zero-order correlation and its associated standardized regression coefficient (Hoffman, 1960, 1962; Pratt, 1987), the increase in the squared multiple correlation that is associated with adding the variable to the model (the variable's *usefulness*; Darlington, 1968), and the t statistic associated with each variable (Bring, 1994; Darlington, 1990). Many reviews of relative weight measures have concluded that these measures fail to address the problem of correlated predictors, because they do not adequately reflect the contribution of a predictor by itself and in combination with other predictors (e.g., Budescu, 1993; Green & Tull, 1975; Kruskal & Majors, 1989).

Budescu (1993) suggested the average increase in R^2 associated with a variable across all possible sub-models (C_{x_i}) as a quantitative measure of relative weight. In other words, relative weight is reflected by the average usefulness of a variable when it is included with each possible combination of predictors. The sum of these average usefulnesses equals R^2, so the relative weight of each variable can be expressed as the proportion of predictable variance for which it accounts. This measure is computationally equivalent to a measure suggested previously by Lindeman et al. (1980), and seems to be effective in quantifying relative weight. The average increase in R^2 associated with the presence of a variable across all possible models is a meaningful measure that fits the definition of relative weight presented earlier. This measure averages a variable's direct effect (considered by itself), total effect (conditional on all predictors in the full model), and partial effect (conditional on all subsets of predictors; Budescu, 1993).

Budescu's method has a serious shortcoming, however, in that it is very difficult and time-consuming to compute, especially as the number of predictor variables increases. This method requires the computation of R^2s for all possible sub-models. Given p predictors, there are (2^p - 1) sub-models. There are only 31 sub-models for five predictors, but 1023 for 10 predictors, and 32,767 for 15. Computational requirements increase exponentially and are staggering for models with more than 10 predictors (Neter et al., 1985). Lindeman et al. (1980) stated that their method may not be feasible when p is larger than five or six. Budescu (1993) suggested using the RSQUARE option in SAS to obtain the necessary squared multiple correlations, but CPU time increases dramatically and all possible submodels are not provided if there are more than 10 predictor variables in the model (SAS Institute, 1990).

Because the computational requirements of Budescu's method make it unlikely that it would be applied to many situations for which it is valuable, it is worthwhile to consider alternative procedures that might yield similar results with less computation. Gibson (1962) and Johnson (1966) suggested that the relative weights for a set of variables can be approximated by creating a set of variables that are highly related to the original set of variables, but are uncorrelated with each other. The dependent variable can then be regressed on the new uncorrelated variables to approximate the relative weight of the original variables. The process of obtaining the best-fitting (in the least squares sense) set of orthogonal variables is as follows: Assume one has an $n \times 1$ vector of criterion scores ($\mathbf{y}$) and an $n \times p$ matrix of predictor scores ($\mathbf{X}$) of full rank, all of which are expressed in standard score form. Assuming $\mathbf{X}$ has at least as many rows as columns, it is possible to find the singular value decomposition of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{P\Delta Q}' \tag{1}$$

where $\mathbf{P}$ contains the eigenvectors of $\mathbf{XX}'$, $\mathbf{Q}$ contains the eigenvectors of $\mathbf{X}'\mathbf{X}$, and $\mathbf{\Delta}$ is a diagonal matrix containing the singular values of $\mathbf{X}$. The singular values are the square roots of the eigenvalues of $\mathbf{X}'\mathbf{X}$ and $\mathbf{XX}'$. The diagonal values in $\mathbf{\Delta}$ are unique except for sign, which can arbitrarily be taken to be positive, and $\mathbf{P}$ and $\mathbf{Q}$ are unique as long as no diagonal elements of $\mathbf{\Delta}$ are equal to zero (Johnson, 1966). If no two predictor variables in $\mathbf{X}$ are perfectly correlated with each other, $\mathbf{X}$ is of full rank and no diagonal elements of $\mathbf{\Delta}$ will be equal to zero. This is assumed to be the case in what follows.

Johnson (1966) showed that the best-fitting orthogonal approximation of $\mathbf{X}$ is obtained by

$$\mathbf{Z} = \mathbf{PQ}'. \tag{2}$$

The columns of $\mathbf{Z}$ are the best-fitting approximations to the columns of $\mathbf{X}$ in that they minimize the residual sum of squares between the original variables and the orthogonal variables (Johnson, 1966). The vector of beta weights when regressing $\mathbf{y}$ on $\mathbf{Z}$ is obtained by

$$\begin{aligned} \boldsymbol{\beta}^* = (\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{y} = (\mathbf{QP}'\mathbf{PQ}')^{-1}\,\mathbf{QP}'\mathbf{y} &= \mathbf{I}^{-1}\mathbf{QP}'\mathbf{y} \\ &= \mathbf{QP}'\mathbf{y}. \end{aligned} \tag{3}$$

The squared elements of $\boldsymbol{\beta}^*$ represent the proportion of predictable variance in $\mathbf{y}$ accounted for by $\mathbf{Z}$, and are considered to be close approximations to the relative weights of the columns of $\mathbf{X}$.

The $\mathbf{Z}$ variables are only approximations of the original variables, however, and may not be close representations if two or more original variables are highly correlated. For example, consider the correlation matrix at the top of Table 1. The x variables are highly intercorrelated, so the correlations between the x variables and their corresponding z variables are relatively low (see the middle of Table 1). Because of this, the partitioning of the predictable variance in y among the z variables does not

Table 1
Illustration of the Transformation to Best-Fitting Orthogonal Variables when Original Variables are Highly Correlated

Correlations Between Dependent Variable and Original Predictors

	y	x_1	x_2	x_3
y	1.00			
x_1	.50	1.00		
x_2	.40	.95	1.00	
x_3	.30	.85	.90	1.00

Correlations Between Original Variables (x) and Orthogonal Variables (z)

	z_1	z_2	z_3
x_1	.758	.521	.393
x_2	.521	.720	.458
x_3	.393	.458	.797

Relationships Between Orthogonal Variables (z) and Dependent Variable (y)

	Zero-Order Correlation	Relative Contribution to R^2
z_1	.549	93.5%
z_2	.143	6.4%
z_3	.023	0.2%

Note: Relative contribution to R^2 is calculated by dividing the squared zero-order correlation by R^2 and multiplying by 100. R^2 = .323.

adequately represent the relative weight of the x variables, given the zero-order correlations between y and each x. For example, z_3 only accounts for 0.2% of the predictable variance in y (bottom of Table 1). This would hardly be an accurate reflection of the relative weight of x_3, which has a zero-order correlation with y of .30.

Realizing the limitations of this procedure, Green et al. (1978) suggested a method by which the orthogonal variables could be related back to the original variables to better estimate the relative weight of the original variables. This was done by first regressing the columns of **Z** on **X**:

$$(4)\quad \mathbf{\Gamma} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Z} = (\mathbf{Q\Delta P'P\Delta Q'})^{-1}\,\mathbf{Q\Delta P'PQ'} = (\mathbf{Q\Delta}^2\mathbf{Q}')^{-1}\,\mathbf{Q\Delta Q}' = \mathbf{Q\Delta}^{-1}\,\mathbf{Q}',$$

where $\mathbf{\Gamma}$ is a matrix of regression weights of $\mathbf{Z}$ on $\mathbf{X}$. Note that $\mathbf{Z}$ is simply a linear transformation of $\mathbf{X}$ (i.e., $\mathbf{Z} = \mathbf{X\Gamma}$; Gibson, 1962). The relative contribution of the j^{th} original variable to the k^{th} orthogonal variable is then calculated as:

$$(5)\qquad \gamma_{jk}^{*2} = \frac{\gamma_{jk}^2}{\sum_{j=1}^{p} \gamma_{jk}^2}$$

where γ_{jk}^2 is the square of the general entry in $\mathbf{\Gamma}$. The relative weight of each original variable is obtained by

$$(6)\qquad \delta_j^2 = \sum_{k=1}^{p} \gamma_{jk}^{*2}\,\beta_k^{*2},$$

where β_k^{*2} is the square of the general entry in $\boldsymbol{\beta}^*$ (Equation 3). The sum of the δ_j^2s is equal to R^2.

This procedure is much simpler computationally than the Budescu (1993) or Lindeman et al. (1980) methods, but it has received criticism (Boya & Cramer, 1980; Jackson, 1980). Jackson (1980) criticized the procedure of regressing the orthogonal variables on the original variables to determine the relative contribution of each original variable to each orthogonal variable (see Equation 4), because the resulting weights (γ_{jk}^2) "are coefficients from regressions on correlated variables....(they) cannot meaningfully and unambiguously assign importance to the x_j's any more than could the β_j's from a regression of y on the x_j's" (Jackson, 1980, pp. 114-115). In a reply, Green and his colleagues acknowledged this criticism, but could only respond that their measure was at least better than previous methods of allocating importance (Green, Carroll, & DeSarbo, 1980).

A New Measure of Relative Weight

Calculation

Although Green et al.'s (1978) measure lacks meaning because regression coefficients are calculated on correlated variables, this problem can be solved by taking a different approach. Green et al. attempted to relate the orthogonal variables back to the original variables by using the set of weights for deriving the orthogonal variables from the original correlated predictors. Because we are going from the orthogonal variables back to the original variables, however, the more appropriate set of weights are those which derive the original variables from the orthogonal variables. In other words, instead of regressing **Z** on **X**, we regress **X** on **Z**. These weights are represented by the inverse of **Γ** (cf. Gibson, 1962):

$$(7) \qquad \mathbf{\Lambda}^* = (\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{X} = (\mathbf{QP}'\mathbf{PQ}')^{-1}\,\mathbf{QP}'\mathbf{P\Delta Q}' = \mathbf{I}^{-1}\,\mathbf{Q\Delta Q}'. \\ = \mathbf{Q\Delta Q}'$$

$\mathbf{\Gamma}^{-1}$ is renamed $\mathbf{\Lambda}^*$ to simplify notation. This procedure assigns regression coefficients to the uncorrelated **Z** variables instead of to the correlated **X** variables, so the problem of correlated predictors is not reintroduced with this method.

Because the **Z** variables are uncorrelated, the relative contribution of each z to each x is represented by the squared standardized regression coefficient (which is the same as the squared zero-order correlation) of each z for each x, represented by the squared column elements of $\mathbf{\Lambda}^*$ (λ_{jk}^{*2}). **X** is a linear transformation of **Z**, so the sum of the λ_{jk}^{*2}s is equal to one. Because the λ_{jk}^{*2}s are correlations, any particular λ_{jk}^{*2} represents the proportion of variance in z_k accounted for by x_j, just as it represents the proportion of variance in x_j accounted for by z_k. In other words, $\mathbf{Z}'\mathbf{X} = \mathbf{X}'\mathbf{Z}$, as shown by the following equations:

$$(8) \qquad \mathbf{Z}'\mathbf{X} = \mathbf{QP}'\mathbf{P\Delta Q}' = \mathbf{Q\Delta Q}'$$

$$(9) \qquad \mathbf{X}'\mathbf{Z} = \mathbf{Q\Delta P}'\mathbf{PQ}' = \mathbf{Q\Delta Q}'.$$

The proportionate contribution of a particular x_j to y can be estimated by multiplying the proportion of variance in each z_k accounted for by x_j (λ_{jk}^{*2}) by the proportion of variance in y accounted for by z_k (β_k^{*2}; see Equation 3), and summing the products. In other words, each z_k is a linear combination of the x_js, so the variance in y accounted for by z_k can be easily partitioned among the x_js according to the proportion of variance in z_k accounted for by each x_j. Because each x_j accounts for some of the variance in each z_k, the sum of these partitioned variances across all z_ks represents the relative weight for x_j. Using the notation $\mathbf{\Lambda}^{*[2]} = ||\lambda_{jk}^{*2}||$ and $\boldsymbol{\beta}^{*[2]} = ||\beta_k^{*2}||$, the new vector of relative weights, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$, is represented by

$$\boldsymbol{\varepsilon} = \boldsymbol{\Lambda}^{*[2]}\, \boldsymbol{\beta}^{*[2]}. \tag{10}$$

The sum of the relative weights is again equal to the model's squared multiple correlation, so the final relative weights can be expressed as percentages of the predictable variance affiliated with x_j by dividing ε_j by R^2 and multiplying by 100. The procedure for calculating $\boldsymbol{\varepsilon}$ from a correlation matrix is presented in the Appendix.

This measure overcomes the limitations of the Gibson (1962) and Johnson (1966) measure while avoiding the problem of correlated variables introduced by the Green et al. (1978) measure. The procedure can be applied to the correlation matrix in Table 1 to illustrate the calculation of $\boldsymbol{\varepsilon}$ and how the predictable variance in a dependent variable is more proportionately distributed among a set of predictor variables with this measure. Using the *xz* correlations ($\boldsymbol{\Lambda}^*$) and *yz* correlations ($\boldsymbol{\beta}^*$) from Table 1, $\boldsymbol{\varepsilon}$ is calculated as follows:

$$\begin{array}{rclclclcl}
 & & \underline{\quad x_1 \quad} & & \underline{\quad x_2 \quad} & & \underline{\quad x_3 \quad} & & \\
\beta_1^{*2} & = & (.758^2 \times .549^2) & + & (.521^2 \times .549^2) & + & (.393^2 \times .549^2) & = & .549^2 \\
+\ \beta_2^{*2} & = & (.521^2 \times .143^2) & + & (.720^2 \times .143^2) & + & (.458^2 \times .143^2) & = & .143^2 \\
+\ \beta_3^{*2} & = & (.393^2 \times .023^2) & + & (.458^2 \times .023^2) & + & (.797^2 \times .023^2) & = & .023^2 \\
R^2 & = & \varepsilon_1 = .179 & + & \varepsilon_2 = .093 & + & \varepsilon_3 = .051 & = & R^2 = .323 \\
 & & 55.4\% & + & 28.7\% & + & 15.9\% & = & 100\%
\end{array}$$

Recall that the *z* variables by themselves were assigned relative weights of 93.5%, 6.4%, and 0.2%, which were considered too asymmetric to adequately reflect the relative weights of the *x* variables. The distribution of weights is somewhat more symmetric when the δ^2 procedure is used, as x_1 gets a weight of 66.6%, x_2 gets a weight of 31.9%, and x_3 gets a weight of 1.6%. Because the relative weights should reflect the effect of each variable when considered in the context of the other predictors and by itself, and the zero-order correlation between x_3 and y is a relatively large .30, most observers would probably agree that the weight for x_3 is still too low relative to the others. When ε is used, however, x_1, x_2, and x_3 are assigned weights of 55.4%, 28.7%, and 15.9%, respectively. These are argued to be much more meaningful values, because they reflect the effect of each variable by itself and in combination with the others. When the Budescu procedure is applied to the correlation matrix in Table 1, the relative weights are 59.4%, 25.8%, and 14.8%, respectively. These weights are much closer to ε than to δ^2 or the weights for the *z* variables alone.

Comparisons

For illustrative purposes, a real data set was used to contrast ε with (a) the squared zero-order correlation between the predictor j and the dependent variable y ($r^2_{yx_j}$), (b) the squared standardized regression coefficient (β^2_j), (c) the product of the standardized regression coefficient and its associated zero-order correlation ($\beta_j r_{yx_j}$; Pratt, 1987), (d) the t statistics associated with each predictor (Darlington, 1990), (e) the squared standardized regression coefficient of y regressed on the least squares orthogonal equivalent of j (β^{*2}_j; Gibson, 1962), (f) δ^2_j (Green et al., 1978), and (g) the mean increase in R^2 across all possible subset regressions (C_{x_j}; Budescu, 1993). The correlation matrix was taken from Borman et al. (1995), and is shown in Table 2. The predictor variables are ratee technical proficiency, dependability, friendliness, obnoxiousness, and the extent to which the ratee is seen as a show-off. The dependent variable is overall job performance ratings made by peers. Borman et al. used path analysis with a larger set of variables, but only variables with direct effects on ratings were included in this illustration, so multiple regression is appropriate.

Relative weights as calculated by the different measures are shown in Table 3. The squared zero-order correlations are reasonable estimates of relative weight, but they do not sufficiently reflect the contribution of the variable in the context of each of the other variables. To fully reflect the relative weight of a variable, the direct effect should be combined with the joint effects it has with other variables (Green & Tull, 1975). The

Table 2
Correlation Matrix for Predictor Importance Measure Comparisons

	Ratings	Tech. Prof.	Dependable	Friendly	Obnoxious	Show-off
Ratings	1.00					
Tech. Prof.	.25	1.00				
Dependable	.32	.14	1.00			
Friendly	.26	.06	.81	1.00		
Obnoxious	-.26	-.18	-.52	-.58	1.00	
Show-off	-.05	.00	-.45	-.45	.73	1.00

Note: N = 631. Partially reproduced from "Effects of Ratee Task Performance and Interpersonal Factors on Supervisor and Peer Performance Ratings," by W. C. Borman, L. A. White, & D. W. Dorsey, 1995, *Journal of Applied Psychology, 80*, p. 173. Copyright 1995 by the American Psychological Association. Adapted with permission.

Table 3

Comparison of Eight Measures of Predictor Importance

	$r^2_{yx_j}$		β^2_j		$\beta_j r_{yx_j}$		t		β^{*2}_j		δ^2_j		C_{x_j}		ε_j	
Variable	RW	Pct.	RW	Pct.	RW	Pct.[a]	RW	Pct.[a]	RW	Pct.	RW	Pct.	RW	Pct.	RW	Pct.
Tech. Prof.	.062	20.5	.023	7.4	.038	--	4.03	--	.044	22.8	.043	22.2	.042	21.8	.044	22.9
Dependable	.102	33.7	.093	30.0	.098	--	4.84	--	.070	36.2	.059	31.0	.057	29.7	.057	29.7
Friendly	.068	22.4	.002	0.7	-.012	--	-0.72	--	.015	8.0	.027	13.9	.026	13.6	.027	14.3
Obnoxious	.068	22.4	.102	32.9	.083	--	-5.35	--	.052	27.0	.046	24.2	.044	23.0	.045	23.2
Show-off	.003	1.0	.090	29.0	-.015	--	5.51	--	.012	6.0	.017	8.7	.023	12.0	.019	9.9

Note: Dependent variable was ratings (see Table 2). Percentages were calculated by dividing individual values by their sum and multiplying by 100. R^2 = .192.

[a]Percentages could not be calculated because of negative values.

regression coefficients fail to take the direct effect of each variable into account. Because of the high correlation between dependable and friendly, the regression coefficients make it appear that friendly contributes almost nothing to ratings. Show-off has a very large regression coefficient despite its very low correlation with ratings, resulting in a suppression effect that is difficult to interpret (Borman et al., 1995). Technical proficiency also contributes very little compared to what would be expected. The product $\beta_j r_{yx_j}$ is even less interpretable. Because the regression coefficients of friendly and show-off have signs that are opposite the signs of their zero-order correlations, relative weights are negative. The *t* statistics also have some negative values, and suffer from too much weight placed on show-off and not enough on friendly.

Transforming the predictors to their least squares orthogonal equivalents (β_j^{*2}) effectively reduces the relative weight of show-off, but the high correlation between dependable and friendly results in a small proportion of the predictable variance being assigned to friendly and 4.5 times as much being assigned to dependable. C_{x_j}, δ^2, and ε allocate the proportion of predictable variance accounted for by dependable and friendly at a ratio of about 2:1. These three measures were all very similar in the relative weights they assigned to each predictor. For these measures, dependable had the largest weight, technical proficiency and obnoxious had about equal weights, friendly had a moderate weight, and show-off had the smallest weight. These three measures appear to be the most effective in capturing both the direct and indirect effects of the predictor variables on ratings.

Convergent Validity

No unambiguous measure of relative weight exists to which ε can be compared. It is presented in this article, however, as a more computationally efficient measure than that proposed by Budescu (1993) and a more meaningful measure than that proposed by Green et al. (1978). If the method proposed by Budescu (1993) is accepted as an appropriate measure of importance as defined here, the performance of ε can be evaluated by observing the convergent validity of ε with C_{x_j} across several regression models.

Thirty-one different data sets, both real and simulated, were used to evaluate ε as a measure of relative weight. The number of predictor variables ranged from 3 to 10 (mean = 6.84, *SD* = 2.10). The average intercorrelation between predictor variables ranged from .10 to .70 (mean = .40, *SD* = .18). Similarity between relative weights was calculated by computing the mean absolute deviation across predictors between weights derived from different methods. A summary of the results is shown in Table 4. Across all data sets, the smallest mean absolute deviation between weights computed by the ε measure and the C_{x_j} measure (when expressed as percentages of the predictable variance) was 0.05, and the largest was 1.99. The average across all data sets was 0.56 (SD = 0.38). Mean differences between weights derived by each method were unrelated to the number of predictor variables (partial r controlling for average predictor intercorrelation = -.07, *ns*), but were highly related to the average predictor intercorrelation (partial r controlling for number of predictor variables = .67, $p < .001$). In other words, differences between ε and C_{x_j} tend to be small, but differences tend to increase as the average correlation between the predictors increases. This is probably because regression weights are less stable under high multicollinearity, so both methods would tend to produce less reliable results.

When the criterion is similarity to relative weights derived by the Budescu method, ε was superior to δ^2. The mean absolute deviation from C_{x_j} was smaller for ε than for δ^2 in 28 of the 31 data sets studied, and the mean deviation across data sets was significantly smaller for ε than for δ^2 (mean = 1.12, SD = 0.91), t (30) = 4.76, $p < .001$. The difference between the C_{x_j} - δ^2 difference and the C_{x_j} - ε difference was correlated .70 with the average predictor intercorrelation. In other words, as the average correlation between predictors increases, δ^2 diverges from C_{x_j} to a greater extent than does ε.

Possible Criticisms

In addition to the point made by Jackson (1980), Boya and Cramer (1980) had two other criticisms of the Green et al. (1978) method that are also relevant to the new method. First, this method is not invariant to orthogonalizing procedures. In other words, Equations 7 and 10 can be applied to an infinite number of orthogonalizations of **X** and a set of weights will be derived that add up to the model's squared multiple correlation. For example, a principal components or a Gram-Schmidt orthogonalization would yield uncorrelated variables, but would not yield the same relative importance weights. Although it is possible to orthogonalize in infinitely many ways, the reason for using the least squares orthogonalization is that it more closely approximates the original variables than does any other orthogonalization procedure, in the sense that the original variables are maximally correlated with the orthogonal variables. Many different orthogonalization procedures were tried in

Table 4
Absolute Deviation Descriptive Statistics Comparing δ^2 and ε

		C_{x_j} - δ^2				C_{x_j} - ε			
Predictors (p)	Average Inter-r	Mean	SD	Min	Max	Mean	SD	Min	Max
3	.29	0.53	0.25	0.29	0.79	0.48	0.21	0.31	0.72
3	.37	0.14	0.10	0.03	0.21	0.91	0.43	0.50	1.36
4	.40	1.56	1.16	0.27	2.85	0.18	0.13	0.03	0.34
5	.23	0.27	0.24	0.05	0.63	0.24	0.19	0.00	0.48
5	.33	0.85	0.60	0.32	1.71	0.51	0.46	0.13	1.26
5	.37	1.34	1.21	0.38	3.36	0.86	0.82	0.10	2.14
5	.47	0.98	0.58	0.07	1.55	0.44	0.35	0.05	0.90
5	.56	2.01	1.58	0.28	4.03	0.98	0.93	0.21	2.44
5	.62	2.67	2.32	0.59	5.35	0.92	0.89	0.07	2.22
5	.69	2.63	2.34	0.52	6.05	0.75	0.52	0.12	1.32
6	.38	0.83	0.74	0.30	2.19	0.59	0.42	0.15	1.22
7	.10	0.09	0.08	0.01	0.18	0.05	0.04	0.00	0.11
7	.15	0.23	0.10	0.12	0.42	0.41	0.25	0.05	0.78
7	.15	0.24	0.18	0.03	0.47	0.19	0.18	0.03	0.51
7	.20	0.34	0.23	0.10	0.64	0.13	0.10	0.01	0.31
7	.30	0.76	0.52	0.24	1.45	0.20	0.16	0.08	0.55
7	.33	0.67	0.50	0.13	1.26	0.68	0.27	0.26	1.11
7	.37	0.91	0.96	0.22	2.99	0.79	0.45	0.34	1.54
7	.38	0.71	0.33	0.04	0.95	0.32	0.29	0.03	0.75
7	.40	1.34	0.90	0.44	2.58	0.35	0.23	0.12	0.84
7	.50	2.06	1.39	0.69	4.00	0.63	0.42	0.02	1.21
7	.60	2.93	1.97	1.01	5.70	1.14	0.78	0.02	2.21
7	.65	1.03	0.88	0.32	2.58	0.62	0.41	0.14	1.18
7	.70	3.89	2.61	1.37	7.61	1.99	1.30	0.16	4.05
10	.16	0.65	0.73	0.04	1.94	0.31	0.37	0.05	1.20
10	.16	0.64	0.99	0.02	3.07	0.45	0.50	0.02	1.58
10	.22	0.95	1.09	0.08	3.65	0.29	0.21	0.01	0.65
10	.22	0.70	0.84	0.01	2.35	0.49	0.57	0.04	1.80
10	.24	0.51	0.65	0.03	2.31	0.27	0.26	0.02	0.84
10	.52	0.94	0.98	0.06	3.11	0.52	0.38	0.02	1.10
10	.69	1.35	0.89	0.26	3.31	0.62	0.44	0.04	1.33
	Mean	1.12	0.90	0.27	2.56	0.56	0.42	0.10	1.23

Note: All relative weights and deviations are expressed in terms of percentages.

developing ε, and the only way to evaluate them was to compare the results with those obtained by an independent procedure (i.e., the Budescu method). The more closely the orthogonal variables approximated the original variables, the better the relative weights performed relative to the Budescu weights. It therefore made sense that the most successful transformation would be the least squares orthogonalization given by Gibson (1962) and Johnson (1966), because this yields a set of orthogonal variables that are maximally similar to the original variables. Comparisons of the different methods to Budescu's method verified this empirically. The orthogonalization method is therefore not arbitrary. It is the method that yields orthogonal variables that are most similar to the original variables and the method by which results are most similar to results obtained by Budescu's procedure.

The other additional criticism of δ^2 given by Boya and Cramer (1980) was that importance weights are not invariant under linear transformations of the variables. In other words, if some of the predictor variables are replaced by linear combinations of those variables, δ^2 for the unchanged variables does not stay the same. This is also true of ε, although the magnitude of the difference is relatively small. This may be considered a minor point, however, because lack of invariance under linear transformations is a property of many methods of estimating relative weight, not just δ^2 and ε. The only measures that are invariant under this condition are the beta weights, part and partial correlations with y, and Darlington's (1968) usefulness index, and these have all been shown to have more serious problems as measures of relative weight. The Budescu and ε measures are aids in the interpretation of regression equations, so it seems appropriate to accept a minor deficiency for a major gain in interpretability.

Recommendations

Any measure of the relative weight of predictors in a regression equation will have some ambiguity in the case of correlated variables, because the "true" relative weight is always unknown. The average increase in R^2 across all possible submodels (Budescu, 1993; Lindeman et al., 1980) and the ε measure presented in this article do, however, overcome the difficulties associated with many other measures that have been suggested, and they consistently yield results that make conceptual sense (i.e., they adequately reflect the proportional contribution of a predictor to the prediction of a dependent variable). It is also encouraging that two measures that have very different definitions and calculations produce very similar solutions. From a measurement standpoint, it probably does not matter which procedure is used because the differences between results are not generally large enough that they should be of concern. From a practical standpoint, however, the new method is almost always preferable over the Budescu method, because of the comparative ease with which it can be applied.

This procedure is meant to be an aid in the interpretation of regression results, not as a replacement for multiple regression. It is clearly not appropriate for many situations in which multiple regression is used. For example, ε should not be used to identify the best subset of variables for prediction purposes. The three or four variables with the highest ε values will not necessarily be the variables that jointly yield the highest R^2. Two variables that are highly correlated and have similar correlations with the dependent variable will usually have similar ε values, but the second variable does not raise R^2 much beyond what the first variable accounts for. Similarly, this procedure should not be used to identify suppressor variables, interaction effects, or redundant measures. It is not a substitute for what multiple regression already does well.

An example of an appropriate use for ε is in describing the information processing strategies of individuals rating the performance of others. Policy capturing is a method of applying multiple regression to statistically describe the implicit weighting strategies used by raters when making an overall performance evaluation (e.g., Hobson, Mendel, & Gibson, 1981; Zedeck & Kafry, 1977). The typical procedure used in policy capturing research is to develop performance profiles of hypothetical ratees in which dimension scores are designed to be uncorrelated. The artificiality of this task has led Hobson and Gibson (1983) to question the construct validity of captured rating policies and to recommend that future rater policy capturing studies be conducted exclusively in field settings with actual ratings. The method described here makes it possible to do this kind of research in a field setting, because the requirement that performance dimensions be uncorrelated is not necessary.

When applying this methodology, it is still necessary to consider the correlations between predictors and the constructs being measured. If two predictors are really measuring the same thing, a composite should be formed, or one should be dropped, because the relative weight will be spread out between them. Keeping both predictors could artificially inflate the weight of a construct that is measured by only one variable.

It is also important to remember that most variables in behavioral science to which this procedure could be applied are measured with some amount of error. Differences in the reliability of measurement can affect relative weights, because the magnitude of the correlation between two variables depends to some extent on the quality of the measures.

Unreliability tends to suppress correlations, so correlations between variables that are relatively less reliable will be suppressed more than correlations between variables that are relatively more reliable. In interpreting relative weights, the relative reliabilities of the input variables should always be considered, and small differences between relative weights should not be considered meaningful.

Summary

The new measure ε addresses the correlated-predictor difficulties associated with the measures proposed by Gibson (1962), Johnson (1966), and Green et al. (1978). It also yields results that are very similar to the average contribution to R^2 across all possible sub-models (Budescu, 1993; Lindeman et al., 1980). It is much more computationally efficient, however, especially with large numbers of predictors. These properties make it a potentially useful technique to apply to the special situation in which an estimate of the proportionate contribution to R^2 is desired for a set of predictors.

References

Afifi, A. A. & Clarke, V. (1990). *Computer aided multivariate analysis* (2nd ed.). New York: Van Nostrand Reinhold.

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80*, 168-177.

Boya, Ö. Ü. & Cramer, E. M. (1980). *Some problems in measures of predictor variable importance in multiple regression*. Unpublished manuscript, University of North Carolina at Chapel Hill.

Bring, J. (1994). How to standardize regression coefficients. *American Statistician, 48,* 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,* 542-551.

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

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

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

Goldberger, A. S. (1964). *Econometric theory*. New York: John Wiley & Sons, Inc.

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

Green, P. E., Carroll, J. D., & DeSarbo, W. S. (1980). Reply to "A comment on a new measure of predictor variable importance in multiple regression." *Journal of Marketing Research, 17,* 116-118.

Green, P. E. & Tull, D. S. (1975). *Research for marketing decisions* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Grisaffe, D. (1993, February). Appropriate use of regression in customer satisfaction analyses: A response to William McLauchlan. *Quirk's Marketing Research Review*, pp. 11-17.

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

Heeler, R. M., Okechuku, C., & Reid, S. (1979). Attribute importance: Contrasting measurements. *Journal of Marketing Research, 16,* 60-63.

Hobson, C. J. & Gibson, F. W. (1983). Policy capturing as an approach to understanding and improving performance appraisal: A review of the literature. *Academy of Management Review, 8,* 640-649.

Hobson, C. J., Mendel, R. M., & Gibson, F. W. (1981). Clarifying performance appraisal criteria. *Organizational Behavior and Human Performance, 28,* 164-188.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116-131.

Hoffman, P. J. (1962). Assessment of the independent contributions of predictors. *Psychological Bulletin, 59,* 77-80.

Jaccard, J., Brinberg, D., & Ackerman, L. J. (1986). Assessing attribute importance: A comparison of six methods. *Journal of Consumer Research, 12,* 463-468.

Jackson, B. B. (1980). Comment on "A new measure of predictor variable importance in multiple regression." *Journal of Marketing Research, 17,* 113-115.

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

Kruskal, W. (1987). Relative importance by averaging over orderings. *American Statistician, 41,* 6-10.

Kruskal, W. & Majors, R. (1989). Concepts of relative importance in recent scientific literature. *American Statistician, 43,* 2-6.

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

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models* (2nd ed.). Homewood, IL: Irwin.

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

SAS Institute (1990). *SAS/STAT user's guide, Version 6* (4th ed., Vol. 2). Cary, NC: SAS Institute, Inc.

Theil, H. (1987). How many bits of information does an independent variable yield in a multiple regression? *Statistics and Probability Letters, 6,* 107-108.

Walsh, A. (1990). *Statistics for the social sciences*. New York: Harper & Row.

Williams, E. J. (1978). Linear hypotheses: Regression. In W. H. Kruskal & J. M. Tanur (Eds.), *International encyclopedia of statistics* (pp. 523-541). New York: Free Press.

Zedeck, S. & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational Behavior and Human Performance, 18,* 269-294.

Accepted April, 1999.

Appendix

Calculation of Relative Weight by the New Method

Finding the singular value decomposition of **X** can often exceed the memory requirements of many PCs when **X** is large, so the preferred method is to use the correlation matrix to calculate orthogonal variables. The procedure for calculating orthogonal variables is from Gibson (1962). A correlation matrix **R** computed from one criterion variable (y) and p predictors ($x_1, x_2, \ldots, x_p$) can be separated into the intercorrelations between the predictors ($\mathbf{R}_{xx}$) and the correlations between y and each of the predictors ($\mathbf{R}_{xy}$). A decomposition of the following form exists for $\mathbf{R}_{xx}$:

(A1) $$\mathbf{R}_{xx} = \mathbf{X}'\mathbf{X} = \mathbf{Q\Delta P'P\Delta Q'} = \mathbf{Q\Delta}^{[2]}\mathbf{Q}',$$

where **Q** contains the eigenvectors of $\mathbf{R}_{xx}$, and $\mathbf{\Delta}^{[2]}$ is a diagonal matrix containing the eigenvalues of $\mathbf{R}_{xx}$ in its diagonal cells. It is then possible to obtain the following square factorization of $\mathbf{R}_{xx}$:

(A2) $$\mathbf{R}_{xx}^{1/2} = \mathbf{\Lambda}^{*} = \mathbf{Q\Delta Q}',$$

which Johnson (1966) showed to yield the set of orthogonal factors that have the highest degree of one-to-one correspondence with the original predictors. These are the correlations between the orthogonal variables and the original variables. The correlations between the orthogonal variables and the criterion variable are obtained by:

(A3) $$\boldsymbol{\beta}^{*} = \mathbf{R}_{xz}^{-1}\,\mathbf{R}_{xy} = \mathbf{\Lambda}^{*-1}\,\mathbf{R}_{xy},$$

because $\mathbf{R}_{xz} = \mathbf{X}'\mathbf{Z} = \mathbf{Q\Delta P'P\Delta Q'} = \mathbf{Q\Delta Q}' = \mathbf{R}_{xx}^{1/2}$. The squared elements of the $\mathbf{\Lambda}^{*}$ matrix and the $\boldsymbol{\beta}^{*}$ vector can then be plugged into Equation 10 to yield the vector $\boldsymbol{\varepsilon}$.