Statistica Sinica 3(1993), 407-420

DECOMPOSITION OF R^2 IN MULTIPLE REGRESSION WITH CORRELATED REGRESSORS

Abraham Genizi

Agricultural Research Organization

Abstract: Let correlated regressors and a dependent variable have a joint distribution, and assume that a suitable regression model has been found. A decomposition of R^2 into components corresponding to the regressors is proposed. The components are to serve as descriptive intuitive statistics indicating the relative importance of each regressor with respect to its overall effect on the dependent variable. When it is possible to partition the set of regressors into mutually orthogonal subsets, the sum of components of this decomposition in any subset is equal to the multiple R^2 due to that subset. Each component consists of a subcomponent "specific" to that regressor, and of "common" subcomponents with each of the other regressors. A measure of deviation of the set of regressors from orthogonality is given to help in assessing the amount of approximation used in the decomposition.

Key words and phrases: Multiple correlation coefficient, Decomposition of explained variance, Relative importance of regressors, Geometrical representation, Measure of deviation from orthogonality.

1. Introduction

The question of the "percentage of variance of the dependent variable 'explained' by each of the explanatory variables" is frequently raised and discussed in scientific papers of various disciplines. This is a relevant and important issue for many cases in which a model is investigated after its construction has been completed successfully. However, no satisfactory solution has been found to this problem for the most common case, in which the explanatory variables are not orthogonal to each other. Moreover, it is well known (see e.g. Williams (1978)), that in such cases no decomposition of R^2 exists with a meaningful allocation of the overall regression effect to the individual regressors. Kruskal (1984), and Kruskal and Majors (1989) discussed the philosophical problems involved in an attempt to decompose a joint effect of two or more interdependent factors into its components, and reviewed several measures given in the literature for the "relative importance" of variables in various contexts. In the present work we outline the type of problems for which it may be justified to decompose R^2 into

components associated with the regressors; then, a decomposition will be presented with some desirable properties. Due to these properties we advocate the components to serve as approximate measures of the relative importance of the regressors in a given regression.

The regression problems for which a decomposition may be relevant are those in which the uncontrolled regressors and the dependent variable have a joint distribution, because the 'contribution' of a regressor that is controlled depends on the range over which it is varied. This restriction may be somewhat relaxed to include regressors which may be controlled, in cases where it may be expected that the ranges of the regressors, and in fact the whole covariance matrix of the regressors, will remain unchanged in the future. Actually, this may be the justification for presenting R^2 itself in many cases. Also, if the effect of a regressor on the dependent variable is through a second regressor, then interest would usually be on the marginal effect of the second regressor after accounting for the first. In general, no cause-effect interrelationships between regressors are assumed, because our aim is a simultaneous rather then sequential decomposition.

We are looking for decompositions after a satisfactory model has been found and only within the framework of this model. The component of R^2 associated with a regressor may change as a result of any change in the set of regressors. Thus, no comparisons of components will be valid across different models, and components will be inappropriate for variable selection purposes. Also, the interpretation of the components will be narrower than in the orthogonal case, e.g. the component of variable x_i will not necessarily be the deduction from R^2 obtained by holding x_i fixed. The components will satisfy some reasonable criteria and will be accompanied by a measure of deviation of the regressor set from orthogonality. They should be considered as descriptive intuitive statistics indicating the relative importance of each of the regressors with respect to their overall effect on the dependent variable.

2. The Model and a Criterion for Decomposition

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$ and $\boldsymbol{y}$ be $p+1$ linearly independent vectors of n observations on p explanatory variables $x_1, x_2, \ldots, x_p$ and the dependent variable y, which have a joint distribution. Since we are interested only in correlations, we assume, without loss of generality, that all $p+1$ vectors are standardized with 0 means. The vector of least squares solution, $\boldsymbol{b}$, to the model $E(\boldsymbol{y}) = \boldsymbol{X}\beta$ is $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$. The vector fitted to the observed $\boldsymbol{x}_i$'s, $\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b} = \Sigma_i \boldsymbol{x}_i b_i$, is well known to be the projection of $\boldsymbol{y}$ onto the subspace $V(\boldsymbol{X})$ of R^n spanned by $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$. The multiple correlation coefficient estimate R is the cosine of angle between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$, and its square $R^2 = R^2_{y(\mathbf{X})} = \hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}/\boldsymbol{y}'\boldsymbol{y}$.

Naturally, the components of any decomposition in the case of uncorrelated (orthogonal) regressors should be proportional to the squared (simple) correlation coefficients between $\boldsymbol{y}$ and the corresponding regressors. Moreover it will required to satisfy the following

Criterion for orthogonal compatibility (OC): for a set of regressors partitioned into mutually uncorrelated subsets, the sum of components in any subset should be equal to the multiple R^2 between $\boldsymbol{y}$ and that subset of regressors.

The term 'component of a subset' will be used for the sum of components of the regressors in that subset of regressors. For a decomposition satisfying Criterion OC it also follows in the situation of mutually uncorrelated subsets, that a nonsingular linear transformation on any subset will change neither the component of that subset nor those of the regressors in any other subset.

Three simple statistics satisfy Criterion OC if all $\boldsymbol{x}_i$'s are orthogonal but not in general. They are
(i) the squared (simple) correlation coefficient $r^2_{y\boldsymbol{x}_i} = (\boldsymbol{y}'\boldsymbol{x}_i)^2/(\boldsymbol{y}'\boldsymbol{y} \cdot \boldsymbol{x}_i'\boldsymbol{x}_i)$;
(ii) the marginal reduction of $\boldsymbol{y}'\boldsymbol{y}$ due to $\boldsymbol{x}_i$ following reductions due to all other $\boldsymbol{x}_j$'s $(j \neq i)$, $R^2_{y\boldsymbol{x}_i.\mathbf{X}_{-i}} = R^2 - R^2_{y(\mathbf{X}_{-i})}$, where $\boldsymbol{X}_{-i}$ is the matrix $\boldsymbol{X}$ with column $\boldsymbol{x}_i$ deleted; and
(iii) the squared length of the vector component $(\hat{\boldsymbol{y}}_i)$ of $\boldsymbol{y}$ in the direction of $\boldsymbol{x}_i$ divided by $\boldsymbol{y}'\boldsymbol{y}$, that is, $\hat{\boldsymbol{y}}_i'\hat{\boldsymbol{y}}_i/\boldsymbol{y}'\boldsymbol{y} = b_i^2\boldsymbol{x}_i'\boldsymbol{x}_i/\boldsymbol{y}'\boldsymbol{y}$ which is proportional to the squared standardized partial regression coefficient.

Indeed, if the subspace of the regressors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is orthogonal to the other regressors, but $\boldsymbol{x}_1$ is not orthogonal to $\boldsymbol{x}_2$, it can be seen easily that the sum of the two components for any of these three statistics in the subspace $V(\boldsymbol{x}_1, \boldsymbol{x}_2)$, is either larger or smaller than R^2 between $\boldsymbol{y}$ and this subspace, which is $\hat{\boldsymbol{y}}_{12}'\hat{\boldsymbol{y}}_{12}/\boldsymbol{y}'\boldsymbol{y}$, where $\hat{\boldsymbol{y}}_{12}$ is the projection of $\boldsymbol{y}$ into this subspace.

Another simple statistic is presented sometimes as the decomposition of R^2. It is proportional to
(iv) $\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i = b_i\boldsymbol{x}_i'\boldsymbol{y}$, or equivalently to the (signed) length of the projection of $\hat{\boldsymbol{y}}_i$ on $\hat{\boldsymbol{y}}$. Although it satisfies Criterion OC, it is not suitable to serve as a decomposition of R^2, because some of its components in specific situations may turn out to be negative. This can be illustrated in two dimensions, in some of the cases when the projection $\hat{\boldsymbol{y}}_{12}$ of $\boldsymbol{y}$ on $V(\boldsymbol{x}_1, \boldsymbol{x}_2)$ falls outside the acute angles, between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Thus, although both $r_{y\boldsymbol{x}_1}$ and $r_{y\boldsymbol{x}_2}$ may be positive, the regression coefficient of the smaller r, say of $\boldsymbol{x}_2$ and consequently $\hat{\boldsymbol{y}}_2'\hat{\boldsymbol{y}}_{12}$, will be negative. They are said to serve as "corrections" to the too high values given to the corresponding coefficients of x_1. Pratt (1987) proposed $\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i$ as a measure of relative importance of the regressors. In his opinion, the possibility of negative values is "not a defect in definition", but it "signifies a situation too complex

for a single measure". He derived this measure by an axiomatic approach and gave to it several interpretations. (See the Discussion for comparisons with the measure proposed here.)

There is another way to define a decomposition of R^2, which satisfies Criterion OC (see also Kruskal (1984), Pratt (1987)). Its component for x_i is

$$\text{(v)} \qquad \bar{r}_i^2 = \frac{1}{p}\sum_{k=0}^{p-1}\binom{p-1}{k}^{-1}\sum_{\boldsymbol{X}_k} R^2_{yx_i.\mathbf{X}_k}, \quad i = 1, 2, \ldots, p, \tag{1}$$

where the second summation goes over all possible groups of k variables, $\boldsymbol{X}_k$, out of $x_1, x_2, \ldots, x_p$ excluding x_i, $R^2_{yx_i.\mathbf{X}_k} = R^2_{y(x_i\mathbf{X}_k)} - R^2_{y(\mathbf{X}_k)}$ and $R^2_{yx_i.\mathbf{X}_0} = r^2_{yx_i}$. This is derived easily if we take all possible 'hierarchical' or stepwise decompositions of R^2 into elements of type $R^2_{yx_i.\mathbf{X}_k}$. (For example, for $p = 3$, six decompositions of the type $R^2 = R^2_{yx_1.x_2x_3} + R^2_{yx_2.x_3} + R^2_{yx_3}$ are obtained by permutation of the indices 1, 2 and 3.) Each decomposition contains p components, with exactly one component of the type $R^2_{yx_i.\mathbf{X}_k}$ for each i $(1 \le i \le p)$. Averaging these hierarchical decompositions separately for elements of type $R^2_{yx_i.\mathbf{X}_k}$ for each i $(i = 1, 2, \ldots, p)$ gives the above decomposition. This averaging process over all permutations is carried out in the absence of a "natural" ordering of the regressors which would yield the $R^2_{yx_i.\mathbf{X}_k}$ of that ordering as measures of their relative importance (Kruskal (1987)). This decomposition involves a large amount of calculations even for moderately large p's.

In the next Section our proposal for decomposition in the parameter space using a statistical approach will be presented and in Section 4 a description of the sample R^2 decomposition will be given from a geometrical point of view, followed by discussion and a simulation study comparing some measures of relative importance of regressors.

3. The Proposed Decomposition

Let us assume temporarily, that $x_1, x_2, \ldots, x_p$ all have zero expectation, unit variance and known correlation matrix $\boldsymbol{\Sigma}$, $\operatorname{cor}(x_i, x_j) = \operatorname{cov}(x_i, x_j) = \Sigma_{ij}$. Assume also that $E(y) = 0$, $\operatorname{var}(y) = 1$ and the correlations $\operatorname{cor}(y, x_i) = \rho_{yx_i}$, $i = 1, 2, \ldots, p$, are also known. It should be noted that the problem of decomposing $R^2 = \boldsymbol{\rho}'_{y\mathbf{X}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}_{y\mathbf{X}}$, where $\boldsymbol{\rho}_{y\mathbf{X}} = (\rho_{yx_1}, \ldots, \rho_{yx_p})'$, into components associated with the regressors, exists even in the present set up of known $\boldsymbol{\rho}_{y\mathbf{X}}$ and $\boldsymbol{\Sigma}$.

We are given the regression equation $\mathrm{E}y = \Sigma_i \beta_i x_i$. Let $z_1, z_2, \ldots, z_p$ be the (unique) linear transformation of $x_1, x_2, \ldots, x_p$ which minimizes $\Psi = E\Sigma_i(z_i - x_i)^2$ subject to the condition that the z_i be uncorrelated with expectation 0 and variance 1. This is a transformation to a new uncorrelated set of regressors

z_i, which are nearest to the corresponding x_i in the sense that the expected sum of squared deviations Ψ between the pairs (x_i, z_i) is minimized. Let the transformation be written in the form $x_i = \sum_j a_{ij} z_j$, $i = 1, 2, \ldots, p$, and the regression in the new variables as

$$Ey = \Sigma_j c_j z_j, \quad \text{with} \quad c_j = \Sigma_i \beta_i a_{ij}, \tag{2}$$

then by the definition of $z_1, z_2, \ldots, z_p$ it follows that

$$a_{ij} = \rho_{z_i x_j} \quad \text{and} \quad c_j = \rho_{y z_j}. \tag{3}$$

Our proposal assigns to x_i, in the decomposition of $\boldsymbol{R}^2$, the component

$$v_i = \Sigma_j a_{ij}^2 c_j^2. \tag{4}$$

v_i is preferred over taking c_i^2 itself as the component assigned to x_i, because although each z_j is closest to the corresponding x_j in the above sense, it is still correlated in general to the other x_j's with the coefficients a_{ij}; therefore a component of a reasonable decomposition should be composed of a linear combination of all the c_j^2s with coefficients which reflect the magnitude of these correlations. Further aspects of this proposal will be discussed in later sections.

It remains to obtain the transformation matrix $\boldsymbol{A} = [a_{ij}]$. Since $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z}$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ and $\boldsymbol{z} = (z_1, z_2, \ldots, z_p)'$, we have $E\boldsymbol{x}\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Sigma}$, the correlation matrix of the x's. (Note that $\boldsymbol{x}$ is a random $p \times 1$ vector, whereas the $\boldsymbol{x}_i'$s are $n \times 1$ vectors of observations on the variable x_i, $i = 1, \ldots, p$.) We have to choose $\boldsymbol{A}$ among the matrix square roots of $\boldsymbol{\Sigma}$ in a way that it should minimize Ψ. It will now be shown that $\boldsymbol{A}$ must be the symmetrical square root of $\boldsymbol{\Sigma}$, i.e.

$$\boldsymbol{A} = \boldsymbol{Q}' \boldsymbol{D}_\alpha \boldsymbol{Q}, \tag{5}$$

where $\boldsymbol{\Sigma} = \boldsymbol{Q}' \boldsymbol{D}_\alpha^2 \boldsymbol{Q}$ is the principal decomposition of $\boldsymbol{\Sigma}$, $\boldsymbol{Q}$ is an orthogonal matrix whose rows are the eigenvectors of $\boldsymbol{\Sigma}$ and $\boldsymbol{D}_\alpha^2$ is diagonal with the eigenvalues $\alpha_1^2, \alpha_2^2, \ldots, \alpha_p^2$ of $\boldsymbol{\Sigma}$ along its diagonal.

The general form of a matrix square root of $\boldsymbol{\Sigma}$ is $\boldsymbol{A} = \boldsymbol{Q}' \boldsymbol{D}_\alpha \boldsymbol{G}$, with an arbitrary orthogonal matrix $\boldsymbol{G}$. We may express Ψ as

$$\begin{aligned} \Psi &= E(\boldsymbol{z} - \boldsymbol{x})'(\boldsymbol{z} - \boldsymbol{x}) = \boldsymbol{I} + \boldsymbol{\Sigma} - 2E\boldsymbol{z}'\boldsymbol{x} = \boldsymbol{I} + \boldsymbol{\Sigma} - 2E\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z} \\ &= \boldsymbol{I} + \boldsymbol{\Sigma} - 2\mathrm{tr}E\boldsymbol{z}\boldsymbol{z}'\boldsymbol{A} = \boldsymbol{I} + \boldsymbol{\Sigma} - 2\mathrm{tr}\boldsymbol{A}. \end{aligned}$$

Thus, to minimize Ψ we have to maximize $\mathrm{tr}\boldsymbol{A} = \mathrm{tr}\boldsymbol{Q}'\boldsymbol{D}_\alpha\boldsymbol{G} = \mathrm{tr}\boldsymbol{D}_\alpha\boldsymbol{G}\boldsymbol{Q}'$. To prove that $\boldsymbol{G} = \boldsymbol{Q}$ maximizes Ψ we state the following lemma.

Lemma. *If $\boldsymbol{D}$ is a diagonal matrix with positive diagonal elements, the trace of $(\boldsymbol{DH})$ is maximized over the set of all orthogonal matrices $\{\boldsymbol{H}\}$ when $\boldsymbol{H} = \boldsymbol{I}$ the identity matrix.*

This follows easily from $\mathrm{tr}(\boldsymbol{DH}) = \Sigma_i d_{ii} h_{ii}$ and $h_{ii} \leq 1$ for orthogonal $\boldsymbol{H}$, with equality signs occurring for all i iff $\boldsymbol{H} = \boldsymbol{I}$.
The lemma implies that $\mathrm{tr}\boldsymbol{D}_\alpha \boldsymbol{G}\boldsymbol{Q}'$ with all $\alpha_i > 0$, is maximized when the orthogonal matrix $\boldsymbol{G}\boldsymbol{Q}' = \boldsymbol{I}$, i.e., when $\boldsymbol{G} = \boldsymbol{Q}$. It will be required without loss of generality that $\alpha_i > 0$ $(1 \leq i \leq p)$, because otherwise, to maximize $\mathrm{tr}\boldsymbol{A}$, all rows of $\boldsymbol{G}$ and $\boldsymbol{Q}$ corresponding to negative α_i must satisfy $\boldsymbol{g}_i'\boldsymbol{q}_i = -1$, or $\boldsymbol{g}_i = -\boldsymbol{q}_i$, resulting in the same $\boldsymbol{A}$ as given by (5).

Thus, $\boldsymbol{A}$ is symmetric and $\boldsymbol{A}^2 = \boldsymbol{\Sigma}$, hence $\Sigma_k a_{ik}^2 = \Sigma_k a_{kj}^2 = 1$ for $1 \leq i \leq p$ and $1 \leq j \leq p$, because these sums are equal to the diagonal elements of the correlation matrix $\boldsymbol{\Sigma}$. It follows that the sum of components

$$\Sigma_i v_i = \Sigma_i \Sigma_k a_{ik}^2 c_k^2 = \Sigma_k \Big(\Sigma_i a_{ik}^2\Big) c_k^2 = \Sigma_k c_k^2 = \Sigma_k \rho_{yz_k}^2 = R^2, \tag{6}$$

because the presentation of the regression as a regression on the uncorrelated set $z_1, z_2, \ldots, z_p$ gives an exact decomposition of $\boldsymbol{R}^2$.

To prove that decomposition (4) satisfies the OC criterion, suppose the $\boldsymbol{x}$ consists of, say, m uncorrelated subsets of variables. By rearranging $\boldsymbol{x}$, $\boldsymbol{\Sigma}$ will become a block diagonal matrix, and $\boldsymbol{A}$ and $\boldsymbol{\Sigma}^{-1}$ will be also block diagonal with blocks at the same positions. Let the subsets be given by $V_1, V_2, \ldots, V_m$, where $i \in V_k$ if x_i belongs to the kth subset. Thus, $\boldsymbol{R}^2 = \rho_{y\mathbf{x}}' \boldsymbol{\Sigma}^{-1} \rho_{y\mathbf{x}}$, will decompose into these subsets, say, $\boldsymbol{R}^2 = \boldsymbol{R}_1^2 + \boldsymbol{R}_2^2 + \ldots + \boldsymbol{R}_m^2$, where $\boldsymbol{R}_k^2$ is constructed only by ρ_{yx_j}'s with $j \in V_k$. Also from (2) and (4) it can be seen, that the z_i and the components v_i with $j \in V_k$, will also be formed only from a_{ji} with both indices belonging to V_k. Hence, by the same reasoning as used above for $\boldsymbol{R}^2$ itself, it follows that

$$\sum_{i \in V_k} v_i = \boldsymbol{R}_k^2. \tag{7}$$

It should be noted that (6) and (7) were proved using the special form (5) of $\boldsymbol{A}$, and they will not hold in general for any square root of $\boldsymbol{\Sigma}$. The principal decomposition of $\boldsymbol{\Sigma}$ is essentially unique when all eigenvalues are different. However, even when not all α_i are different, $\boldsymbol{A}$ as given by (5) and $\boldsymbol{A}^{-1}$ are always unique.

The decomposition (4) can be written using the relation $\boldsymbol{z} = \boldsymbol{A}^{-1}\boldsymbol{x}$ as

$$v_i = \Sigma_j \Big(\Sigma_k \rho_{yx_k} a^{kj}\Big)^2 a_{ij}^2, \tag{8}$$

where a^{kj} is the kjth element of $\boldsymbol{A}^{-1}$. This form involves only the correlations ρ_{yx_k} and $a_{ij} = \rho_{x_i x_j}$.

In the following, the general case will be considered when the expectation and variance of the variables are unknown. In this setup, the component assigned to x_i in the decomposition of R^2 will be estimated by (8) or (4), where the matrix $\boldsymbol{A} = [a_{ij}]$ now stands for the symmetric square root of the matrix of sample correlations $r_{x_i x_j}$ and r_{yx_k} replaces ρ_{yx_k}.

4. The Sample R^2 Decomposition and its Geometrical Representation

Given a sample of $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$, and $\boldsymbol{y}$, we assume, without loss of generality, that they are standardized with zero means and unit lengths. In analogue to the previous procedure, we construct an orthonormal basis $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_p$ for the p dimensional subspace $V(\boldsymbol{X})$ in R^n spanned by $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$, with the following properties: (a) $\Sigma_i \boldsymbol{x}_i' \boldsymbol{z}_i$ is maximized; (b) $\boldsymbol{x}_i' \boldsymbol{z}_j = \boldsymbol{x}_j' \boldsymbol{z}_i$ for all $i, j = 1, 2, \ldots, p$; and (c) $\delta = \|\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{Z}'\boldsymbol{X}\|$ is minimized, where $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_p)$. Property (a) states that $\boldsymbol{z}_i$'s are closest to their corresponding $\boldsymbol{x}_i$'s in the sense that the sum of $\cos(\boldsymbol{x}_i, \boldsymbol{z}_i)$ is maximized. Property (b) shows that the $\boldsymbol{z}_i$'s are positioned around the $\boldsymbol{x}_i$'s in a symmetrical configuration so that for any pair i, j the angle between $\boldsymbol{x}_i$ and $\boldsymbol{z}_j$ is equal to that between $\boldsymbol{x}_j$ and $\boldsymbol{z}_i$. Finally, property (c) states that $\boldsymbol{z}_i' \boldsymbol{x}_j$ is closest to $\boldsymbol{x}_i' \boldsymbol{x}_j$ in the sense that their sum of squares of deviation over all $i, j = 1, 2, \ldots, p$ is minimized.

Let $\boldsymbol{X}$ be presented by the basis in the form $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{A}$; then, analogous to the previous derivation and using the lemma, it turns out from any of the requirements (a) or (c), that $\boldsymbol{A}$ is the symmetric matrix square root of $\boldsymbol{X}'\boldsymbol{X} = \mathbf{R}_{XX}$, the matrix of sample correlations between the $x's$. Requirement (b) is clearly satisfied by using this matrix $\boldsymbol{A}$. In the sample all eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ are different with probability 1.

The decomposition of each $\boldsymbol{x}_i$ along the directions of the $\boldsymbol{z}_j$'s is $\boldsymbol{x}_i = \Sigma_j \boldsymbol{z}_j \boldsymbol{z}_j' \boldsymbol{x}_j$. Similarly, each correlation coefficient $r_{yx_i} = \boldsymbol{y}'\boldsymbol{x}_i$ can be written in the form

$$r_{yx_i} = \sum_j \boldsymbol{y}' \boldsymbol{z}_j \boldsymbol{z}_j' \boldsymbol{x}_i = \Sigma_j c_j a_{ji}, \tag{9}$$

where $c_j = \boldsymbol{y}'\boldsymbol{z}_j = r_{yz_j}$.

In view of properties (a), (b) and (c) of the set of vectors $\boldsymbol{z}_j$, we may consider $\boldsymbol{y}'\boldsymbol{z}_i\boldsymbol{z}_i'\boldsymbol{x}_i = c_i a_{ii}$ as the "specific" component of $\boldsymbol{x}_i$, and $\boldsymbol{y}'\boldsymbol{z}_j\boldsymbol{z}_j'\boldsymbol{x}_i = c_j a_{ji}$ as the "common" component of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ $(j \neq i = 1, 2, \ldots, p)$ in the correlation coefficient between $\boldsymbol{y}$ and $\boldsymbol{x}_i$.

Likewise, we define the component assigned to $\boldsymbol{x}_i$ in the decomposition of $R^2_{y(\mathbf{X})}$ as

$$r^2_{yx_i}(\boldsymbol{X}) = \sum_j (\boldsymbol{y}' \boldsymbol{z}_j \boldsymbol{z}_j' \boldsymbol{x}_i)^2 = \Sigma_j (c_j a_{ji})^2. \tag{10}$$

Here we changed the component's notation to indicate its dependence on the whole set of regressors. It contains, as above, a specific component due to $\boldsymbol{x}_i$ and common components with all other $\boldsymbol{x}_j's$ $j \neq i$.

Using $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A}^{-1}$, we can express $\boldsymbol{c} = (c_1, \ldots, c_p)$, too, as functions of correlations only, in the form $\boldsymbol{c} = \boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{r}_{\mathbf{X}y}$; hence $r^2_{yx_i}(\boldsymbol{X})$ can be expressed in the form

$$r^2_{yx_i}(\boldsymbol{X}) = \Sigma_j \left(\Sigma_k r_{yx_k} a^{kj}\right)^2 a^2_{ji}, \tag{11}$$

where a^{kj} is the kjth element of $\boldsymbol{A}^{-1}$.

Proofs that this is a decomposition, i.e., $\Sigma_i r^2_{yx_i}(\boldsymbol{X}) = R^2$, and that it satisfies Criterion OC when component sets of the sample vectors are mutually orthogonal are completely analogous to those given above for the population, and will not be repeated here.

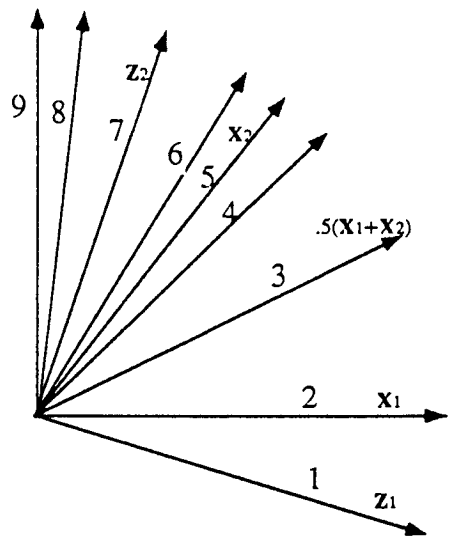
5. Discussion

(a) Let $\boldsymbol{R}_{XX} \to \boldsymbol{I}$ in the form $c\boldsymbol{R}_{XX} + (1-c)\boldsymbol{I}$, where $c \to 0$. This is equivalent to having $\boldsymbol{D}^2_\alpha \to \boldsymbol{I}$ in the form $c\boldsymbol{D}^2_\alpha + (1-c)\boldsymbol{I}$ while using the *same* matrix $\boldsymbol{Q}$, since $\boldsymbol{Q}'\{c\boldsymbol{D}^2_\alpha + (1-c)\boldsymbol{I}\}\boldsymbol{Q} = c\mathbf{R}_{XX} + (1-c)\boldsymbol{I}$. Therefore, if we also leave $\boldsymbol{P}$ unchanged, the coordinate system $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_p$ may remain, by this method, preserved, while each vector $\boldsymbol{x}_i$ will tend to its corresponding coordinate vector, and $r^2_{yx_i}(\boldsymbol{X})$ will converge to $(\boldsymbol{y}'\boldsymbol{z}_i)^2 = c_i^2 = (\Sigma_k r_{yx_k} a^{ki})^2$. Thus, the special coordinate system has a further property, that it is the set to which the vectors $\boldsymbol{x}_i$ can smoothly tend when uncorrelatedness is approached in the simple way of reducing simultaneously all correlation coefficients towards zero by the same proportion.

(b) A few special cases will now be considered to illustrate the proposed decomposition.

1. $p = 2$. In this case, to construct orthogonal $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ satisfying $\boldsymbol{z}_1'\boldsymbol{x}_2 = \boldsymbol{z}_2'\boldsymbol{x}_1$, the angle between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ has to be enlarged (or constricted) symmetrically with respect to their bisector (Fig.1) to an orthogonal one. Figure 1 also shows the changes occurring in the components of two measures when the position of $\hat{\boldsymbol{y}}$ is changed in relation to a specific configuration of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The cases $\hat{\boldsymbol{y}} = \boldsymbol{x}_1$ and $\hat{\boldsymbol{y}} = \boldsymbol{x}_2$ are unrealistic because in such cases the model will consist of one regressor only. They can be viewed as limiting cases $\hat{\boldsymbol{y}} \to \boldsymbol{x}_1$ or $\hat{\boldsymbol{y}} \to \boldsymbol{x}_2$ when R^2 also tends to 1 and therefore even a small contribution of the other regressor is important. The case when one of the simple correlations is zero, say $r^2_{yx_1} = 0$ ($\hat{\boldsymbol{y}}$ in position 9 in Figure 1), is instructive. Usually one will have, even in this case, $R^2 > r^2_{yx_2}$. Therefore, it is justified to state that $\boldsymbol{x}_1$ has also contributed to R^2 and thus to have a positive component for $\boldsymbol{x}_1$. Indeed, $\boldsymbol{x}_1$ is entitled to one

half of the difference $R^2 - r^2_{yx_2}$. In a similar way the rationale behind the mode of changes in $r^2_{yx_i}(\boldsymbol{X})$ can be seen to be reasonable.



position of $\hat{\boldsymbol{y}}$	$r^2_{yx_1}$	$r^2_{yx_2}$	$r^2_{yx_1}(\boldsymbol{X})$	$r^2_{yx_2}(\boldsymbol{X})$	$\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_1$	$\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_2$
1	.90	.10	.90	.10	1.125	−.125
2	1.00	.36	.82	.18	1.00	.00
3	.80	.80	.50	.50	.50	.50
4	.50	.98	.26	.74	.125	.875
5	.36	1.00	.18	.82	.00	1.00
6	.26	.99	.14	.86	−.07	1.07
7	.10	.90	.10	.90	−.125	1.125
8	.01	.73	.14	.86	−.07	1.07
9	.00	.64	.18	.82	.00	1.00

Figure 1. Squared correlation coefficients and components of R^2 (in units of R^2) according to two definitions, $r^2_{yx_i}(\boldsymbol{X}) = \bar{r}^2_i$ and $\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i$, for different directions of $\hat{\boldsymbol{y}}$ in a specific case with $p = 2$ and $\cos(x_1, x_2) = 0.6$ (deviation measure from orthogonality $\Delta = 0.7$).

In this case $r^2_{yx_i}(\boldsymbol{X})$ $(i = 1, 2)$ can be expressed by simpler terms in the form $r^2_{yx_i}(\boldsymbol{X}) = (r^2_{yx_i} + R^2_{yx_i.x_j})/2$, $i \neq j = 1, 2$. It can be derived in the following way. Due to $a_{12} = a_{21}$ and the positive definiteness of $\boldsymbol{A}$, we also have $a_{11} = a_{22}$ because $a^2_{11} + a^2_{12} = a^2_{21} + a^2_{22} = 1$. Also by (2) and (3) the differences between the components of R^2 and the simple squared correlations are equal i.e. $r^2_{yx_1}(\boldsymbol{X}) - r^2_{yx_1} = 2c_1c_2a_{11}a_{12} = r^2_{yx_2}(\boldsymbol{X}) - r^2_{yx_2} = 2c_1c_2a_{21}a_{22}$. Therefore, $r^2_{yx_1} + R^2_{yx_1.x_2} = r^2_{yx_1} + R^2 - r^2_{yx_2} = r^2_{yx_1}(\boldsymbol{X}) + R^2 - r^2_{yx_2}(\boldsymbol{X}) = 2r^2_{yx_1}(\boldsymbol{X})$ and the result follows.

2. $\hat{\boldsymbol{y}}$ forms equal angles with all $\boldsymbol{z}_i$'s, or $c_1 = c_2 = \cdots = c_p$ $(= c_0, \text{say})$. In

this case $\hat{\boldsymbol{y}}$ is in the center of the set $\{\boldsymbol{z}_i\}$ (i.e., $\hat{\boldsymbol{y}} = R/\sqrt{p}\boldsymbol{Z}\mathbf{1}$) and all $r^2_{yx_i}(\boldsymbol{X})$ are equal because $r^2_{yx_i}(\boldsymbol{X}) = \Sigma_j c_j^2 a_{ij}^2 = c_0^2 \Sigma_j a_{ij}^2 = c_0^2 = R^2/p$. Such equality does not necessarily hold for the $r^2_{yx_i}$'s.

3. All $r_{x_i x_j}$'s are equal to (say) r so that the $\boldsymbol{x}_i$'s form a symmetric pencil of vectors with axis of symmetry $\boldsymbol{X}\mathbf{1}_p$. In this case $\boldsymbol{z}_j$ $(j = 1, 2, \ldots, p)$ will form equal angles with all $\boldsymbol{x}_i$, $i \neq j$, and have the same axis of symmetry. Also, it is easily shown that $\boldsymbol{A} = \sqrt{(1-r)}\{\boldsymbol{I}_p + p^{-1}(q-1)\boldsymbol{J}_p\}$, where $q = \sqrt{(1+pr-r)}/\sqrt{(1-r)}$, and $\boldsymbol{J}_p$ is a $p \times p$ matrix of 1's. Therefore, in this case, $r^2_{yx_i}(\boldsymbol{X})$ can also be explicitly expressed in the form $r^2_{yx_i}(\boldsymbol{X}) = (1-r)(q-1)^2 R^2/p^2 + \{1 + 2(q-1)/p\}\{\boldsymbol{y}'\boldsymbol{x}_i - (1-q^{-1})\Sigma_j \boldsymbol{y}'\boldsymbol{x}_j/p\}^2$, $i = 1, 2, \ldots, p$.

(c) Pratt (1987) discussed a symmetrical situation, which is equivalent to assuming both 2. and 3. above. The components of two subsets consisting of m and n regressors, respectively, are in the ratio of $m : n$, as follows from 2., but in a regression on two variables which are the sums of the m and n regressors, respectively, our measure yields components in a ratio which is a monotonic increasing function of m/n, whereas Pratt's measure $\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i$ maintains the ratio $m : n$. Four out of Pratt's six criteria are also satisfied by our measure. The exceptions are his postulate (A3), which requires the above mentioned ratio $m : n$, and (A6), which requires invariance of subsets' components under nonsingular linear transformation of any one of the other subsets. Neither are these postulates satisfied by $\bar{r}_i^2$ of (1). It can be argued from the geometrical point of view, that this assumption is not so "natural" if the subsets are not mutually orthogonal, because change in position of the individual vectors of the transformed subset in relation to other subsets, changes the overall configuration, which is expressed in the coordinate set $\boldsymbol{Z}$ and hence, in our opinion, in the whole decomposition it implies.

(d) Maximum likelihood estimates of ρ_{yx_j} $(i = 1, 2, \ldots, p)$ under a multivariate normal model in which the x_i's are assumed to be uncorrelated, can also be proposed as candidates (after squaring) for components in the decomposition of R^2. They can be derived in the form $\hat{\rho}_{yx_j} = \Sigma_j q(\boldsymbol{X}'\boldsymbol{X})^{ij}\boldsymbol{y}'\boldsymbol{x}_j = \Sigma_j q \boldsymbol{A}^{ij} c_j$, where q is a factor of proportionality and $(\boldsymbol{X}'\boldsymbol{X})^{ij}$ is the ijth element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Since the ith row of $\boldsymbol{A}^{-1}$ is orthogonal to all columns of $\boldsymbol{A}$ except for the ith, these estimates, after squaring, are proportional to $R^2_{yx_i.\mathbf{X}_{-i}}$, which were discussed in Section 2 and rejected from serving as components in the decomposition of R^2. In fact, our aim is to find a decomposition which would assign to each variable its 'fair' portion of R^2 in the presence of nonzero correlations, instead of assuming that they are zero.

(e) The matrices $\boldsymbol{Q}$ and $\boldsymbol{D}_\alpha$ occur also in principal components regression (e.g. Massy (1965) and Mansfield et al. (1977)), which is done to facilitate and simplify calculations in cases of collinearity, and to base upon it considerations about

reducing the number of regressors.

6. Decomposition in ANOVA Problems

If we relax the requirement of linear independence between $\boldsymbol{x}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ then definition (11) of $r^2_{yx_i}(\boldsymbol{X})$ can still hold, with a^{kj} being the kjth element of the generalized inverse of $\boldsymbol{A}$ obtained by inverting the positive elements of $\boldsymbol{D}_\alpha$ and leaving the zero elements unchanged (a Moore Penrose type generalized inverse). This may be advantageous when there is, *a priori*, a "natural" set of regressors into which the variance of $\boldsymbol{y}$ is to be decomposed despite their linear interdependence.

It is well known (cf. Rodgers and Nicewander (1988)) that, in a one-way ANOVA setting, the squared multiple correlation coefficient between y and the columns of the design matrix $\boldsymbol{X}$ is a monotone function of the F statistic testing the hypothesis of equality of the population means. This is an example of the case where the standardized $\boldsymbol{x}_i$'s are linearly dependent. It is interesting to check the meaning of the individual components of R^2 in terms of contrasts between the populations' means. Let the number of populations be k and the sample size in each population be n. We have $R^2_{yx_i.\mathbf{X}_{-i}} = 0$, $i = 1, 2, \ldots, k$, due to the linear dependence, but the three measures (i), (iii) and (iv) from Section 2, namely $r^2_{yx_i}$, $\hat{\boldsymbol{y}}_i'\hat{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i$ are proportional to the same squared standardized contrast: $b_i^2/S_{yy} = n(y_{i.} - y_{..})^2/\Sigma\Sigma(y_{ij} - y_{..})^2$, with $\Sigma b_i^2/S_{yy} = R^2$, using the customary dot notation. This happens because of the special form of $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}_k - \frac{1}{k}\boldsymbol{J}_k$. For $r^2_{yx_i}(\boldsymbol{X})$ we get $r^2_{yx_i}(\boldsymbol{X}) = [R^2/k + (k-2)b_i^2/S_{yy}]/(k-1)$, which is a weighted average of R^2/k and b_i^2/S_{yy}. Note that for $k = 2$ the four measures coincide.

In the situation when $y_{i.}$ coincides with the mean of the other samples, $y_{i.} = y_{..}$, the three measures associating component zero to $\boldsymbol{x}_i$, give the intuitively correct value, because $\boldsymbol{x}_i$ does not contribute to R^2 at all. On the other hand, when $y_{i.}$ is merely close to the mean of the other samples, then the contribution of population i to R^2 should also reflect the increase in R^2 obtained from $\sum_{j\neq i} b_j^2/S_{yy}$ due to its effect on $y_{..}$. Thus, in general $r^2_{yx_i}(\boldsymbol{X})$ is more justified than the other measures in the ANOVA context.

7. A Measure of Deviation from Orthogonality

Given $\boldsymbol{X}$, we found $\boldsymbol{Z}$ that minimizes $\delta = \|\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{Z}'\boldsymbol{X}\|$, over all choices of orthonormal matrices $\boldsymbol{Z}$, which is : $\delta = \|\boldsymbol{D}_\alpha^2 - \boldsymbol{D}_\alpha\| = \sqrt{\Sigma_i(\alpha_i^2 - \alpha_i)^2} = \sqrt{\Sigma_i \alpha_i^2(\alpha_i - 1)^2}$. When the columns of $\boldsymbol{X}$ are orthogonal, $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$ and all $\alpha_i\,(= +\sqrt{\alpha_i^2})$ are 1, hence $\boldsymbol{Z} = \boldsymbol{X}$ and $\delta = 0$. Since $\Sigma\alpha_i^2 = \mathrm{tr}(\boldsymbol{X}'\boldsymbol{X}) = p$, the supremum of δ over different $\boldsymbol{X}$ matrices occurs when $\max_i \alpha_i^2 \to p$ and all

other α_j's $\to 0$. Thus, we can normalize δ by putting $\Delta = \delta/(p - \sqrt{p})$, yielding $0 \leq \Delta < 1$. The deviation of Δ from zero towards 1 can be used as a measure of deviation of the columns of $\boldsymbol{X}$ from orthogonality. Δ is a function of the characteristic roots α_i^2 of the matrix of correlations; therefore it is a meaningful measure in both the sample and the parameter space. Since the decompositions mentioned in Section 2 all agree when $\boldsymbol{X}$ is orthogonal it is interesting to find out in a quantitative way their trend to disagree as Δ approaches 1. This is done in the next Section.

8. Comparisons of Measures: Simulations

Numerical calculations of the various measures on real data sets are of limited value, for the decision as to which of the measures assesses, fairly to the regressors, relative importances in a sense which is relevant to the data, is a matter of subjective judgment.

A small Monte Carlo simulation was carried out to calculate correlation coefficients between components of six measures of relative importance for $p = 2$, 3 and 4, to investigate their closeness under different conditions. 2000 random $\boldsymbol{X}(p+1) \times p$ matrices and random $p \times 1$ $\boldsymbol{c} = \boldsymbol{z}'\boldsymbol{y}$ vectors were generated, all i.i.d. $U[0,1]$. In each case the measure of deviation from orthogonality, Δ, and the components for six measures were calculated.

Table 1 shows the number (n) of cases that fall in each Δ range and the average of correlation coefficients calculated for each sample between the components $r^2_{yx_i}(\boldsymbol{X})$ and each of the following measures: SIMPLE $= r^2_{yx_i}$, MARGINAL $= R^2_{yx_i.\mathbf{X}_{-i}}$, STANDARD $= \hat{\boldsymbol{y}}_i'\hat{\boldsymbol{y}}_i$, AVERAGE $= \bar{r}_i^2$ from (1), PRATT $= \hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i$ and NNPRATT, which is the same as PRATT, but includes only the cases when all the components are non-negative. The standard deviations are also presented. For $p = 2$, all correlation coefficients were 1, indicating a complete agreement between all measures concerning the order of importance of the two regressors. Therefore, it was omitted from the table. However, the average mean square of deviations between $r^2_{yx_i}(\boldsymbol{X})$ and the other measures changed, for $p = 2$, with increasing Δ from .0004 to .04, with small differences among the six measures. For $p > 2$, even the same ordering with different relative components results in $r < 1$, and the value of r reflects the magnitude of agreement among the measures.

It is evident that as $\Delta \to 1$ there is a sharp decrease in the correlation coefficients. They also decrease somewhat with increasing p, especially for small Δ values. Among the measures, AVERAGE is closest to $r^2_{yx_i}(\boldsymbol{X})$ followed by NNPRATT, PRATT, SIMPLE, MARGINAL and STANDARD in decreasing order. Note that n_{NN}, the number of cases for NNPRATT, decreased sharply with increasing p and Δ, indicating that some negative components for PRATT occur quite frequently. The table shows that for small p and Δ, several measures

Table 1. Average correlation coefficients (±standard deviations) between the components of the proposed measure of relative importance and those of six other measures from a Monte Carlo study with $p = 3$ and 4 and at different ranges of the nonorthogonality measure Δ. The number of cases (n) at each Δ range is also presented.

$\Delta =$	.0–0.3	.3–0.4	.4–0.5	.5–0.6	.6–0.7	.7–0.8	.8–0.9	.9–1.0
measure*				$p = 3$				
SIMPLE	.94±.22	.70±.53	.67±.58	.67±.56	.70±.54	.62±.62	.74±.50	.18±.67
MARGINAL	.97±.10	.80±.40	.64±.61	.63±.60	.61±.60	.52±.70	.49±.74	.66±.60
STANDARD	.97±.07	.70±.48	.46±.70	.54±.62	.45±.64	.34±.73	.28±.73	.44±.63
AVERAGE	1.00±.01	.85±.41	.84±.44	.82±.44	.83±.43	.80±.49	.88±.32	.83±.31
PRATT	1.00±.01	.88±.28	.73±.41	.74±.34	.71±.32	.64±.41	.57±.43	.53±.46
NNPRATT	1.00±.01	.90±.27	.81±.40	.77±.33	.79±.33	.56±.57	.47±.64	.32±.55
$n(n_{NN})$	125(100)	456(174)	554(160)	345(85)	227(38)	146(32)	107(13)	40(8)
				$p = 4$				
SIMPLE	.77±.36	.71±.41	.67±.44	.65±.48	.65±.50	.63±.46	.70±.47	
MARGINAL	.73±.44	.62±.51	.58±.52	.58±.52	.54±.53	.45±.62	.16±.52	
STANDARD	.60±.49	.49±.54	.44±.54	.44±.51	.39±.49	.41±.52	.06±.69	
AVERAGE	.93±.19	.91±.22	.90±.24	.91±.20	.91±.15	.89±.23	.96±.05	
PRATT	.82±.31	.73±.33	.70±.31	.65±.32	.60±.33	.66±.30	.45±.35	
NNPRATT	.90±.25	.83±.29	.80±.30	.70±.37	.66±.45	.87±.08	–	
$n(n_{NN})$	316(92)	673(120)	607(77)	282(29)	97(9)	24(2)	7(0)	

*SIMPLE = $r^2_{yx_i}$, MARGINAL = $R^2_{yx_i.\mathbf{X}_{-i}}$, STANDARD = $\hat{\boldsymbol{y}}_i'\hat{\boldsymbol{y}}_i$, AVERAGE = $\bar{r}_i^2$ PRATT = $\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}_i$; NNPRATT is the same as PRATT, but discards all cases when any component is negative; n_{NN} is the number of cases for NNPRATT.

give essentially the same decomposition and any of them may be used to assess the relative importance of the regressors. For larger p or Δ the measures vary considerably, and one should use a measure only after its relevance has been established.

In the example of Yule's data, for which Pratt (1987) calculated the components of his measure, there are $p = 3$ regressors and $\Delta = 0.34$. Hence, we found that, in accordance with Table 1, all the above measures are highly correlated and gave similar estimates of relative importances.

Acknowledgements

The author is very grateful to the referees for their many helpful comments and suggestions. This research was partly sponsored by the US-Israel Binational Agricultural Research and Development Fund, Project No. US-334-81. Partial

assistance was obtained from Natural Sciences and Engineering Research Council of Canada.

References

Kruskal, W. (1984). Relative importance of determiners. *Questiio* **8**, 39–45.

Kruskal, W. (1987). Relative importance by averaging over orderings. *Amer. Statist.* **41**, 6–10.

Kruskal, W. and Majors, R. (1989). Concepts of relative importance in recent scientific literature. *Amer. Statist.* **43**, 2–6.

Mansfield, E. R., Webster, J. T. and Gunst, R. F. (1977). An analytic variable selection technique for principal component regression. *Appl. Statist.* **26**, 34–40.

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *J. Amer. Statist. Assoc.* **60**, 234–256.

Pratt, J. W. (1987). Dividing the indivisible: using simple symmetry to partition variance explained. In *Proceedings of Second International Tampere Conference in Statistics* (Edited by T. Pukkila and S. Puntanen), 245–260. University of Tampere, Tampere, Finland.

Rodgers, J. L. and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *Amer. Statist.* **42**, 59–66.

Williams, E. J. (1978). Postscripts to "Linear hypothesis: regression". In *International Encyclopedia of Statistics* (Edited by W. H. Kruskal and J. M. Tanur), 537–541. Free Press, New York, NY.

Department of Statistics and Operations Research, Agricultural Research Organization, Bet Dagan 50250, P.O. Box 6, Israel.

(Received June 1991; accepted December 1992)