PSYCHOMETRIKA—VOL. 41, NO. 4
DECEMBER, 1976

REGRESSION WITH QUALITATIVE AND QUANTITATIVE VARIABLES: AN ALTERNATING LEAST SQUARES METHOD WITH OPTIMAL SCALING FEATURES

FORREST W. YOUNG

UNIVERSITY OF NORTH CAROLINA

JAN DE LEEUW

RIJKSUNIVERSITEIT TE LEIDEN

YOSHIO TAKANE

UNIVERSITY OF NORTH CAROLINA

A method is discussed which extends canonical regression analysis to the situation where the variables may be measured at a variety of levels (nominal, ordinal, or interval), and where they may be either continuous or discrete. There is no restriction on the mix of measurement characteristics (i.e., some variables may be discrete-ordinal, others continuous-nominal, and yet others discrete-interval). The method, which is purely descriptive, scales the observations on each variable, within the restriction imposed by the variable's measurement characteristics, so that the canonical correlation is maximal. The alternating least squares algorithm is discussed. Several examples are presented. It is concluded that the method is very robust. Inferential aspects of the method are not discussed.

Key words: linear model, canonical analysis, measurement, multivariate data, survey data, successive block algorithm.

In many areas of social and behavior sciences, the investigator has obtained observations on a variety of variables from a number of respondents and wishes to investigate their structure with the use of multiple (or, occasionally, canonical) correlation techniques. But the investigator has a problem.

This research was partially supported by grants MH10006 and MH26504 from the National Institute of Mental Health to the Psychometric Laboratory of the University of North Carolina. Portions of the research reported here were presented to the spring meeting of the Psychometric Society, 1975. We wish to thank John B. Carroll and Elliot M. Cramer for their critical evaluations of an earlier draft of this report, and Jack Hoadley and John B. Carroll for letting us use their data. Copies of the paper may be obtained from the first author.

Jan de Leeuw is currently at Datatheorie, Central Rekeninstituut, Wassenaarseweg 80, Leiden, The Netherlands. Yoshio Takane can be reached at the Department of Psychology, University of Tokyo, Tokyo, Japan.

Requests for reprints should be sent to Forrest W. Young, Psychometric Laboratory, University of North Carolina, Davie Hall 013 A, Chapel Hill, North Carolina 27514.

He knows that his variables do not measure up, as it were, to the assumptions required by the multiple correlation technique. Some of his variables are measured at the interval level, as is conventionally required, but others are ordinal and yet others are (heaven forbid) only nominal. What does he do in the face of this dilemma? He proceeds as though there were no problem at all. After all, what other alternatives are there? There is no completely appropriate alternative, as the only multiple correlation techniques which permit qualitative variables assume that all variables are qualitative, or that the independent variables are quantitative while the dependent variable is qualitative. In the former situation the variables must all be nominal [Hayashi, 1950], or all ordinal [Lingoes, 1973; de Leeuw, Note 2], with no mixing of levels and no quantitative variables. In the latter situation the dependent variable may be ordinal [de Leeuw, Note 3; Carroll, 1972; Srinivasan, Note 7], nominal (discriminant analysis) or binary [de Leeuw, Note 4]. With the exception of discriminant analysis, no procedures have been proposed which permit several dependent variables that may be qualitative.

The work presented in this paper is designed with the above situation in mind. With the Multiple Optimal Regression by Alternating Least Squares (MORALS) technique, and the corresponding canonical regression technique (CORALS), the investigator with variables defined at a variety of measurement levels can investigate their structure while he respects the various levels of measurement. MORALS optimizes the multiple correlation between a single criterion variable and a set of predictor variables where any of the variables (criterion included) may be nominal, ordinal or interval. The variables do not all have to be measured at the same level; any mixture will do. Also, the process assumed to have generated the data may be either discrete or continuous. As will be explained, MORALS obtains an optimal scaling for each variable within the restrictions imposed by the regression model, the measurement level, and the generating process. The scaling is optimal in the Fisher [1938] sense of optimal scaling: the multiple correlation is maximized.

CORALS is very similar to MORALS, except that it optimizes the canonical correlation between two sets of variables. Since MORALS is a special case of CORALS (where one of the sets of variables consists of a single variable), we refer to the algorithm as the MORALS/CORALS algorithm.

In the companion paper [de Leeuw, Young & Takane, 1976] we discussed in detail the simple additive situation where one has obtained qualitative data in a factorial design and is only interested in the main effects. The previous work, then, was restricted to univariate qualitative data obtained in factorial experiments, and to the additive model with no interaction terms.

In this paper we remove all of these limitations. We discuss the general linear model as it applies to multivariate data where the variables may be defined at any mixture of measurement levels.

As was discussed in the companion paper, the analysis of additivity has usually been introduced in the context of a statistical model which places very stringent and oftentimes unrealistic assumptions on the nature of the data. As a result of these assumptions it is possible to develop inferential procedures. Our purpose is not to develop an inferential procedure, but rather to develop a descriptive procedure to provide the investigator with tools for investigating the additive structure of his qualitative data. While it would be desirable to develop an inferential procedure based on assumptions commensurate with qualitative variables, such a development is not presented in this paper.

1. Mathematical Developments

Notation

We use bold-face capital letters to represent matrices (i.e., $\mathbf{X}$); bold-face lower case letters for vectors ($\mathbf{x}$); and regular lower case letters for scalars (x). Note that all vectors are assumed to be column vectors, with a row vector denoted as the transpose of a column vector ($\mathbf{x}'$). We refer to a specific column vector of a matrix as $\mathbf{x}_i$, a specific element of a matrix as x_{ai} ; and a specific element of a vector as x_a . We further reserve Greek letters for parameters (i.e., β is a vector of parameters) and script letters for functions (i.e., $\mathfrak{I}$ is a transformation).

The problem

Let there be two matrices of observation variables $\mathbf{X}$ and $\mathbf{Y}$, there being n variables $\mathbf{x}_i$ and m variables $\mathbf{y}_j$, all having k observations (we will consider missing data later). We assume that each $\mathbf{x}_i$ and $\mathbf{y}_j$ is measured at some known measurement level, with each element being subject to certain measurement restrictions. There need be no particular relationship of measurement levels between variables, although we assume that all observations on a single variable are at the same measurement level (an assumption to be relaxed later).

Let us further define two vectors of parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and two matrices $\mathbf{X}^*$ and $\mathbf{Y}^*$, where $\boldsymbol{\alpha}$ has n elements α_i , $\boldsymbol{\beta}$ has m elements β_i , $\mathbf{X}^*$ has n columns and k rows, and $\mathbf{Y}^*$ has m columns and k rows. The columns $\mathbf{x}_i^*$ and $\mathbf{y}_i^*$ correspond to the observation variables $\mathbf{x}_i$ and $\mathbf{y}_j$. Furthermore, the columns $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ have two important characteristics: *a*) all $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ are defined at the interval level of measurement; and *b*) each $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ is related to its observation variable $\mathbf{x}_i$ and $\mathbf{y}_j$ by a transformation which completely

satisfies the measurement characteristics of the observation variable. That is,

$$\begin{aligned} \mathbf{x}_i^* &= \mathfrak{Z}_i[\mathbf{x}_i], \\ \mathbf{y}_j^* &= \mathfrak{Z}_j[\mathbf{y}_j], \end{aligned} \tag{1}$$

where $\mathfrak{Z}_i$ and $\mathfrak{Z}_j$ are called the measurement transformations, and are collectively referred to as $\mathfrak{Z}$ when there is no resulting confusion. Of course, the $\mathfrak{Z}$ are subject to restraints by the measurement level and process of their variables. These restraints are discussed in the next section. We may correctly think of $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ as being the observation variables rescaled at the interval level of measurement so that the correlation is optimized. Thus, we will oftentimes refer to $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ as the optimally scaled observations.

The problem we wish to solve can now be stated. We wish to obtain transformations $\mathfrak{Z}$ of each observation variables $\mathbf{x}_i$ and $\mathbf{y}_j$, as well as regression weights α and β, so that the canonical correlation between $\mathbf{X}^*$ and $\mathbf{Y}^*$ is as high as possible. Since maximizing the canonical correlation is equivalent to minimizing the sums of squared differences between two composite variables (under suitable normalization assumptions), we defined composite variables $\mathbf{a}$ and $\mathbf{b}$ such that

$$\begin{aligned} \mathbf{a} &= \mathbf{X}^*\alpha, \\ \mathbf{b} &= \mathbf{Y}^*\beta, \end{aligned} \tag{2}$$

and state our goal as the minimization of

$$\begin{aligned} \lambda^2 &= (\mathbf{a} - \mathbf{b})'(\mathbf{a} - \mathbf{b}), \\ &= (\mathbf{X}^*\alpha - \mathbf{Y}^*\beta)'(\mathbf{X}^*\alpha - \mathbf{Y}^*\beta), \end{aligned} \tag{3}$$

where the normalization restrictions are

$$\begin{aligned} &\text{(a)} \quad \mathbf{1}'\mathbf{x}_i^* = \mathbf{1}'\mathbf{y}_j^* = 0, \qquad (i = 1, \cdots, n;\ j = 1, \cdots, m), \\ &\text{(b)} \quad \frac{1}{k}\mathbf{x}_i^{*\prime}\mathbf{x}_i^* = \frac{1}{k}\mathbf{y}_j^{*\prime}\mathbf{y}_j^* = 1, \\ &\text{(c)} \quad \frac{1}{k}\mathbf{a}'\mathbf{a} = 1, \\ &\text{(d)} \quad \frac{1}{k}\mathbf{b}'\mathbf{b} = 1. \end{aligned} \tag{4}$$

Of course, $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ are also subject to the measurement restrictions of (1). Note that we minimize λ and not λ normalized by the variance of $\mathbf{a}$ (or $\mathbf{b}$). De Leeuw [Note 5] has shown that restriction (*c*) makes such a normalization unnecessary. Note that the canonical correlation between the two sets of variables is defined as

$$R = \frac{1}{k}\mathbf{a}'\mathbf{b}.$$

Restrictions

In this section, we discuss the types of restrictions imposed on the relationship between the raw observations in matrices **X** and **Y** and the rescaled observations represented by matrices **X*** and **Y***; that is, we discuss restrictions of the transformations in (1). The restrictions are of three types; *a*) those concerned with identifying the parameters α and β (the normalization restrictions); *b*) those concerned with the measurement level of the observation variables; and *c*) those concerned with the underlying process which generated the observations on each variable.

The minimal set of restrictions always involves the normalization restrictions imposed by (4). Though these restrictions are trivial and not very interesting, they are of crucial importance since without them unique values for α and β are undefinable. Of greater interest are the restrictions on the transformations $\mathfrak{I}$ (equation 1). These restrictions concern two different aspects of the measurement situation: the level of measurement of the observation variable (i.e., nominal, ordinal or interval), and the nature of the process which generated the observations (discrete or continuous). We do not go into these restrictions in detail here since they have already been discussed in detail in the companion papers [de Leeuw, Young & Takane, 1976; Takane, Young & de Leeuw, in press]. We simply state that the measurement process restrictions involve either the discrete restriction

$$\mathfrak{I}^d : (x_{ai} \sim x_{bi}) \rightarrow (x_{ai}{}^* = x_{bi}{}^*), \tag{5}$$

(where $\sim$ indicates empirical equivalence, i.e., membership in the same observation category) or the continuous restriction

$$\mathfrak{I}^c : (x_{ai} \sim x_{bi}) \rightarrow (x_{ai}{}^- = x_{bi}{}^-) \leq \begin{Bmatrix} x_{ai}{}^* \\ x_{bi}{}^* \end{Bmatrix} \leq (x_{ai}{}^+ = x_{bi}{}^+), \tag{6}$$

(where $x_{ai}{}^-$ and $x_{ai}{}^+$ are lower and upper bounds of an interval of real numbers). Note that the same types of restrictions apply to Y and Y^*.

The measurement level restrictions involve *a*) no added restrictions for nominal variables; *b*) order restrictions

$$\mathfrak{I}^0 : (x_{ai} < x_{bi}) \rightarrow (x_{ai}{}^* \leq x_{bi}{}^*), \tag{7}$$

for the ordinal measurement level; and *c*) linear restrictions

$$\mathfrak{I}^l : x_{ai}{}^* = \delta_0 + \delta_1 x_{ai}\,, \tag{8}$$

or polynomial restrictions

$$\mathfrak{I}^p : x_{ai}{}^* = \sum_{q=0}^{p} \delta_q x_{ai}{}^q, \tag{9}$$

for the interval level of measurement.

Model subspace

The final notion to be introduced is that of the model subspace represented by the matrices $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, which are of the same order as the observation matrices $\mathbf{X}$ and $\mathbf{Y}$. The model subspace notion is most easily introduced by expanding (3) to

$$\lambda^2 = [\mathbf{x}_l{}^*\alpha_l - (\mathbf{b} - \mathbf{a} + x_l{}^*\alpha_l)]'[\mathbf{x}_l{}^*\alpha_l - (\mathbf{b} - \mathbf{a} + x_l{}^*\alpha_l)]. \tag{10}$$

If we define

$$\tilde{\mathbf{x}}_l\alpha_l = [\mathbf{b} - (\mathbf{a} - \mathbf{x}_l{}^*\alpha_l)], \tag{11}$$

then

$$\begin{aligned} \lambda^2 &= [\mathbf{x}_l{}^*\alpha_l - \tilde{\mathbf{x}}_l\alpha_l]'[\mathbf{x}_l{}^*\alpha_l - \tilde{\mathbf{x}}_l\alpha_l] \\ &= \alpha_l{}^2(\mathbf{x}_l{}^* - \tilde{\mathbf{x}}_l)'(\mathbf{x}_l{}^* - \tilde{\mathbf{x}}_l). \end{aligned} \tag{12}$$

Note that the model subspace vector $\tilde{\mathbf{x}}_l$ is proportional to the difference between the two linear components $\mathbf{a}$ and $\mathbf{b}$ when the optimally scaled observations on variable l are removed from the equation:

$$\tilde{\mathbf{x}}_l = [\mathbf{b} - (\mathbf{a} - \mathbf{x}_l{}^*\alpha_l)]\alpha_l{}^{-1}. \tag{13}$$

This explains the name of these variables. Of course, we may define $\tilde{\mathbf{y}}_l$ correspondingly.

2. Algorithm

The MORALS/CORALS algorithm is an alternating least squares (ALS) algorithm. A very closely related ALS algorithm which is appropriate to the ANOVA situation (ADDALS) has been described in the companion paper [de Leeuw, Young, & Takane, 1976]. An ALS algorithm for individual differences multidimensional scaling (ALSCAL) has also been discussed by Takane, Young, and de Leeuw [in press]. These same investigators are also developing ALS algorithms for principal components analysis. All of these algorithms have in common the fact that the data being analyzed may be at a variety of measurement levels. The ALS approach is related to the work of Wold and Lyttkens [1969], de Leeuw [Note 3] and Young [1972].

As is implied by the name, an ALS algorithm is an interative algorithm which alternatives back and forth between two phases, each of which is a least squares procedure. In one of the phases, least squares estimates for the model parameters are obtained while the data transformations are held constant, whereas in the other phase, least squares estimates of the transformations are obtained while the model parameters are held constant. It has been shown by de Leeuw, Young and Takane [1976] that under the appropriate conditions such an iterative process is convergent. The MORALS/CORALS algorithm is such an algorithm. In one of the phases the least

squares estimates of the model parameters α and β are obtained while the transformations $\mathfrak{z}$ are held constant, while in the other phase the least squares estimate of a single transformation $\mathfrak{z}_i$ is obtained while α, β, and the remaining $\mathfrak{z}$ are fixed. Since both phases minimize (3), the algorithm is convergent. In the next two sections we discuss the model estimation phase (which estimates α and β) and then the optimal scaling phase (which estimates $\mathfrak{z}_i$).

We should point out that the model estimation phase is entirely superfluous. It is possible to recast our problem in the indicator matrix format used in our work with the additive model [de Leeuw, Young & Takane, 1976] and to avoid the transformations notation entirely. When viewed in this light it is apparent that the regression weights are unnecessary, as they are absorbed into the scaling of each variable. However, it is our opinion that the presentation mode used here is more desirable pedagogically in the present context.

Model Estimation Phase

In the model estimation phase we desire to obtain least squares estimates (relative to λ^2) of the model parameters α and β under the assumption that all the optimally scaled variables $\mathbf{X}^*$ and $\mathbf{Y}^*$ are held constant. When it is recalled that the $\mathbf{X}^*$ and $\mathbf{Y}^*$ represent the observations rescaled at the interval level, we see that this phase is no different than the classical canonical correlation situation [Hotelling, 1935], where the variables are the $\mathbf{x}_i^*$ and $\mathbf{y}_j^*$ (not the $\mathbf{x}_i$ and $\mathbf{y}_j$). If we define $\mathbf{R}_{xx}$ to be the correlations among the variables in $\mathbf{X}^*$, and define R_{xy} and R_{yy}, correspondingly, then the least squares solution to the canonical correlation equation

$$\mathbf{X}^*\alpha \cong \mathbf{Y}^*\beta, \tag{14}$$

may be obtained by finding the square root of the largest latent root μ of

$$(\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy} - \mu\mathbf{I})\beta = \mathbf{0}. \tag{15}$$

It is then the case that the vector of weights β is the characteristic vector associated with the largest root of (15). The weight vector α is obtained from

$$\alpha = (\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\beta)^{-1/2}. \tag{16}$$

In the multiple correlation situation where, for example, there is only one variable $\mathbf{x}^*$, the least squares estimate of

$$\mathbf{x}^* = \mathbf{Y}^*\beta, \tag{17}$$

for β reduces to

$$\beta = (\mathbf{Y}^{*\prime}\mathbf{Y}^*)^{-1}\mathbf{Y}^{*\prime}\mathbf{x}^*. \tag{18}$$

Optimal scaling phase

We will not discuss the details of the optimal scaling phase as they are the same as in the companion papers [de Leeuw, Young & Takane, 1976; Takane, Young & de Leeuw, in press].

Initial values

As with any iterative algorithm, we must supply the procedure with values to initiate the process, but with MORALS, unlike some algorithms, there is a particularly compelling initialization procedure. We simply assume that the matrices $\mathbf{X}$ and $\mathbf{Y}$ (the raw data) are actually the matrices $\mathbf{X}^*$ and $\mathbf{Y}^*$. This is equivalent to assuming, for the initialization process, that the raw data are measured on an interval scale. (We must assign arbitrary values to the observation categories when a variable is nominal.) We then enter the model estimation phase and solve for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by (15) and (16) (or (18), as the case may be). We then use these estimates of the model parameters to begin the iterative procedure whose flow is discussed in the next section. Note that the initialization procedure is simply the classical canonical or multiple regression procedure. Thus, the very first step of the procedure corresponds to the analysis the researcher would have obtained had he decided that his variables were all quantitatively measured. By comparing the results of the entire iterative process with the results of the very first step of the process, the user can determine what he has gained by the use of his measurement assumptions (except for nominal variables). If, for example, he assumed that his variables were ordinal and discovers that the canonical correlation is essentially the same after the iterative process as it was before, and he also discovers that his monotonic transformations are all essentially linear, then he could safely conclude that he has gained nothing by the assumptions of ordinality, and that it would have been proper to assume that his variables are measured at the interval level. Thus, we see that by using the classical procedure as the initialization process the user can easily investigate whether his assumptions concerning measurement levels are correct (see Takane, Young & de Leeuw, in press, for a discussion of this point).

Algorithm Flow

As described thus far, the MORALS/CORALS procedure consists of two phases, one to estimate the model parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and the other to estimate the data transformation $\mathfrak{I}_i$. The obvious ALS procedure would involve alternating these two phases until convergence is obtained. A careful reader, however, will have by now detected a subtle imbalance between the two phases. Whereas the model estimation phase obtains estimators of the weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for *all* variables, the optimal scaling phase obtains estimates of the transformation *one* variable at a time. Thus, the obvious ALS procedure

must be modified somewhat. Several of the alternatives that we might choose are discussed in the remainder of this section.

One alternative is an iterative procedure where a single iteration is defined as follows. First select a particular variable and optimize λ^2 relative to that variable (as expressed by (13)) by repetitively alternating the two phases until convergence is obtained. Then repeat this procedure for another variable and another, etc., until all variables have been subjected to this procedure. This completes a single iteration, with the iterations being repeated until convergence is obtained. We have not investigated this procedure as it appears to us to be computationally inefficient relative to some of the procedures discussed below.

Another alternative would define an iteration as follows: obtain the transformation $\mathfrak{I}$ which yields the optimal $\mathbf{x}_1^*$. Then replace the previous $\mathbf{x}_1^*$ with the new one, and repeat this process for $\mathbf{x}_2^*$, $\mathbf{x}_3^*$, etc., until all the variables in both sets have been optimally scaled. There is now an entirely new set of $\mathbf{X}^*$ and $\mathbf{Y}^*$, for which the transformation process is repeated. We repeat the transformation estimation until convergence is obtained at which point we have already obtained estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as a by-product of the procedure. We have investigated this algorithm and have found it to be very slow. Thus, we do not pursue it further, although it can be shown to converge on the desired $\mathbf{X}^*$ and $\mathbf{Y}^*$ and on the appropriate weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

The two algorithms that we have investigated most thoroughly are each rather similar to the one just discussed. The iterative structure for the first of these two procedures is as follows. First obtain the transformation $\mathfrak{I}$ which yields the optimal $\mathbf{x}_1^*$, and then replace the previous $\mathbf{x}_1^*$ with the new one. Repeat this process for each of the other $\mathbf{x}_i^*$ ($i = 1, 2, 3 \cdots n$); and solve for weights α_i and β_j. Follow this with estimation and replacement of each $\mathbf{y}_j^*$ ($j = 1, 2, \cdots m$) and another estimation of α_i and β_j. The other algorithm that we have investigated is exactly the same except that the weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are estimated only once on each iteration. Thus, the structure of an iteration for this procedure is to estimate and replace each $\mathbf{x}_i^*$ ($i = 1, 2, \cdots n$), and then each $\mathbf{y}_j^*$ ($j = 1, 2, \cdots m$), and then to estimate the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Both of these algorithms entail an iteration which is much quicker than the first possibility discussed in the previous paragraphs and slightly slower than the second possibility. Many fewer iterations are required with these algorithms than with the second alternative. More important, however, is the fact that these two algorithms place more nearly equal emphasis on the two phases than the former two algorithms. We have been unable to determine any characteristics of the latter two algorithms which would allow us to select one over the other except for the fact that the one involving only one model estimation is slightly more efficient. Thus, we have chosen it to define the iterative flow of MORALS/CORALS.

Missing data

Missing data are allowed for in a manner which does not destroy the ALS property of the MORALS/CORALS algorithm. If some observation x_{ai} (or, implicitly, y_{ai}) is missing, then the computation of the initial weights $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ is changed in a minor manner: we simply estimate each missing x_{ai}* as being the mean of the nonmissing observations in the vector $\mathbf{x}_i$. The computation of initial $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ then proceeds as stated.

The computation of the subspace vector $\tilde{\mathbf{x}}_i$ is as stated above except for the element $\tilde{x}_{ai}$ corresponding to the missing observation. The value of this subspace element is determined by either *a*) setting it equal to the mean of $\mathbf{x}_i$*; or *b*) setting it equal to the element x_{ai}*. The choice of method is left to the user and corresponds to whether he views his missing observations as having been caused by a discrete or continuous process, respectively. If the discrete assumption seems appropriate, then all missing observations on a given variable are assumed to have been caused by the same (discrete) set of events, and are all assigned the mean of the nonmissing values (the least squares discrete estimate). If the continuous assumption is appropriate, then all missing observations on a given variable are assumed to have been caused by different (continuous) sets of events and are all set equal to their corresponding optimally scaled observation (the least squares continuous estimate). Alternatively, we may view the choice as concerning whether all missing observations represent a single category (form means) or separate categories (don't form means).

Partitions

It may sometimes be the case that not all of the observations made on a single variable are comparable. For example, it may be that some of the observations of a variable were made at a different time or under somewhat different conditions than some of the other observations. The result is that a specific observation made at one time (or place) cannot be said to be larger or smaller than one made at the other time (or place), even though the observations have the same measurement characteristics. Or, as another example, the measuring device may have broken during the observation process and been repaired in such a way that the measurements before and after the break-down are no longer directly comparable. If it is the case that not all observations on a specific variable are directly comparable, then we wish to partition the observations into mutually exclusive and exhaustive subsets, and permit separate transformations within each partition.

The partition notion is also useful in precisely the opposite situation. It may be that we have two (or more) variables which are measured on the same scale and that we wish to obtain identical optimal transformations of both variables. In this case we wish to collect the observations on the two

variables into a single superset of observations and obtain a single transformation for the superset.

It should be apparent that defining partitions as either subsets of the observations on a variable or as supersets of observations creates no particular problem. We need only to substitute the desired partitioning of **X** and **Y** for the partitioning which has been implicit in our previous developments (i.e., partitioning by variables).

3. Evaluation

In this section we present three evaluations of the MORALS/CORALS algorithm. The first evaluation involves a Monte Carlo study in which we find that the algorithm can recover known information in the face of systematic and random error. In the second evaluation we obtain useful and interpretable results when the algorithm is applied to a set of real survey data. In the third and final evaluation we obtain a meaningful interpretation of a multidimensional scaling solution in a special case of MORALS/CORALS which corresponds to a previously proposed procedure for interpreting such solutions.

Monte Carlo study

A small Monte Carlo study designed to investigate the robustness of MORALS/CORALS in the face of both systematic and random error is presented in this section. We do not claim that this is a complete or definitive study of the algorithm's behavior in situations likely to be encountered by the typical investigator. Rather, it is simply a small study designed to demonstrate the algorithm's behavior in one common situation, that in which there is a single dependent variable and two independent variables, all of which are ordinal. The design of the study is as follows.

Two "true" independent variables were generated, each with 64 observations. Each observation was sampled from a random uniform distribution on the interval (0–1). A "true" dependent variable was then generated by simply adding together corresponding elements of the independent variables.

Five degrees of nonlinear monotonic distortion and three levels of random error were then defined. The nonlinear distortion involved the three transformations:

$$
\begin{aligned}
y_i &= \bar{y}_i{}^k, \\
x_{i1} &= \bar{x}_{i1}{}^{1/k}, \\
x_{i2} &= (\bar{x}_{i2} - \bar{x}_{\cdot 2})^k \cdot \mathrm{SIGN}\ (\bar{x}_{i2} - \bar{x}_{\cdot 2}),
\end{aligned}
$$

where k varies from 1 to 5, a "bar" indicates "true" values, and $x_{\cdot 1}$ indicates the mean of the true values. Thus when $k = 1$, all variables remain undistorted, and when k varies up to 5, the variables become increasingly more

nonlinear, with each variable being distorted in its own unique manner. Note, however, that all distorted variables are equivalent to the undistorted variables at the ordinal level, since each transformation is strictly monotonic. Random error was introduced by adding a random normal deviate to each element of each of the three distorted variables, where the standard deviation of the deviate is proportional to the standard deviation of the undistorted variable. The proportionality constant defined the three levels of random error, and was equal to zero for the lowest level (no random error), .10 for the intermediate level, and .50 for the highest level. These systematically distorted and randomly perturbed variables were then input to MORALS/CORALS, under the assumption that each variable was ordinal. The results are summarized in Table 1 and Figure 1. The table presents the classical multiple correlation coefficient R_c (which is the one obtained under the interval measurement level assumptions made in the standard multiple correlation analysis, and which is also the initial correlation obtained by MORALS/CORALS), the optimal multiple correlation coefficient R_0 (obtained under ordinal measurement level assumptions), the simple correlation of each optimally scaled variable with its corresponding "true" variable (r_y , r_{x1} , and r_{x2}), and the number of iterations required for convergence of the algorithm. The figure presents the optimal transformations obtained for each variable for the fourth degree of systematic distortion and for all three levels of random perturbation. The horizontal axes are the distorted and perturbed values being analyzed, and the vertical axes are the optimally transformed values. Ideally, the optimal transformations presented in Figure 1 should be the inverse of the transformations used to define the systematic distortion.

We observe that for all levels of systematic distortion the algorithm can perfectly recover the underlying true structure when there is no random error (all optimal multiple correlation are 1.0 and all simple correlations are at least .976). This implies that the algorithm is able to obtain the desired inverse of each transformation defined above, as otherwise the simple correlations would be lower. Note also that in these cases with no random error the classical multiple correlation progressively decreases from 1.0 in the no-distortion case to .626 in the case of most distortion, as is expected.

We also observe that when random error is present, the degree to which the true information is recovered deteriorates, as is expected. For the intermediate level of random error the amount of deterioration does not seem to be affected much by the amount of systematic distortion. However, there appears to be some effect of systematic error for the greatest degree of random error, with the overall poorest recovery occurring when there is a large amount of random error and severe non-linearities in the data. In particular we note that for high random error and for systematic error levels 3 and 5 (but not 4) the solution obtained by MORALS/CORALS is degenerate. The

Table 1

Monte Carlo Study

Random Error Level		Systematic Distortion Level				
		1	2	3	4	5
.00	R_c	1.0000	.9213	.8081	.7079	.6255
	R_o	1.0000	1.0000	1.0000	.9999	.9999
	r_y	1.0000	.9917	.9861	.9833	.9831
	r_{x1}	1.0000	.9869	.9792	.9761	.9763
	r_{x2}	1.0000	.9943	.9918	.9904	.9898
	iter	1	5	7	8	9
.10	R_c	.9882	.9069	.7832	.6939	.5921
	R_o	.9995	.9999	.9994	.9996	.9998
	r_y	.9942	.9867	.9810	.9517	.9869
	r_{x1}	.9931	.9820	.9796	.9139	.9800
	r_{x2}	.9906	.9851	.9814	.9748	.9864
	iter	5	6	13	11	12
.50	R_c	.8486	.7043	.5944	.6498	.5228
	R_o	.9544	.9269	.9998	.9545	.9997
	r_y	.9177	.8753	.3339	.9214	.2721
	r_{x1}	.8801	.8709	.8927	.8805	.8126
	r_{x2}	.8849	.9058	.2785	.9099	.2380
	iter	6	5	21	6	23

optimal transformations obtained for the dependent variable and for the second independent variable consist, essentially, of two values, with the largest value on each vector being transformed into a very large value, and all the remaining values being tied together at a very small value. The four corresponding simple correlations are much lower than any others. It is not clear to us why such a degeneracy should occur for these two particular levels of systematic distortion and not for the intervening level (see Figure 1). However, these results should serve as a warning that in cases of extreme nonlinearities and high amounts of error the MORALS/CORALS algorithm may yield degenerate results. In our experience with real data, however,

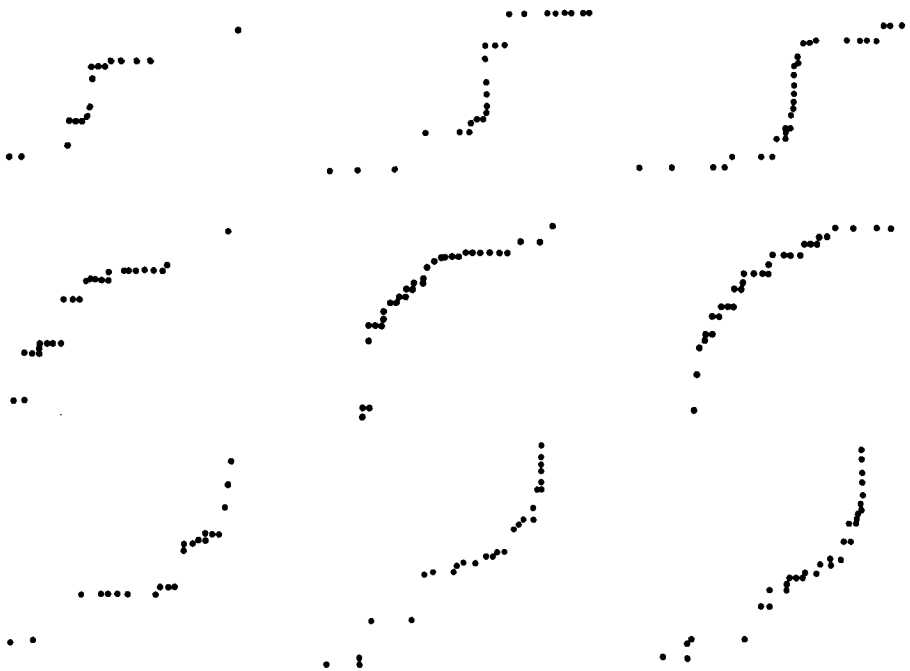

FIGURE 1
Recovered transformations for three levels of random error at the fourth level of systematic distortion.

we have never obtained such degeneracies, suggesting that these extreme Monte Carlo conditions may be uncommon empirically.

Survey data

The application process for prospective graduate students to a quantitative psychology program can be summarized, from the faculty's point of view, by the variables presented in Table 2. Of the 12 variables presented in this table, eight are obtained from the 33 graduate student applications and four are derived by the faculty. Thus the eight applicant variables may be naturally thought of as independent variables, and the four evaluation variables as dependent variables. The eight independent variables concern the degree obtained by the applicant (bachelor's or master's), the applicant's major as an undergraduate (either psychology, mathematics including statistics, or a double math-psych major), the applicant's grade point average, his verbal and mathematical score on the Graduate Record Examination, his score on the Miller Analogies examination, his strongest interest (quantitative psychology or statistics), and his average recommendation rating. The four dependent variables indicate the tentative accept-reject decision

Table 2

Correlation and Regression Weights for Survey Data

	Canonical		Multiple			
	1	2	1	2	3	4
Regression Weights						
Dependent variables						
Tentative decision	.212					
Final decision	.091					
Suitability order	.970	.996	1.000		1.000	
Applicant action	-.081	-.086		1.000		1.000
Independent variables						
Degree	.137	.204	.214	-.065		
Major	-.136	-.184	-.166	.192		
GPA	.192	.276	.260	-.173		
GRE verbal	-.204	-.295	-.272	1.296		
GRE math	-.281	-.409	-.410	-.957	-.745	.420
Millers Analogies	-.164	-.235	-.232	.187		
Interest	.108	.126	.065	-.348		
Recommendation	-.443	-.636	-.574	.238	-.528	.296
Correlations						
Classical	.925	.923	.921	.593	.842	.297
Optimal	1.000	1.000	1.000	.961	.997	.543
Number of cases removed	7	7	7	7	2	2
Number of iterations	1	2	2	13	15	2

made by the faculty at the beginning of the evaluation (including a "defer" decision), the corresponding decision made at the conclusion of the evaluation, the faculty's rank order of all applicants' suitability to the graduate program, and the action taken by the applicants. Note that three of the four dependent variables are weakly equivalent at the ordinal level. The 13 students who ranked most suitable were accepted, and of these the nine highest were tentatively accepted and the remaining four were tentatively deferred. Note also that eight of those students who were initially deferred were later rejected, and that all of these students were ranked lower than those who were finally accepted.

We performed two canonical regressions on these data, one using all four dependent variables, and one using only *suitability* and *action* (due to the relationship between *suitability* and the other two dependent variables noted above). For each analysis we assumed that all variables except *suitability* were discrete, and that the *degree*, *major* and *interest* variables were nominal, while all others were ordinal. For each analysis we removed all of those students who had missing data (the analyses were also performed with these students left in and with their missing data estimated as stated above, with little difference in results).

The results of each analysis are presented in Table 2. For both analyses

the optimal canonical correlation was 1.0, obtained in either 1 or 2 iterations. Thus it is possible to transform the dependent and independent variables in a manner which allows a linear combination of one set to be perfectly linearly related to a linear combination of the other set. The pattern of weights is very informative. Among the dependent variables *suitability* is very heavily weighted, and the remaining receive almost no weight. This is true for both analyses and suggests that a very strong multiple correlation based only on *suitability* exists (a suggestion which will prove to be correct). The pattern of dependent variable weights is easily interpreted. We should expect the *suitability* weight to be very high because the suitability rank order is developed by the faculty from the applications summarized by the independent variables. So, if the independent variables summarize all of the information in the application, and if the rank order is consistent with that information, then the regression weight should be very high. The weights on the *tentative* and *final* decision variables should be low because their information is already contained in the *suitability* rank order, as noted above. Finally, the *action* variable should have a low weight because if reflects many variables not included in the analysis, such as whether or not the applicant was accepted by other graduate programs, how strongly he desired to be in each program, etc. Turning now to the independent variable weights, we see that the pattern of weights is the same for the two analyses, with the *GRE-math* and *recommendation* variables receiving the heaviest weight. It is reassuring that the pattern of weights is stable between the two analyses.

Due to the perfect relationship derived in both canonical analyses, we decided to perform several multiple regression analyses, using either *suitability* or *action* as the dependent variable. The results of these analyses are also presented in Table 2. (The same measurement assumptions and initial category values were used in these analyses as in the canonical analyses.)

We first note that the optimal multiple correlation is still perfect for the analysis using *suitability* as the dependent variable, but is somewhat lower for the *action* analysis, a result which is consonant with the dependent variable weights in the canonical analyses above. It is also interesting to note that the classical multiple correlation is initially quite strong for *suitability* but rather weak for *action*. We next note that the regression weights for the *suitability* analysis display the same pattern as the weights in the canonical analyses, but that the weight pattern for the *action* analysis has changed. This should be expected due to the very large weight for *suitability* and the low weight for *action* in the canonical analyses.

Finally, we performed one more pair of multiple correlation analyses, each involving only the two independent variables receiving the most weight in the three analyses in which *suitability* was a dependent variable. This new pair of analyses differed according to whether the dependent variable was

suitability or *action*. The results are also presented in Table 2, and are as anticipated. The optimal multiple correlation for the *suitability* analysis is no longer perfect, but is nearly so (.997), and the optimal multiple correlation for the *action* analysis is very low (.5427). Furthermore, the regression weights for the *suitability* analysis are essentially the same as those in the previous analyses with this dependent variable.

The final aspect of these analyses to be emphasized is the set of optimal scale values assigned to each nominal variable (see Table 3). (We do not present the scale values for binary variables since they are arbitrarily set to one and zero.) The first point to be made is that the scale values are rather stable for all analyses except the multiple correlation analysis of *action*. This result is in line with previous results. We next note that these variables do not receive very heavy regression weights in any of the analyses, so any interpretation of the variables by themselves does not add much to our understanding of the total process. However, for the three analyses in which *suitability* was a dependent variable, *major* received a negative weight, indicating that larger (optimally scaled) *major* values are associated with smaller (optimally scaled) *suitability* values. When we note that smaller *suitability* values indicate greater suitability (the highest rank is 1), we see that applicants with mathematics backgrounds tend to be rated as most suitable, and those with psychology backgrounds as least suitable, while those having double majors are in between. In a corresponding fashion we can interpret the optimal scale values for *interest* (which has a positive regression weight in the three analyses in which *suitability* is involved) as showing that applicants who are most interested in quantitative psychology tend to be judged as most suitable, cognitive psychology as next most suitable, and statistics as least suitable. Finally, the positive regression weight for the binary *degree* variable, and the coding of this variable (1 for Bachelors, 2 for Masters) indicates that applicants with Bachelors degrees are more suitable.

Thus, from all of these analyses taken together, we conclude that it is possible to describe the faculty's judgments of the suitability of a prospec-

Table 3

Optimal Scale Values

Variable	Category	Canonical 1	Canonical 2	Multiple 1	Multiple 2	Initial Values
Major	Psychology	1.023	1.016	1.008	1.333	1
	Math/Psych	1.803	1.856	1.921	2.133	2
	Mathematics	3.169	3.129	3.075	-0.554	3
Interest	Quantitative	1.007	.996	.998	.892	1
	Cognitive	1.984	2.010	2.005	2.531	2
	Statistics	3.012	2.992	2.996	2.400	3

tive graduate student on the basis of only that student's score on the quantitative section of the Graduate Record Examination, and on his strength of recommendation. In addition we conclude that it is not possible to adequately describe the characteristics of those applicants who are to become members of the graduate training problem on the basis of the variables summarizing their applications. Furthermore, the faculty tend to prefer students interested in quantitative psychology over those interested in cognitive psychology (and least like those interested in statistics), tend to judge those with a mathematical background as more suitable than those without, and tend to prefer students without masters degrees.

MDS interpretation

One of the special uses of the MORALS/CORALS approach is to interpret multidimensional scaling (MDS) configurations by projecting external information into the MDS space. Cliff and Young [1968] suggested that multiple correlation techniques could be used to project information into the MDS space in the form of a vector in the space which correlated most highly with the external information. Carroll [1972] suggested that multiple correlation techniques could also project information into an MDS space in the form of a point such that the distances between the projected point and the MDS points correlated most highly with the external information. Both of these suggestions assumed that the external variable is measured at the interval level of measurement, an assumption which is usually untenable, but generally ignored. Carroll and Chang [1970], however, have recently extended their proposal to cover variables defined at the ordinal level by a procedure precisely equivalent to the special case of MORALS to be discussed [Note 2].

Hoadley [Note 6] has obtained a MDS solution for 100 U.S. Senators based on their voting records in the Senate. He also obtained a variety of additional information including ratings of each Senator by liberal, conservative and special interest groups, proportion of votes which supported the Senator's party position on various issues, specific votes on certain key issues, etc. He also computed Guttman scales on several different topics of national concern.

MORALS/CORALS was used with these data in the following manner. Several multiple regressions were performed, all of which assumed that the two dimensions of the MDS space were defined at the interval level of measurement, and that the vector of information being regressed into the MDS space was qualitative. The results of these analyses are presented in Table 4, along with the measurement characteristics of each dependent variable. A geometric representation of one of these analyses is presented in Figure 2. This is a representation of both the MDS space and the regression of the Guttman scale concerning the degree to which each Senator voted in support

Table 4

Correlations and Regression Weights for MDS Interpretation

Variable	Measurement Level	Measurement Process	Correlation Initial	Correlation Final	Weights Initial dimension 1	Weights Initial dimension 2	Weights Final dimension 1	Weights Final dimension 2	Number of Iterations
War Powers	binary	d	.4819	.6003	-.476	.086	-.599	.051	1
Agricultural Subs	binary	d	.4426	.5922	-.341	.291	-.456	.389	1
Party	binary	c	.9121	1.0000	.549	.716	.549	.716	(0)
G Agri. Scale	ordinal	d	.7900	.8416	.708	-.367	.779	-.338	2
G Civil Liberty	ordinal	d	.9255	.9333	-.764	+.541	-.765	.552	1
G Gvt. Reg. Business	ordinal	d	.8812	.8918	-.879	-.049	-.889	-.056	1

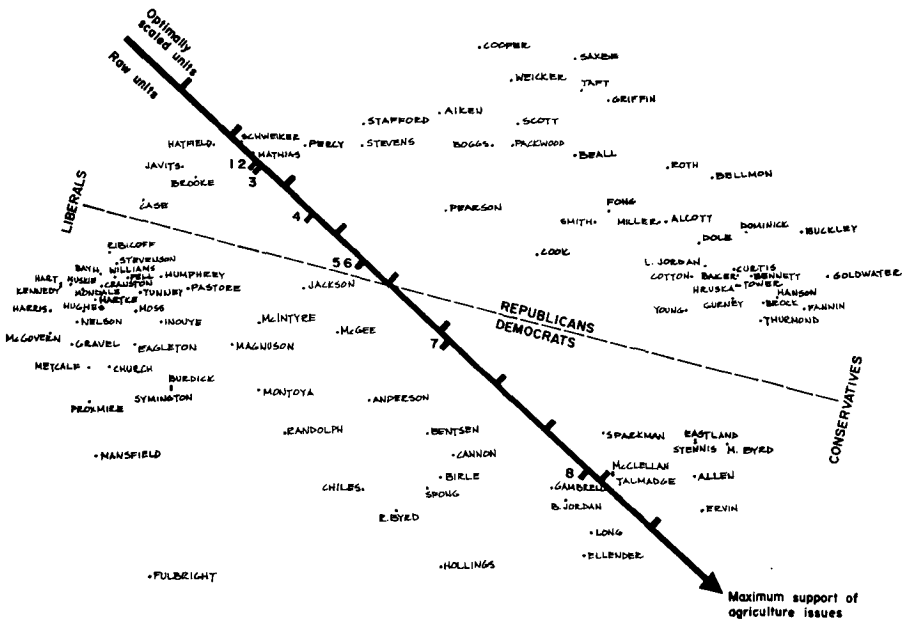

FIGURE 2
Configuration of U. S. Senators with the optimally regressed agriculture support vector.

of agricultural issues. The vector through the space is the direction determined by MORALS/CORALS such that when the points in the space are projected onto the vector, the resulting projections are as strongly correlated with the optimally scaled units as possible. Note that we have presented the optimally scaled units along one side of the vector, and the raw Guttman scale units along the other. The relationship between each of these sets of units is the optimal transformation 3^{do}. The interpretation of this figure is straight forward: the conservatives (who are in the lower right part of the space) support agricultural issues most strongly, and the liberals least strongly. The strength of support is mostly a function of the liberal-conservative dimension (the horizontal direction) and only a weak function of the party split (the vertical direction). This interpretation corresponds to the weighting scheme presented in Table 4 (Dimension 1 is horizontal, Dimension 2 vertical).

4. Discussion

The results just presented lead us to conclude that the MORALS/CORALS approach to multiple/canonical correlation with a mixture of qualitative and quantitative data is capable of obtaining meaningful results in a variety of situations. Thus our approach not only extends correlation

analysis to commonly occurring situations which were not previously covered, but also appears to do so in an efficacious manner.

Reliability and Validity

Due to the great flexibility of the MORALS/CORALS procedure, one may question whether the optimal transformations and linear combinations obtained for a particular set of observations will be replicated by a new set of observations in an otherwise identical empirical situation. That is, does the procedure provide reliable estimates of the optimal transformations and of the linear combinations, or will these estimates fluctuate wildly from sample to sample?

Certainly, in some situations the procedure will yield unreliable estimates, and in others it will yield reliable estimates. The issue centers around the ratio of the number of parameters being estimated to the number of observations, with this ratio being most favorable in the classical regression situation (i.e., when all variables are at the interval level of measurement), and least favorable when all variables are nominal, especially when there are relatively few observations in each category. As has been noted in the companion paper [de Leeuw, Young & Takane, 1976, section 4], a necessary condition for a unique solution is that there be at least two observations in at least one category. For the present situation this concept applies to every nominal variable. If at least one nominal variable consists entirely of unique categories (i.e., there is only one observation in each category), then the solution is not uniquely determined; a very large number of arbitrary quantifications of the categories yield correlation coefficients of 1.0. Since the solution is not unique, it is unreliable. Thus, one should avoid nominal variables which have observation categories represented by only one observation. There are other types of situations which also do not yield unique solutions, some of which are discussed in the companion paper.

Generally, one should suspect the reliability of his results in the same situations in which he suspects the reliability of the results of a classical multiple or canonical regression, except that the problems are compounded when the measurement levels of the variables are weakened. Thus, if one has a small number of observations, and a relatively large number of variables of which several are qualitative, then he should suspect the reliability of the results. The survey example given above (Section 3) is an example in which the reliability of the results are suspect. In this example there are a total of four dependent variables and eight independent variables. All of these variables are qualitative, nine being ordinal and three nominal. Note that there are only 33 cases, seven of which are removed from most of the analyses. Thus, the first canonical analysis involved 12 variables and 26 cases; the second, ten variables and 26 cases. The first two multiple correlation analyses each involved nine variables and 26 cases, and the last two, three variables

and 31 cases. It is not surprising that essentially perfect multiple and canonical correlation coefficients are obtained in four of the six analyses. Due to the variable/case ratio one might even suspect the results if all of the variables were quantitative. We do note, however, that the transformation and the weight patterns were interpretable, although the interpretation process was was highly subjective.

The MDS example, involving 99 "cases" (Senators) and only three variables, would seem to be reliable. Only one of the variables was qualitative; it was also ordinal. Thus it would appear to be the case that this analysis should be nearly as reliable as its classical counterpart. The ease of interpretation in this example, and the congruence of the interpretation with the investigators' expectations also attest to the reliability of this example. (Naturally, if reliability is of paramount importance, then any of the standard procedures for empirically determining reliability can be used, including such procedures as split-half analyses and repeated observation designs.)

Three schools of thought about additivity

There have been three essentially separate traditions centered around the additive model. The oldest and most widely known tradition falls under the rubric analysis of variance, with the central focus being on the development of inferential tests of the significance of the model's components. Nearly as old, but much less widely known, is the optimal scaling tradition, with the central focus being on the quantification of qualitative variables within the context of the additive model. The most recent tradition is conjoint measurement, which focuses on the axiom systems underlying the additive model. Our work is clearly within the optimal scaling tradition, and makes no pretenses about the development of inferential tests or axiomatic systems. We are simply interested in scaling qualitative variables so that they will be as linearly related as possible.

In the analysis of variance tradition, the additive model is associated with assumptions about the distributional properties of the observations and/or the errors, additional assumptions about null and alternative hypotheses, and formal procedures for accepting or rejecting the null hypotheses. In terms of the assumptions which must be satisfied for the significance tests to be valid, the analysis of variance tradition requires that the error components be independently normally distributed, generally with zero means and equal variances. The formal theory of parametric statistics (and thus of the analysis of variance) does not concern either the scale level of the variables or the axiomatic system underlying the model. The only critical aspect is the distributional assumption. Moreover, when the null hypotheses are invariant over monotonic transformations of the dependent variable (as is usually the case), the tests can be easily generalized to any other dependent variable which is monotonically equivalent, if the distributional

assumptions of the tests are met. Thus there is a tradition within the analysis of variance literature of monotonically transforming variables so that the distributional assumptions will be met.

In the conjoint measurement tradition, the additive model is conceptualized as representing the "true" nature of the process which generated the dependent measures. Measurement theory is concerned with a) specifying the exact nature of the conditions under which the dependent measures will be precisely representable by the model, and *b*) identifying the types of transformation of the dependent variables which are allowed within the previously identified conditions. In the case of the additive model, conjoint measurement has postulated a set of axioms concerning the structure of the data which are the necessary and sufficient conditions for the additive model to precisely describe the data. Not all of the axioms are empirically testable, but if those that can be tested are satisfied, then there is a monotonic transformation of the dependent variable which allows the variable to be precisely described by the additive model. Thus the central focus of the conjoint measurement tradition is specifying conditions under which monotonic transformations of the dependent variable permit an additive description of the variable.

In the optimal scaling tradition, the basic goal is to obtain a transformation of the dependent (and independent) variables so that the additive model fits as well as possible, as should be obvious from the body of this paper. This focus is similar to one of the conjoint measurement goals, but the difference is that those working in the conjoint measurement tradition are interested in specifying the condition under which the simple additive model will describe the data perfectly, whereas the optimal scaling tradition centers on actually providing the best description.

Note that the conjoint measurement and optimal scaling traditions are similar in that they are both descriptive. Each is concerned with describing the structure of the data with the additive model, one centering on whether a perfect description is possible, and the other one providing the best possible description. The analysis of variance tradition, on the other hand, is inferential: it is concerned with making inferences beyond the data on the basis of the additive model. All three traditions are concerned with data transformations, but the essential difference between the analysis of variance tradition and the other two is that in the former a transformation is desired to improve the adequacy of the inferential process, whereas in the latter a transformation is desired to improve the adequacy of the descriptive process.

Perhaps it should be emphasized that a useful function is performed within each of the traditional approaches to additivity, and that perhaps the best approach is a combination of all three. It seems to the present authors that each approach by itself gives an incomplete picture of the data being analyzed, and that the most complete picture is derived when all three ap-

proaches are used. Surely we best understand a particular set of data when we know the degree to which it satisfies the axioms underlying the additive model, the transformations which make the data most additive, and the significance of the components of the additive model.

REFERENCE NOTES

1. Carroll, J. Douglas. Personal communication, 1976.
2. de Leeuw, J. *Canonical analysis of categorical data.* Unpublished doctoral dissertation, University of Leiden, The Netherlands, 1973.
3. de Leeuw, J. *The linear nonmetric model* (Research Note RN003-69). Leiden, The Netherlands: University of Leiden, Data Theory, 1969.
4. de Leeuw, J. *Nonmetric discriminant analysis* (Research Note RN006-68). Leiden, The Netherlands: University of Leiden, Data Theory, 1968.
5. de Leeuw, J. *Normalized cone regression.* Unpublished paper, University of Leiden, Data Theory, The Netherlands, 1975.
6. Hoadley, J. *Spatial analysis of Senate voting patterns.* Unpublished master's thesis, University of North Carolina, 1974.
7. Srinivasan, V. *Linear programming computational procedures for ordinal regression.* Unpublished manuscript, Graduate School of Business, Stanford University, 1974.

REFERENCES

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences.* Seminar Press, New York, 1972.

Carroll, J. D. & Chang, J. J. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition. *Psychometrika*, 1970, *35*, 283–319.

Cliff, N. & Young, F. W. On the relation between unidimensional judgments and multidimensional scaling. *Organizational Behavior and Human Performance*, 1968, *3*, 207–208.

de Leeuw, J., Young, F. W., & Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 1976, 471–503.

Fisher, R. *Statistical methods for research workers* (10th ed.). Edinburgh: Oliver and Boyd, 1938.

Hayashi, C. On the quantification of qualitative data from the mathematico: Statistical point of view. *Annals of the Institute of Statistical Mathematics*, 1950, *2*, 35–47.

Hotelling, H. The most predictable criterion. *Journal of Educational Psychology*, 1935, *26*, 139–142.

Lingoes, J. C. *The Guttman–Lingoes nonmetric program series.* Ann Arbor, Michigan: Mathesis Press, 1973.

Shepard, R. N. & Kruskal, J. B. A nonmetric variety of linear factor analysis. *Psychometrika*, 1974, *39*, 123–157.

Takane, Y., Young, F. W., & de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, in press.

Wold, H. & Lyttkens, E. (Eds.) Nonlinear iterative partial least squares (NIPALS) estimation procedures (group report). *Bulletin of the International Statistical Institute*, 1969, *43*, 29–51.

Young, F. W. A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling, Vol. 1*. New York: Seminar Press, 1972.

Young, F. W. Methods for describing ordinal data with cardinal models. *Journal of Mathematical Psychology*, 1975, *12*, 416–436.

Manuscript received 12/1/75
Final version received 4/20/76