PSYCHOMETRIKA—VOL. 51, NO. 1, 101–112
MARCH 1986

ASSESSING THE CORRESPONDENCE OF ONE OR MORE VECTORS TO A SYMMETRIC MATRIX USING ORDINAL REGRESSION

THOMAS J. REYNOLDS
THE UNIVERSITY OF TEXAS AT DALLAS
KENNETH H. SUTRICK
MURRAY STATE UNIVERSITY

A statistical model for interpreting psychological scaling research, based on the heuristic work of Reynolds (1983), is developed. This new approach has certain advantages over the standard property fitting approach (Chang and Carroll, 1969) currently used to interpret multidimensional scaling spaces (Shepard, 1962; Torgerson, 1965). These advantages are (a) the ability to directly assess the correspondence of a descriptor vector(s) to a symmetric matrix, and (b) to provide a method in which only ordinal properties of such descriptors are required: thus standard rating, ranking, or sorting data collection methods can be used as the basis to interpret the multidimensional space resulting from the distance data.

Introduction

Multidimensional scaling (MDS) analysis (Shepard, 1962; Torgerson, 1965) is a general term descriptive of a series of algorithms used typically in behavioral research whose objective is to maximally reproduce pairwise proximities or distances between n stimuli in a Euclidean space of a given dimensionality. These $C(n, 2)$ pairwise measures of psychological distance, where $C(n, 2)$ is the number of combinations of n objects taken two at a time, will be denoted $\{Y_{ij}\}$. Typically the $\{Y_{ij}\}$ are judged on a nine point dissimilarity scale or the subject rank orders all pairs providing a scale of length $C(n, 2)$. Importantly, MDS procedures assume only that the scale of the proximity measures is monotonic. The goal of MDS is, of course, to represent the pairwise judgements in as few dimensions as possible. The determination of relative goodness-of-fit of various dimensional solutions is made on the basis of a stress measure (Kruskal, 1964), using rule of thumb interpretation.

The data to be used in the interpretation of the resulting multidimensional representation typically consists of vectors $\mathbf{Z}_i$, which are characteristics or descriptor attributes of the i-th stimulus. The goal of the interpretative phase is to determine to what extent Y_{ij} corresponds to $\mathbf{Z}_i$ and $\mathbf{Z}_j$. The standard methodological approach is to treat each attribute (i.e., each coordinate of $\mathbf{Z}$) as a dependent variable regressing on the Cartesian coordinates of the Euclidean space output from the MDS analysis. This property fitting approach yields direction cosines (computed from the β's) that permit a vector representing the descriptor to be plotted in the Euclidean stimulus space (Chang & Carroll, 1969). This set of vectors (one for each descriptor) as well as their relative fit to the stimulus space is then used to make interpretive inferences as to the underlying basis of the perceptual discrimination between the stimulus objects.

Reynolds (1983) points out a number of limitations of the above analytic procedure, namely:

1. The fact that the regression is performed on the recovered stimulus coordinates rather than directly on the actual psychological distances injects an indeterminant degree

Requests for reprints should be sent to Thomas J. Reynolds, School of Management, University of Texas at Dallas, Richardson, TX 75080.

0033-3123/86/0300-6101$00.75/0

of error. Reynolds however does not address the fact that many researchers believe that the use of MDS actually results in a more stable representation of the bulky and more unstable set of input judgments.

2. The use of linear regression assumes an interval sale for the descriptor ratings which is always in question when human judgement tasks are utilized.

Since the assumption of an ordinal scale is more appropriate we take an ordinal regression approach in analyzing this data. The goal is to determine how well ordinal relationships between pairs Y_{ij} and pairs $Y_{i'j'}$ are predicted by ordinal relationships between pairs $\mathbf{Z}_i$, $\mathbf{Z}_j$ and pairs $\mathbf{Z}_{i'}$, $\mathbf{Z}_{j'}$. Essentially the problem is one of comparing a symmetric matrix of pairwise judgements between stimuli to a vector of stimuli ratings. A measure of fit is based on comparing all $C(C(n, 2), 2)$ pairs of pairs of subscripts "ij". The measures of fit bear a formal resemblance to the Goodman-Kruskal gamma (1954), Kendall's tau (1955) and Somers' D (1962), but do not possess the same statistical properties.

The unique and truly interesting aspect of this problem is that the $\{Y_{ij}\}$ which are the dependent variables have $C(n, 2)$ degrees of freedom while the $\{\mathbf{Z}_i\}$ which are the independent variables have n degrees of freedom. This difference in degrees of freedom gives our measures of fit quite different statistical properties than the traditional measures.

The statistical properties (for example limiting normality) are obtained by adapting U-statistics theorems (Hoeffding, 1948) to this somewhat more complicated situation. Both the single and multiple vector cases, assuming equal weighting, are discussed. An example is detailed.

Association With One Independent Variable

In this section we consider the case where there is only one independent descriptor variable and the pairwise proximity matrix. The statistical model for this data is described: the prediction rule for determining ordinal relationships among the $\{Y_{ij}\}$ is given, the measures of fit are specified, and the statistical properties of the measures of fit are discussed.

The statistical model is presented in the following two assumptions.

Assumption 1

The attribute ratings are a set of n independent identically distributed random variables $\{Z_i\}$.

Assumption 2

The measures of psychological distance $\{Y_{ij}\}$ are a set of $C(n, 2)$ random variables of the form

$$Y_{ij} = f(Z_i, Z_j) + \varepsilon_{ij}, \qquad i, j = 1, 2, \ldots, n, \, i < j,$$

where the $\{\varepsilon_{ij}\}$ are a set of independent identically distributed measurement errors, mutually independent of $\{Z_i\}$, and the function f satisfies:

$$f(Z_i, Z_j) = f^*(\min \{Z_i, Z_j\}, \max \{Z_i, Z_j\})$$

where f^* is a bivariate function such that $f^*(w_1, w_2) \geq f^*(w_3, w_4)$ if $w_1 \leq w_3 \leq w_4 \leq w_2$. This definition of f results in a positive correlation between $\{Y_{ij}\}$ and $\{Z_i\}$. A negative correlation results if f^* satisfies $f^*(w_1, w_2) \leq f^*(w_3, w_4)$.

Since the $\{Z_i\}$ are a set of n independent variables, they have n degrees of freedom. We shall say that the $\{Y_{ij}\}$ have $C(n, 2)$ degrees of freedom since in addition to the variability of the $\{Z_i\}$ they contain $C(n, 2)$ independent bits of error from the $\{\varepsilon_{ij}\}$.

Such functions f^* do exist. For example, if the data is interval rather than ordinal then the conditions of f^* are satisfied if $f^*(w_1, w_2) = f^{**}(|w_2 - w_1|)$, for some monotonically increasing function f^{**}. However, for such an f^{**} the $\{Y_{ij}\}$ depend on the univariate metric distance $|w_2 - w_1|$. In data collection procedures where the subject rank orders or sorts all pairs only the ordinal nature of the $\{Y_{ij}\}$ can be assumed.

We will not make the stronger assumption of the existence of an f^{**}. Also the size of Y_{ij} should depend on the relative distance between Z_i and Z_j. This relative distance is determined by min $\{Z_i, Z_j\}$ and max $\{Z_i, Z_j\}$, which when used in a bivariate function f^* does not include any metric assumptions such as those of an f^{**}.

No other smoothness conditions on the random variables will be assumed and we shall not assume a particular parametric form for the function f. Our procedures are general and will work as long as Assumption 2 is satisfied. To simplify the notation it will be useful to define random vectors $\{\mathbf{X}_{ij}\}$, $\{\mathbf{Z}_{ij}\}$ and random variables Z_{ij}^w, Z_{ij}^u as

$$\mathbf{X}_{ij} = \begin{bmatrix} Y_{ij} \\ Z_i \\ Z_j \end{bmatrix}, \tag{1}$$

$$\mathbf{Z}_{ij} = \begin{bmatrix} Z_i \\ Z_j \end{bmatrix}, \tag{2}$$

$$\begin{aligned} Z_{ij}^w &= \min\{Z_i, Z_j\}, \qquad \text{and} \\ Z_{ij}^u &= \max\{Z_i, Z_j\}. \end{aligned} \tag{3}$$

In this regression situation we want to determine how ordinal relations among the $\{Y_{ij}\}$ are predicted by ordinal relations among the $\{Z_i\}$. Due to the ordinal nature of the data, a prediction of a relation will be made if $Z_{ij}^w \le Z_{i'j'}^w \le Z_{i'j'}^u \le Z_{ij}^u$ and either $Z_{ij}^w \ne Z_{i'j'}^w$ or $Z_{ij}^u \ne Z_{i'j'}^u$. Define a function b which indicates when a relation can be predicted as

$$b(\mathbf{Z}_{ij}, \mathbf{Z}_{i'j'}) = \begin{cases} 1 & \text{if } Z_{ij}^w \le Z_{i'j'}^w \le Z_{i'j'}^u \le Z_{ij}^u \\ & \text{and either } \quad Z_{ij}^w \ne Z_{i'j'}^w \quad \text{or} \quad Z_{ij}^u \ne Z_{i'j'}^u \\ -1 & \text{if } Z_{i'j'}^w \le Z_{ij}^w \le Z_{ij}^u \le Z_{i'j'}^u \\ & \text{and either } \quad Z_{ij}^w \ne Z_{i'j'}^w \quad \text{or} \quad Z_{ij}^u \ne Z_{i'j'}^u \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The following ordinal prediction rule is straightforward in this context.

Prediction Rule 1. Predict $Y_{ij} > Y_{i'j'}$ if $b(\mathbf{Z}_{ij}, \mathbf{Z}_{i'j'}) > 0$, predict $Y_{ij} < Y_{i'j'}$ if $b(\mathbf{Z}_{ij}, \mathbf{Z}_{i'j'}) < 0$.

A measure of fit can be defined by comparing all $C(C(n, 2), 2)$ pairs of pairs ij and counting the number of correct ordinal predictions. Formally this is accomplished by defining functions B^k, h_{1k}, h_{2k}, h_{3k}. The subscript k denotes the number of independent variables. The same measures of fit are applicable in the univariate and multivariate cases. In the univariate case, $k = 1$, the function B^k satisfies:

$$B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = B^1(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = b(\mathbf{Z}_{ij}, \mathbf{Z}_{i'j'}). \tag{5}$$

The functions h_{1k}, h_{2k}, h_{3k} are defined as

$$h_{1k}(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = \begin{cases} 1 & \text{if } (Y_{ij} - Y_{i'j'}) \cdot B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$h_{2k}(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = \begin{cases} 1 & \text{if } (Y_{ij} - Y_{i'j'}) \cdot B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) < 0 \\ 0 & \text{otherwise} \end{cases}$$

$$h_{3k}(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = \begin{cases} 1 & \text{if } (Y_{ij} - Y_{i'j'}) \cdot B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Pairs ij are concordant with pairs $i'j'$ if $h_1 = 1$, discordant if $h_2 = 1$, and tied if $h_3 = 1$. Ties are those pairs where the order of one or both variables cannot be determined. A count of the number of concordant pairs, discordant pairs, and ties is given by

$$U_{\ell kn} = \frac{1}{C_n} \sum_{D_n} h_{\ell k}(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}), \tag{6}$$

respectively, for $\ell = 1, 2, 3$, where $C_n = C(C(n, 2), 2)$ and D_n is the set of all allowable subscripts

$$D_n = \{(ij, i'j') \mid i < j, i' < j', i < i'\} \cup \{(ij, i'j') \mid i < j, i' < j', i = i', j < j'\}.$$

To help explicate the rather involved notation, consider an upper triangular matrix of $\{Y_{ij}\}$ where, for example, $n = 5$. In this case, a Y_{ij} is compared to $Y_{i'j'}$, which is either lower and/or to the right.

A number of measures of fit are possible (see Simon, 1978). We shall look at two used in Reynolds (1983):

$$T_{tk} = \frac{U_{1kn} - U_{2kn}}{U_{1kn} + U_{2kn} + U_{3kn}}$$

and (7)

$$T_{rk} = \frac{U_{1kn} - U_{2kn}}{U_{1kn} + U_{2kn}}$$

where $k = 1$ in the univariate case. The measure T_{rk} bears a formal resemblance to the Goodman-Kruskal gamma (1954) and T_{tk} is somewhat like Kendall's tau (1955). However, the difference in degrees of freedom between the independent and dependent variables gives the measures in (7) quite different statistical properties.

Additionally, a measure to be considered that reflects in part a compromise between Reynolds' indices T_{tk} and T_{rk}, resembling Somers' D (1962), is

$$T_{jk} = \frac{U_{1kn} - U_{2kn}}{U_{1kn} + U_{2kn} + \frac{1}{2}U_{3kn}}. \tag{8}$$

The properties of T_{tk}, T_{jk}, T_{rk} are determined by the properties of $U_{\ell kn}$ in (6). The properties of $U_{\ell kn}$, such as limiting normality, can be derived along the lines of, but not identical to, U-statistics theorems. For a review of U-statistics, see for example Randles and Wolfe (1979). There is additional work required in this case, because due to the differing degrees of freedom the usual theorems do not apply. The fact that the T-statistics above can be approximated by the normal distribution is delimited in the following two theorems. A sketch of their proofs is outlined in the Appendix.

Theorem 1. Under Assumptions 1 and 2, if $U_{\ell kn}$ is given by (6) then

$$n^{1/2}(U_{\ell kn} - EU_{\ell kn}) \xrightarrow{\mathcal{L}} N(0, 16v_{\ell k})$$

where $v_{\ell k} = \text{Cov}\,[h_{\ell k}(\mathbf{X}_{12}, \mathbf{X}_{34}), h_{\ell k}(\mathbf{X}_{15}, \mathbf{X}_{67})]$, $\ell = 1, 2, 3$ and $\mathbf{X}_{ij}$ is given by (1).

Since the functions $h_{\ell k}$, $\ell = 1, 2, 3$ take on the values zero or one, and hence have finite mean and variance, no additional smoothness conditions are required on the $\{\mathbf{X}_{ij}\}$ for limiting normality.

Theorem 2. Under Assumptions 1 and 2 if T_{rk} is given by (7) then

$$\frac{n^{1/2}(T_{rk} - ET_{rk})}{(\mathrm{Var}\ T_{rk})^{1/2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

A similar result holds for statistics T_{tk} and T_{jk}.

Theorem 2 can be used to construct goodness-of-fit test statistics, given an estimate of variance. An estimate of variance can be obtained by the jackknife procedure, which recomputes statistic T successively leaving out one stimulus at a time. Define

$$D_n^{(-s)} = \{(ij, i'j') \mid i < j,\ i' < j'\ i < i' \quad \text{and} \quad i, j, i', j' \neq s\}$$
$$\cup \{(ij, i'j') \mid i < j,\ i' < j',\ i = i', j < j' \quad \text{and} \quad i, j, i', j' \neq s\}$$

and define

$$U_{\ell kn}^{(-s)} = \frac{1}{C_{n-1}} \sum_{D_n^{(-s)}} h_{\ell k}(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}).$$

The jackknifed version of statistic T_{rk} has the form

$$T_{rk}^{(-s)} = \frac{U_{1kn}^{(-s)} - U_{2kn}^{(-s)}}{U_{1kn}^{(-s)} + U_{2kn}^{(-s)}}, \tag{9}$$

and the jackknifed estimate of variance is

$$\hat{\lambda}_{T_{rk}}^2 = \sum_{s=1}^{n} (T_{rk}^{(-s)} - \bar{T}_{rk})^2 \tag{10}$$

where

$$\bar{T}_{rk} = \frac{1}{n} \sum_{s=1}^{n} T_{rk}^{(-s)}. \tag{11}$$

The following theorem presents a test statistic resulting from the above statements.

Theorem 3. Assume that Assumptions 1 and 2 hold. Under the null hypothesis $H_0 : f \equiv 0$, that is, there is no order relation between Y_{ij} and Z_i, Z_j, the test statistic

$$\hat{\theta}_{rk} = \frac{T_{rk}}{\hat{\lambda}_{T_{rk}}} \tag{12}$$

has a limiting standard normal distribution as $n \to \infty$, where T_{rk} is given by (7) and $\hat{\lambda}_{T_{rk}}$ is given in (10).

For the sketch of a proof see the Appendix.

A similar result holds for statistics T_{jk} and T_{tk}.

Regression With Multiple, Equally Weighted Independent Variables

In this section we treat the case where the observations on the independent variable $\{Y_{ij}\}$ are a function of more than one attribute descriptor rating. The observations on the independent variables are a set of vectors $\{\mathbf{Z}_i\}$,

$$\mathbf{Z}_i = \begin{bmatrix} Z_i^{(1)} \\ Z_i^{(2)} \\ \vdots \\ Z_i^{(k)} \end{bmatrix}. \tag{13}$$

There are two hypotheses of interest. The first is testing whether the whole set of k independent variables has predictive power. The second is whether adding a $(k+1)$-st independent variable significantly increases the predictive power.

Assumption 3

The random vectors $\{\mathbf{Z}_i\}$ are a set of independent identically distributed random vectors.

Assumption 4

The random variables $\{Y_{ij}\}$ are of the form

$$Y_{ij} = f_k(Z_i^{(1)}, Z_j^{(1)}, Z_i^{(2)}, Z_j^{(2)}, \ldots, Z_i^{(k)}, Z_j^{(k)}) + \varepsilon_{ij}$$
$$i, j = 1, \ldots, n, \quad i < j,$$

where the $\{\varepsilon_{ij}\}$ are a set of independent identically distributed measurement errors, mutually independent of the $\{\mathbf{Z}_i\}$, where f_k is a $2k$-variate function such that for each pair of arguments $Z_i^{(\ell)}, Z_j^{(\ell)}, \ell = 1, \ldots, k$, it satisfies the assumptions of f in Assumption 2. That is, for all $\ell = 1, \ldots, k$ if

$$\mathbf{Z}_{ij}^{(\ell)} = \begin{bmatrix} Z_i^{(\ell)} \\ Z_j^{(\ell)} \end{bmatrix}, \tag{14}$$

and $b(\mathbf{Z}_{ij}^{(\ell)}, \mathbf{Z}_{i'j'}^{(\ell)}) = 1$ with b given by (4) then

$$f_k(\cdot, \cdot, \cdots, Z_i^{(\ell)}, Z_j^{(\ell)}, \cdots, \cdot, \cdot) \geq f_k(\cdot, \cdot, \cdots, Z_{i'}^{(\ell)}, Z_{j'}^{(\ell)}, \cdots, \cdot, \cdot).$$

Again we are interested in predicting the ordinal relations of the $\{Y_{ij}\}$ as a function of the vectors $\{\mathbf{Z}_i\}$. An ordinal prediction rule, where each independent variable is equally weighted, is obtained by defining the function B^k as

$$B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) = \sum_{\ell=1}^{k} b(\mathbf{Z}_{ij}^{(\ell)}, \mathbf{Z}_{i'j'}^{(\ell)}) \tag{15}$$

where b is given in (4) and $\mathbf{Z}_{ij}^{(\ell)}$ by (14).

Prediction Rule 2. Predict $Y_{ij} > Y_{i'j'}$ if $B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) > 0$ and predict $Y_{ij} < Y_{i'j'}$ if $B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) < 0$.

The measures of fit for Prediction Rule 2 are given by (7) and (8). Theorems 1, 2, and 3 also hold in the multivariate case with similar proofs. The test statistic (12) is a test of the predictive power of the combined set of independent variables.

The final test to be discussed is the significance of adding an independent variable. If there are k independent variables and a $(k+1)$-st variable is added, a test statistic can be based on

$$\hat{\theta}_{r,k,k+1} = \frac{T_{r,k+1} - T_{rk}}{\hat{\lambda}_{r,k,k+1}} \tag{16}$$

where $T_{r,k+1}$, T_{rk} are given by (7) and

$$\hat{\lambda}^2_{r,k,k+1} = \Sigma\, [T_{rk}^{(-s)} - T_{r,k+1}^{(-s)} - \bar{T}_{rk} + \bar{T}_{r,k+1}]^2$$

where $T_{rk}^{(-s)}, T_{r,k+1}^{(-s)}, \bar{T}_{rk}, \bar{T}_{r,k+1}$ are given by (9) and (11).

Theorem 4. Under Assumptions 3 and 4 the test statistic (16) has an approximate $N(0, 1)$ distribution.

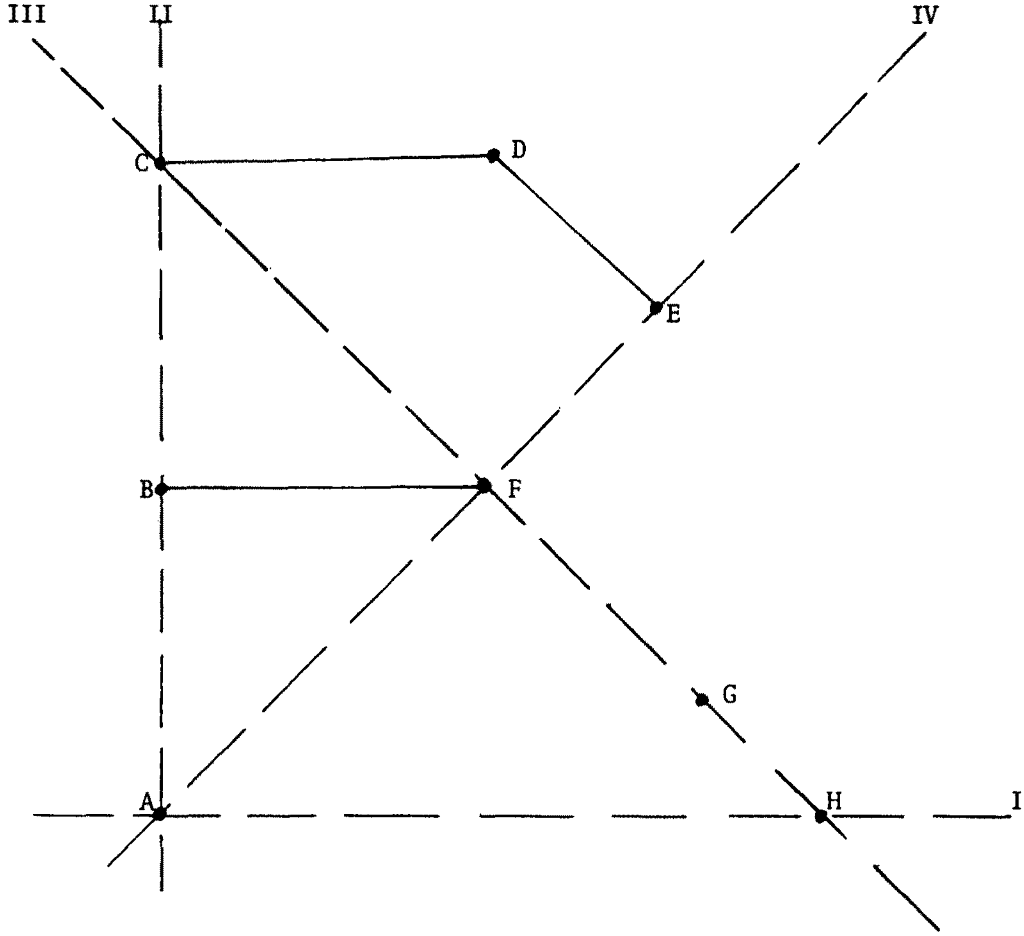

FIGURE 1
"R" configuration with eight imbedded stimulus points

For a sketch of the proof see the Appendix.

The test statistic (16) gives an approximate test. A more exact test under Assumptions 3 and 4 is delimited in Reynolds and Sutrick (1984). Problems occur when ties are broken or made going from k to $k + 1$ variables. For example the function B^k in (15) might satisfy $B^k = 0$ but $B^{k+1} > 0$. These cases must be kept track of carefully in an exact test. Similar results hold for statistics T_{tk}, T_{jk} in (7) and (8).

Example

An artificial two dimensional representation of eight points embedded in an "R" configuration, seen in Figure 1 (taken from Reynolds, 1983) will be used to illustrate the computations. To be reported are the computation of the univariate measures ($k = 1$) of association (T_{t1}, T_{j1}, T_{r1}) given by (7) and (8) and their tests of significance, the bivariate equal weighting case of combining descriptor vectors given by Prediction Rule 2 and (7) and (8), and the test for significance for adding a predictor given in (16).

The four vectors denoted in Figure 1, representing two orthogonal parts, reflect the basis of the descriptor scores $\{\mathbf{Z}_i\}$ reported in Table 1. For example, consider Descriptor III starting from the upper left moving to the lower right: stimulus C is a 1, B and D are

TABLE 1

Descriptor Scores Corresponding to Axes of "R" Configuration

Stimuli	I	II	III	IV
A	4	5	3	4
B	4	3	2	3
C	4	1	1	2
D	3	1	2	1
E	2	2	3	1
F	3	3	3	2
G	2	4	4	2
H	1	5	5	2

2's, A, E, and F are 3's, G is 4 and H is 5. The distance measures $\{Y_{ij}\}$, seen in Table 2, represent a rank ordering of the Euclidean distances between all pairs of stimulus points.

The tau indices T_{t1}, T_{j1}, and T_{r1}, from (7) and (8) for each of the four descriptor vectors are reported in Table 3 along with their respective test statistics, $\hat{\theta}$, from (12). As detailed, the calculation of $\hat{\theta}$ requires the bootstrap estimate of the variance, $\hat{\lambda}_T$, for whichever index is being evaluated, as described in (10). To illustrate using the T_{r1} measure, each of the stimuli would have to be pulled out and a new T_{r1} statistic for the reduced set, denoted $T_{r1}^{(-s)}$, computed. For descriptor vector I, where $k = 1$, the computations result in:

$$T_{r1}^{(-1)} = .8333 \qquad T_{r1}^{(-5)} = .7345$$

$$T_{r1}^{(-2)} = .6071 \qquad T_{r1}^{(-6)} = .6923$$

$$T_{r1}^{(-3)} = .7064 \qquad T_{r1}^{(-7)} = .7217$$

$$T_{r1}^{(-4)} = .7886 \qquad T_{r1}^{(-8)} = .6250.$$

TABLE 2

Rank Order Transformation of "R" Distances

Stimuli	A	B	C	D	E	F	G	H
A	-	2	5	7	6	3	4	5
B		-	2	3	4	2	4	7
C			-	2	4	3	6	8
D				-	1	2	4	7
E					-	1	2	4
F						-	1	3
G							-	1
H								-

TABLE 3

Univariate T Measures and Normal Scores

	I	II	III	IV
T_t	.441	.405	.353	.220
$\hat{\theta}_t$	3.183	2.842	1.625	.827
T_j	.548	.510	.450	.275
$\hat{\theta}_j$	3.382	3.019	1.627	.820
T_r	.723	.691	.619	.367
$\hat{\theta}_r$	3.606	3.327	1.626	.808

The estimate of the standard deviation of T_{r1}, $\hat{\lambda}_{T_{r1}}$, following (10) is

$$\hat{\lambda}_{T_{r1}} = [\Sigma\,(T_{r1}^{(-s)} - \bar{T}_{r1})^2]^{1/2} = .2004$$

where $\bar{T}_{r1} = (1/n)\,\Sigma\,T_{r1}^{(-s)}$. And, the test statistic is following (12),

$$\hat{\theta}_{r1} = \frac{T_{r1}}{\hat{\lambda}_{T_{r1}}} = \frac{.7228}{.2004} = 3.61.$$

The interpretation of $\hat{\theta}_{r1}$ under H_0 can be roughly assessed from its limiting normal distribution, although we have only eight stimuli.

The equal weighting combination, for $k = 2$, of descriptors given by Prediction Rule 2 is presented in Table 4. The computer implementation of this rule was accomplished by defining two vectors of length $C(C(n, 2), 2)$. The first vector represents the dependent variable and has as its components sgn $[Y_{ij} - Y_{i'j'}]$, where *sgn* is the sign function

$$sgn\,[x] = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ -1 & \text{if} \quad x < 0. \end{cases}$$

The second vector representing the independent variables has components *sgn* $[B^k(\mathbf{X}_{ij}, \mathbf{X}_{i'j'})]$. The measures of fit T_{rk}, T_{jk}, T_{tk} are measures of correlation between the two vectors. It is useful to think of the second vector as a composite vector combining the independent variables. A component of the composite vector, for $k = 2$, is a function of $b(\mathbf{Z}_{ij}^{(1)}, \mathbf{Z}_{i'j'}^{(1)})$ and $b(\mathbf{Z}_{ij}^{(2)}, \mathbf{Z}_{i'j'}^{(2)})$. The component is zero if the functions b are both zero or opposite in sign, and is the sign of the nonzero b otherwise. The orthogonal axes combine to produce a better fit than do the other combinations. Also reported in Table 4 are the recomputed $\hat{\theta}$ values, from (12), for the newly constructed composite vector.

The jackknifed values of the combination of Descriptors I and II for statistic T_{r2},

TABLE 4

Composite Vector T's and Normal Scores*

		I	II	III	IV
	I	---	7.010	3.095	4.701
	II	.540	---	3.440	3.580
T_t	III	.468	.440	---	5.323
	IV	.467	.423	.456	---
	I	---	7.659	3.224	5.263
	II	.661	---	3.665	3.996
T_j	III	.565	.543	---	6.287
	IV	.567	.520	.575	---
	I	---	8.796	3.370	6.042
	II	.827	---	4.036	4.818
T_r	III	.711	.711	---	7.995
	IV	.724	.676	.778	---

*T values in lower diagonal and normal scores in upper diagonal.

using (9), lead to

$$\begin{array}{ll} T_{r2}^{(-1)} = .8589 & T_{r2}^{(-5)} = .8340 \\ T_{r2}^{(-2)} = .7642 & T_{r2}^{(-6)} = .8219 \\ T_{r2}^{(-3)} = .8000 & T_{r2}^{(-7)} = .8548 \\ T_{r2}^{(-4)} = .8492 & T_{r2}^{(-8)} = .7806 \end{array} \tag{17}$$

$$\hat{\lambda}_{T_{r2}} = [\Sigma\,(T_{r2}^{(-s)} - \bar{T}_{r2})^2]^{1/2} = .0941$$

and

$$\hat{\theta}_{r2} = \frac{T_{r2}}{\hat{\lambda}_{T_{r2}}} = 8.80.$$

An examination of the composite (I and II) jackknifed values in (17) shows rather stable behavior of T_{r2} leading to a relatively small standard deviation. This in conjunction with a very high value of T_{r2}(.827), due to the error-free example, yields a highly signficiant normal score of 8.80. Thus, the composite increase to .827 can be compared to the fit of its components, .723 and .691, respectively, for Descriptors I and II. The normal scores of the respective T_{r1} values are $\hat{\theta}_{r1} = 3.61$ and $\hat{\theta}_{r1} = 3.33$. In this example case, then, the combination of the descriptors appears to provide a near perfect fit to the R-configuration, one that is clearly superior to either of its components.

The test for statistically assessing the significance of the increase in tau with the addition of a second descriptor (II), following (16), is computed by jackknifing the values of $T_{r2} - T_{r1}$. The estimate of the standard deviation is:

$$\hat{\lambda}_{r12} = [\Sigma\,(T_{r2}^{(-s)} - T_{r1}^{(-s)} - (\bar{T}_{r2} - \bar{T}_{r1}))^2]^{1/2} = .1228.$$

The test statistic, $\hat{\theta}_{r12} = (T_{r2} - T_{r1})/\hat{\lambda}_{r12} = 0.85$, in this case, would not be considered significant, despite the significant increase noted above. This is due to the fact, quite simply, that instead of testing the significance with respect to a null hypothesis of zero, the test is made with a null of the existing T_{r1} of the descriptor already selected.

Conclusion

The problem of directly relating a vector to a symmetric matrix, specifically, reflecting pairwise estimates of distance as could be input to MDS, has been addressed by a new measurement approach. The goal, of course, is improved interpretation of the judgments that underlie perceptual difference. The new approach, to be termed cognitive differentiation analysis (CDA), measures the correspondence of the descriptor vectors to the discriminations between pairs of stimuli summarized in the $n(n-1)/2$ pairwise judgment matrix. Importantly, the assumption placed on the underlying scale of the descriptors is ordinal, thereby permitting data gathering methodologies where an interval assumption would be suspect, such as ranking or sorting along a given dimension.

The possibility of inferring only ordinal relationships in the descriptor judgments gives the model another distinct advantage over the current methodology. Of potential interest may be the joint application of these methods with the traditional vector fitting approach thereby gaining a more complete understanding of the relationship between the two methodologies.

The statistics developed to quantify the relationship bear a formal resemblance to other standard nonparametric measures of association. The unique and novel feature of the statistical issues addressed here, involving the direct comparison of a vector to a matrix, is the obvious difference of the respective degrees of freedom. The approach followed here uses U-statistics to provide a basis to develop test statistics that have a limiting normal distribution. Estimates of variance, then, are obtained by a jackknife procedure.

The generalization of the method to a second descriptor involves the use of a composite vector, under assumptions of equal weighting, that can be assessed in either of two ways. The composite vector can be evaluated with a significance test using no association as a null hypothesis, as in the simple univariate case. Or, it can be tested for significance using the zero order relation of the first descriptor as the null. Results from an artificial example, under a no-error condition, suggests the latter is probably too conservative.

Appendix

In this appendix, the main points of the proofs of Theorems 1 to 4 will be given. For details see Reynolds and Sutrick (1984).

Sketch of Proof for Theorem 1. The terms of $U_{\ell kn}$ are dependent variables so the method of proof is to approximate $U_{\ell kn}$ by a sum of independent random variables and then apply the central limit theorem. One can show

$$nE\left[U_{\ell kn} - \sum_{i=1}^{n} E(U_{\ell kn} \mid Z_i)\right]^2 \to 0 \quad \text{as} \quad n \to \infty.$$

Since convergence in quadratic mean implies convergence in distribution, $U_{\ell kn}$(suitably scaled) has the same limiting distribution as $\Sigma E(U_{\ell kn} \mid Z_i)$, which satisfies the conditions of the Lindeberg-Feller central limit theorem and is limiting normal.

Sketch of Proof for Theorem 2. Since $U_{\ell kn}$ suitably scaled converges to a normal, one can use the delta-method to write $T_{rk} = a_1 U_{1kn} + a_2 U_{2kn} + \varepsilon$ where a_1, a_2 are constants and ε is an error term which is small, hence T_{rk} is essentially a linear combination of limiting normals and hence limiting normal.

Sketch of Proof for Theorem 3. Under the null hypothesis $H_0 : f \equiv 0$ $P(h_1 = 1) = P(h_2 = 1)$, hence using a symmetry of distribution argument one can conclude $ET_{rk} = 0$. One can show that $\hat{\lambda}^2_{T_{rk}}$ converges in probability to Var (T_{rk}) and the theorem follows from Theorem 2 and Slutsky's theorem.

Sketch of Proof of Theorem 4. Basically the same proof as that of Theorem 3 applies.

References

Chang, J. J., & Carroll, J. D. (1969). How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation. Murray Hill, NJ: Bell Telephone Laboratories.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association, 49*, 732–764.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics, 19*, 293–325.

Kendall, M. G. (1955). *Rank correlation methods* (2nd ed.). London: Charles W. Griffin.

Klahr, D. (1969). A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. *Psychometrika, 34*, 319–330.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53*, 814–861.

Kruskal, J. B. (1964). Multidimensional scaling of optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1–27.

Randles, R. H., & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics.* New York: John Wiley & Sons.

Reynolds, T. J. (1983). A Nonmetric approach to determine the differentiation power of attribute ratings with respect to pairwise similarity judgements. Paper presented at the AMA Educators Conference in Research Methods and Casual Modeling, Sarasota, FL.

Reynolds, T. J., & Sutrick, K. H. (1984). *Ordinal regression when the independent and dependent variables have different degrees of freedom* (Working Paper Series No. 15-9-84). Dallas: University of Texas.

Shepard, R. N. (1962). The Analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika, 27*, 125–140.

Simon, G. A. (1978). Efficacies of measures of association for ordinal contingency tables. *Journal of the American Statistical Association, 73*, 545–551.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27*, 799–811.

Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika, 30*, 379–393.

Manuscript received 4/16/84
Final version received 7/25/85