Dimensions and Clusters: A Hybrid Approach to Classification

Harvey A. Skinner
Addiction Research Foundation

A hybrid strategy is described for integrating the dimensional and discrete clusters approaches to classification research. First, a parsimonious set of dimensions is sought through a multiple replications design. The computations employ a two-stage least squares solution that is based on a sequential application of the Eckart and Young (1936) decomposition. Second, relatively homogeneous subgroups are identified within this low dimensional space using a clustering or density search algorithm. To facilitate interpretation of the final solution, an ideal type concept is introduced that is similar to the "idealized individual" interpretation of multidimensional scaling. Depending upon the model chosen, the independent contribution of elevation, scatter, and shape parameters may be differentiated in defining profile similarity.

At the most fundamental level, classification methods attempt to fit a geometric structure to a given data set. Popular structures include hierarchical trees, discrete clusters, and dimensional models. These structures provide convenient heuristics for helping the investigator describe and understand a certain substantive area. Also, they facilitate communication among researchers. Degerman (1972) provides a review of various geometric representations, as well as orthogonal and nested composites of these structures. Initially, quantitative methods developed

for classification research or "numerical taxonomy" (Sneath & Sokal, 1973) concentrated upon single structures, although the possibilities for fitting hybrid models were certainly recognized (Torgerson, 1965). As researchers attempt to analyze more complex data, considerable interest is growing in methods for simultaneously fitting two or more structures. Indeed, Carroll (1976) speculates that the use of multiple-tree structures and hybrid models is a future trend in data analysis.

The classification problem involves three fundamental aspects. First, some definition of similarity must be expressed along with an index for quantifying the degree of resemblance among individuals (operational taxonomic units). Inherent to the first stage is the assessment of relevant attributes (characters) from which alternative similarity coefficients may be calculated. Euclidean distance and correlation are two frequently used indices in the behavioral sciences. The second stage entails the allocation of individuals into definable subgroups through application of some clustering algorithm. A distinction may be drawn among *hierarchical models*, in which individuals are classified into successive subsets to form a tree structure (Blashfield, 1976); *clustering strategies*, which either directly or iteratively allocate individuals into subgroups (Hartigan, 1975); and *ordination methods*, which display relationships among

individuals in a parsimonious multidimensional space (Rohlf, 1972). Finally, the identification stage involves the classification of new samples according to the taxonomy developed in the first and second stages. Multiple discriminant analysis is often used in this respect.

Each classification method and corresponding geometric structure reflects a markedly different conceptualization of what constitutes a "type." This distinction among methods can have profound implications for a particular research domain. For example, a vast literature surrounds the controversy regarding whether psychiatric syndromes are categorical, dimensional, or hierarchical in nature (Kendell, 1975; Strauss, 1973). At least five theoretical alternatives have been postulated for a typology of depression (Kendell, 1976): (1) they are distinct illnesses (clusters); (2) they are variants of a single depressive illness; (3) they are poles of a psychotic/neurotic continuum; (4) they represent two orthogonal dimensions; and (5) they are arranged in a hierarchy. Given the diversity of classification procedures and criteria, a major issue for researchers is selection of the most appropriate technique. This choice among methods must judge the appropriateness of the geometric structure to underlying theoretical considerations (Skinner, 1977a).

In this paper a general data analysis strategy is described for integrating various elements of the classification problem. First, a dimensional model is developed in which relationships among individuals are represented as points or vectors in a multidimensional space. The primary aim is to achieve a satisfactory solution in the minimum number of dimensions. Emphasis is placed upon (1) identification of a preliminary set of dimensions and (2) replication of these basis dimensions across samples. In the second stage, subsets are formed by grouping individuals that are "close together" in the space. This may be accomplished by application of a clustering algorithm to interpoint distances, or simply by visual inspection of the dimensional plot. Thus, the present hybrid strategy assim-

lates the continuous property of a dimensional representation with the search for discrete clusters in this low dimensional space.

Although a dimensional analysis such as principal components is sometimes conducted as a preliminary to another classification strategy (e.g., Wolfe, 1978), the current approach underscores the substantive interpretation of these dimensions. To motivate the theoretical development, an ideal type concept is proposed that is similar to the "idealized individual" interpretation of individual differences in multidimensional scaling (Cliff, 1968).

Dimensional Representation

The principal methodological task is to identify a "best" set of dimensions for depicting the relationships among individuals. A least squares estimate is derived through a sequential application of the Eckart and Young (1936) decomposition. The first stage identifies dimensions separately within each of m samples. Then, through a replication design the second stage evaluates convergence among these dimensions across samples. Finally, the generalizability of these basis dimensions to new samples is evaluated.

Consider a data matrix X giving the observations of N individuals on k attributes. Further, the N individuals may be partitioned by row into m samples of size n_j ($j = 1, \dots, m$). It is assumed that the k attribute measures are expressed in a compatible scale (e.g., standard scores). Otherwise, certain attributes may assume excessive weight in determining the degree of resemblance among individuals (Sneath & Sokal, 1973, pp. 152–157). Let X_j represent the $n_j \times k$ data matrix for sample j ; $\bar{X}_j$ denote the row centered matrix (each individual's mean is subtracted from profile attributes); and Z_j represent the row standardized matrix (row mean = 0.0, standard deviation = 1.0). For convenience only, assume that $n_j > k$ and that entries in X_j , $\bar{X}_j$, and Z_j have been scaled by $1/\sqrt{k}$.

The preliminary scaling operations upon X_j determine whether the investigator wants to dif-

ferentiate the contribution of elevation, scatter, and shape in defining profile similarity (Cronbach & Gleser, 1953). Shape denotes the actual pattern of "ups and downs" across attribute measures in the profile, scatter describes the dispersion of scores about the mean, and elevation is the mean score over the k attributes in the profile. As discussed by Skinner (1978), in moving from X_j to $\bar{X}_j$ differences in profile elevation are eliminated; in moving to Z_j all profiles are equated with respect to scatter. Thus, X_j confounds all three resemblance parameters, $\bar{X}_j$ confounds scatter and shape, but Z_j only contains information pertaining to shape. Computationally, this scaling parallels the differences among raw cross-products ($X_j X'_j$), covariances ($\bar{X}_j \bar{X}'_j$), and correlations ($Z_j Z'_j$) as the resemblance index among individuals.

This distinction among elevation, scatter, and shape is quite relevant to certain research problems. For example, Sneath (1975) has proposed a vector model of disease states in clinical medicine. The concept of size (elevation) provides a measure of the severity of a disease, and shape depicts the clinical pattern or type of disease. Similarly, in the model of psychopathology developed by Skinner and Jackson (1978), a distinction is made between MMPI profile shape (type of syndrome) and elevation (degree of maladjustment).

Within-Sample Analysis

The computational strategy will be described using the row standardized data matrix Z_j . This matrix is decomposed according to the Eckart and Young (1936) theorem, that is,

$$Z_j = U_j \Gamma_j W'_j \quad [1]$$

where $U_j (n_j \times k)$ and $W_j (k \times k)$ are matrices whose columns are, respectively, the left-hand and right-hand eigenvectors of $Z_j (U'_j U_j = W'_j W_j = I)$, and Γ_j is a diagonal matrix of singular values. Horst (1965, p. 81) calls Equation 1 the "basic structure" of Z_j , whereas the term "sin-

gular value decomposition" is used in numerical analysis (Stewart, 1973, p. 317). Computer algorithms for this decomposition are available (e.g., Dobson & Hepper, 1974; Golub & Reinsch, 1970).

The principal component weightings A_j among individuals are given by

$$A_j = U_j \Gamma_j \quad [2]$$

The scaling of Z_j for Equation 1 means that A_j is equivalent to results that would be obtained from a traditional Q-technique analysis of the $n_j \times n_j$ correlation matrix R_j among individuals (i.e., $R_j = Z_j Z'_j = A_j A'_j$). Upon inspecting the distribution of eigenvalues Γ_j^2 and adoption of some criterion, such as Cattell's (1966) scree test, only the first t_j components ($t_j \leq k$) are retained. Consequently, A_j is now of order $n_j \times t_j$ and W_j is $k \times t_j$.

For most research applications the investigator should consider a simple structure transformation of A_j . The aim is to ensure that each component is interpretable primarily with respect to a specific subset of individuals. Otherwise, this dimension will have little empirical import. The task is to find a $t_j \times t_j$ nonsingular transformation matrix T of unit-length columns (Hakstian, 1974). Then, the primary pattern matrix B_j and primary structure matrix C_j of order $n_j \times t_j$ are given by

$$\begin{aligned} B_j &= A_j (T')^{-1} \quad \text{and} \\ C_j &= A_j T. \end{aligned} \quad [3]$$

Analytic criteria for determining this oblique transformation are given by Harris and Kaiser (1964) and by Katz and Rohlf (1975), among others. The primary component intercorrelation matrix Φ of order $t_j \times t_j$ is given by $\Phi = T' T$. For the restricted case of an orthogonal transformation such as the varimax criterion, Equation 3 simplifies to $B_j = C_j = A_j T$, and the component intercorrelations yield an identity matrix $\Phi = T' T = I$.

Finally, principal component scores are generated that relate each attribute to the first t_j component weightings, that is,

$$Y_j = W_j T. \quad [4]$$

Y_j describes the projection of each attribute measure on principal components of the person-space (Gabriel, 1971). Skinner (1977b) has introduced the term "modal profile" to describe a column vector of Y_j . Each modal profile gives the defining vector of k attribute scores for individuals who have a salient weight on this specific dimension in B_j or C_j .

By retaining only the first t_j components, the effects of measurement error may be largely discarded in the residual E_j (Gleason & Staelin, 1973). From the Eckart and Young (1936) theorem, B_j and Y_j provide a rank t_j least squares estimate of Z_j . Then

$$Z_j = B_j Y_j' + E_j \quad [5]$$

where trace $E_j' E_j$ is a minimum for a given rank t_j solution.

Replication Across Samples

Let t represent the total number of within-sample dimensions retained from the m samples (i.e., $t = \sum t_j$). The $k \times t$ supermatrix Y is formed by adjoining horizontally the m Y_j matrices. Assume that Y is scaled by $1/\sqrt{k}$. Next, consider a single rank i component score matrix U that fits the m sets of within-sample scores Y according to a least squares criterion. Guidelines for determining the number of components i are discussed below. These k by i component scores U represent an estimate of a population component score matrix that is hypothesized to underlie the m samples (cf. Cliff, 1966; Meredith, 1964). The solution is given by an Eckart and Young (1936) decomposition of Y . Assuming $k \geq t$, then

$$Y = U \Gamma W' \quad [6]$$

where $U(k \times t)$ and $W(t \times t)$ are matrices whose columns are, respectively, the left-hand and right-hand eigenvectors, and Γ is a diagonal matrix of corresponding singular values. The degree of fit of the model to the m sets of within-sample dimensions Y is measured by the trace of Γ for a given rank i solution ($i = 1, \dots, t$).

The $t \times t$ matrix of principal component weightings is given by

$$A = W \Gamma \quad [7]$$

where the entries in A provide a resemblance measure that compares the shape of within-sample component scores (modal profiles) from Y_j with a respective column vector of U . Let

$$A = Y' U. \quad [8]$$

The next step is to consider a transformation of the principal component weightings A and component scores U from their principal axes orientation. Note that the degree of fit of the model is invariant under nonsingular transformations of A and U . One approach is to seek simple structure by attempting to locate component axes that have a salient weighting from *only one* within-sample dimension in each sample, i.e.,

$$B = A(T')^{-1} \quad [9]$$

where $B(t \times i)$ is the primary pattern matrix and T is an $i \times i$ nonsingular transformation. Corresponding component scores are transformed by

$$P = U T \quad [10]$$

to yield an estimate of the common dimensions underlying the m samples. The primary com-

ponent intercorrelation matrix Φ of order i is given by

$$\Phi = T' T. \quad [11]$$

Optional Strategy for Determining the Number of Common Dimensions

The present approach seeks a robust classification through sequential analysis of several data samples and through emphasis on a reduced rank solution to minimize the effects of measurement error. One guideline for determining the number i of common dimensions is to examine the distribution of singular values Γ from Equation 6. The scree test (Cattell, 1966) or some other preferred method could be employed to ascertain i . Alternatively, an empirical strategy is proposed. However, this suggestion is optional and is not central to the basic model.

Random Data

Consider m matrices ($n_i \times k$) similar to Z_j , except that each entry is derived from a random number generating function. These random data matrices, designated $\hat{Z}_j$, may be analyzed according to the rationale of Equations 1 through 11. Let

$$\begin{aligned} \hat{Z}_j &= \hat{U}_j \hat{\Gamma}_j \hat{W}_j' \\ &= \hat{B}_j \hat{Y}_j' + \hat{E}_j \end{aligned} \quad [12]$$

where $\hat{B}_j$ are the entity component weightings and $\hat{Y}_j$ are the within-sample component scores. Note that the number of components retained (t_j) parallels the real-data analysis of Z_j . The common dimensions $\hat{P}$ are estimated by an Eckart and Young (1936) decomposition of $\hat{Y}$, i.e.,

$$\hat{Y} = \hat{U} \hat{\Gamma} \hat{W}' \quad [13]$$

where $\hat{P}$ is formed from the first i columns of $\hat{U}$.

The singular values Γ reflect the extent to which within-sample component scores $\hat{Y}_j$ match across samples due to chance. Thus, in the tradition of Horn (1965), one strategy for determining the number of common dimensions is to compare the singular values Γ from Equation 6 based on real data with respective singular values $\hat{\Gamma}$ due to chance.

Permuted Data

With the random data matrices both individuals and attributes are randomly generated in $\hat{Z}_j$. A stronger test (cf. Pinneau & Newhouse, 1964, p. 273) is to consider the individuals fixed in each sample but to assume that the attribute measures are randomly determined. Consider a permuted data matrix $\bar{Z}_j$, which is simply the original data matrix Z_j in which the columns have been randomly rearranged without changing the column headings. Note that intercorrelations among the individuals within each sample are not disturbed by this permutation. The left-hand eigenvectors U_j and singular values Γ_j of $\bar{Z}_j$ are identical to Equation 1, i.e.,

$$\begin{aligned} \bar{Z}_j &= U_j \Gamma_j \bar{W}_j' \\ &= B_j \bar{Y}_j' + \bar{E}_j. \end{aligned} \quad [14]$$

Only the right-hand eigenvectors $\bar{W}_j$ and within-sample component scores $\bar{Y}_j$ are permuted. An Eckart and Young (1936) decomposition of $\bar{Y}$ estimates the modal profiles $\bar{P}$, where the attributes are assumed random. Let

$$\bar{Y} = \bar{U} \bar{\Gamma} \bar{W}' \quad [15]$$

where $\bar{P}$ is formed from the first i columns of $\bar{U}$.

To implement the permutation strategy, the within-sample component scores Y_j of Equation 4 may be used. That is, randomly permuting the rows of Y_j producing $\bar{Y}_j$ is identical to results from applying the same permutation by column

to Z_j forming $\bar{Z}_j$, and then solving for $\bar{Y}_j$ by Equations 1 and 4. This procedure may be repeated using "fresh" random numbers for each permutation run to yield a sampling distribution for the singular values $\bar{\Gamma}$.

In the spirit of guidelines emanating from the multidimensional scaling literature for determining the "true" dimensionality underlying a proximity matrix (Spence & Graef, 1974; Stenson & Knoll, 1969), confidence intervals for the real singular values Γ (or eigenvalues Γ^2) may be estimated. Figure 1 displays a plot of three sets of eigenvalues (real, random, average of 100 permutation runs), as well as a line two standard deviations above the mean permuted values $\bar{\Gamma}^2$.

The data are based on six samples ($N = 661$) of patients who had taken a personality inventory (Skinner, Reed, & Jackson, 1976). Assuming that estimates of the eigenvalues due to chance are normally distributed, it would be expected that 97.5% of the eigenvalues would fall below the two standard deviation upper boundary if the patients were responding randomly to the test. The first eight real eigenvalues Γ^2 are clearly above the confidence interval. When the real eigenvalues intersect a selected confidence boundary, the investigator must entertain the hypothesis that the remaining within-sample component scores are matched solely on the basis of chance.

Figure 1

Distribution of eigenvalues for real (Δ^2), perturbed, ($\bar{\Delta}^2$) and random ($\hat{\Delta}^2$) data

Generalizability of the Common Dimensions

The third stage evaluates the degree to which the basis dimensions P are representative of the derivation and new samples. Individuals from the row standardized data matrix Z_j may be projected into this dimensional space by

$$Q_j = \frac{Z_j P}{k} \quad [16]$$

where Q_j is the $n_j \times i$ matrix of component structure coefficients. Then, the matrix $F_j(n_j \times i)$ of component pattern coefficients is given by

$$F_j = Q_j \Phi^{-1} \quad [17]$$

where Φ is the component intercorrelation matrix of Equation 11. Of course, in the case of an orthogonal transformation in Equations 9 and 10, the structure and pattern coefficients are identical, since $\Phi^{-1} = I$.

Observe that

$$Z_j = F_j P' + E_j \quad [18]$$

where E_j is the $n_j \times k$ residual matrix. A goodness-of-fit measure of the model to sample j is given by trace $(E_j E_j')$. The entries in F_j (or Q_j) provide coordinates for an ordination plot, whereby a graphic display may be constructed of the relationships among entity profile due to similarity in profile shape. Note that Equation 16 is analogous to the factor extension procedure originally proposed by Dwyer (1937).

A second measure of fit for the model is to consider classification hit rates. That is, a dimension (column vector of P) has empirical import only to the extent that it is characteristic of a subset of entities in the sample. Cliff (1968) raised the same issue in cautioning the avoidance of idealized individual vectors that do not correspond to the judgments of real subjects. It is conceivable that the basis dimensions may account for a reasonable portion of the variance,

and yet there may be few entities in the sample with salient weightings. Using an entrance criterion, such as a highest weighting in F_j above .50, classification hit rates may be computed for each dimension.

For example, classification percentages and variance-explained estimates are given in Table 1 for three samples of alcoholic patients. Seven modal profiles derived from the two male groups (Skinner, Jackson, & Hoffmann, 1974) were also fitted to the (independent) female sample. Both measures of fit were somewhat larger in the derivation (male) samples, with the classification hit rate around 70% and the variance-explained estimate approximately 60%. In all cases, the measures of fit based on the real modal profiles P substantially exceeded the permuted $\bar{P}$ and random $\hat{P}$ estimates.

General Models

Let M_j represent the $n_j \times 1$ vector of row means and S_j denote the $n_j \times 1$ vector of row standard deviations. Relationships among observed profiles for individuals in sample j may be expressed as a function of

$$X_j = f \left\{ M_j, S_j, (F_j P' + E_j) \right\} \quad [19]$$

where M_j is a vector of *elevation* parameters, S_j is a vector of *scatter* parameters, and F_j is an $n_j \times i$ matrix of *shape* parameters from Equation 17. The F_j matrix describes the degree to which an individual's profile resembles the component score vectors in P . Skinner (1977b, 1978) discusses properties of this general model, where emphasis is placed upon differentiating the contribution of elevation, scatter, and shape parameters.

For research problems where the differentiation among elevation, scatter, and shape is not applicable, the data analyses could be based upon X_j . In this case, the equations would corre-

Table 1
Measures of Fit

Measure of Fit and Group	<i>N</i>	Common <i>P</i>	Permuted $\bar{P}$	Random $\hat{P}$
Classification Percentage				
Male Alcoholics (A)	141	68.1%	5.0%	6.4%
Male Alcoholics (B)	141	70.9%	7.1%	7.1%
Female Alcoholics	122	50.8%	4.1%	9.0%
Variance Explained				
Male Alcoholics (A)	141	59.8%	27.8%	24.7%
Male Alcoholics (B)	141	61.8%	27.2%	25.5%
Female Alcoholics	122	54.6%	27.8%	25.6%

spond to *raw cross-products* (Skinner, 1978), and the general model may be stated as

$$X_j = G_j P' + E_j \quad [20]$$

where G_j is an $n_j \times i$ component weighting matrix. Note that in the analysis of raw cross-products the first principal component often resembles the entity means M_j or person main effects (Gollob, 1968). Alternatively, the investigator may wish to partial out the influence of profile elevation and to proceed with the row-centered data matrix $\bar{X}_j$. Then, the equations would correspond to a *covariance* analysis. The general model may be depicted as

$$X_j = f \left\{ M_j, \{H_j P' + E_j\} \right\} \quad [21]$$

where H_j is a $n_j \times i$ component weighting matrix. Note that the H_j matrix would display relationships among individuals due to both profile scatter and shape.

in sample j . If most of the covariation can be summarized in two (or possibly three) dimensions, then one can visually search for distinct clusters in this space. However, a more rigorous strategy, especially with higher dimensionality, is to apply a clustering algorithm. Consider the component weighting matrix F_i of Equation 17. Interpoint distances among individuals in this i dimensional space may be computed by

$$D_{12}^2 = \sum_i (F_{1n} - F_{2n})^2 \quad [22]$$

where d_{12}^2 is the squared euclidean distance between Subject 1 and Subject 2. These distance indices for all n_i individuals may be used as input to a clustering algorithm such as K - Means (Hartigan, 1975) to identify homogeneous clusters in the dimensional space. Furthermore, relationships among the clusters may be examined with respect to their salience on these basis dimensions or modal profiles (Skinner, 1977b). For example, two clusters may represent opposite poles of a dimension characterized by the severity of depressive symptoms (Kendell, 1976).

Search For Homogeneous Subsets

The entries in F_j , G_j , or H_j provide coordinates for an ordination plot of the individuals

Ideal Type Construct

In an important clarification of the Tucker and Messick (1963) points-of-view model, Cliff

(1968) introduced an "idealized individual" concept as a means for interpreting each point of view in the scaling experiment. That is, each idealized individual (and corresponding vector of mean judgments) is used to represent or to summarize the actual judgments of a subset of subjects who are near the idealized individual vector in the person-space. Furthermore, this heuristic was recommended by Cliff (1968) when applying Tucker's (1966) individual differences model to learning curve data. In brief, Cliff contends that "the outcome of points-of-view or learning curve analysis is a set of few response vectors . . . which may be substituted for the actual responses of any individual without great loss of information" (p. 231).

The present classification strategy is based upon a vector model that has several methodological and conceptual similarities to the Tucker and Messick (1963) and Tucker (1966) procedures. In the strategy described in this paper, each point-of-view concept corresponds to a subgroup of individuals (depicted in F_j , G_j , or H_j) who exhibit a similar profile of attribute measures. The centroid of each homogeneous cluster is analogous to Cliff's (1968) "idealized individual" concept. This centroid could be estimated by averaging the coordinates (F_j , G_j , or H_j) of individuals that define the cluster. Given this $1 \times i$ centroid vector designated C_s for cluster s , the corresponding vector of attributes may be computed by

$$D_s = C_s P^t \quad [23]$$

where D_s is the $1 \times k$ vector of attribute measures. That is, the attribute profile D_s may be interpreted as the characteristic profile for individuals categorized according to subgroup s .

Two dimensions based on various deviant groups (Skinner et al., 1976) are plotted in Figure 2. The Differential Personality Inventory scales (Jackson & Messick, 1971) that characterize each pole are depicted, as well as two hypothetical clusters, A and B. The centroid coordinates for Cluster A are (.65, .65) and for

Cluster B (.20, -.80). Following Equation 23, the ideal type vector D_A may be computed for Cluster A in this two-dimensional example simply by postmultiplying the centroid vector coordinates (.65, .65) by the first two column vectors in P . Similarly, an ideal type vector D_B may be determined for Cluster B. These ideal type concepts may be used to form the basis of a theoretical model (Skinner, 1977a, 1977b). Note that Sneath (1977) has proposed a statistical test for assessing the distinctness of clusters in Euclidean space.

The attribute profile D_s represents an ideal type construct. The notion of ideal types as intervening variables has a considerable history (Dahlstrom, 1972; Hempel, 1965). Each ideal type may be defined as a hypothetical profile pattern that is characteristic of a subset of individuals in the population. Certain ideal types are hypothesized in a population with each individual more or less related to each type. That is, an individual's profile may be depicted as a weighted mixture of these ideal types. Kretschmer (1925, p. 93), in an interesting metaphor, has termed this mixture the "constitutional alloy." This conceptual development is similar in rationale to the centroid organism construct in biological research (Sneath & Sokal, 1973, p. 195) and to linear typal analysis described by Overall and Klett (1972, p. 219).

Finally, an interesting possibility for enhancing the interpretation of the classification is to include certain "contrived" profiles as marker variables with sample j in Equations 19, 20, or 21. These contrived persons could represent types established in previous research, as well as low base rate individuals or "freaks" (cf. Sandler, 1952). The location of these marker persons within the i dimensional space could provide important benchmarks for interpreting relationships among individuals.

Example

An example is taken from two samples, I ($n_1 = 123$) and II ($n_2 = 122$), of male delinquent ado-

Figure 2

Two hypothetical clusters (A and B) plotted on Dimension 1
versus Dimension 2

lescents who had completed the Basic Personality Inventory (BPI; Jackson, 1976). The BPI is an 11-scale structured inventory of psychopathology that was developed using a construct validation paradigm (Jackson, 1971). The data were analyzed according to Equations 1 through 5. Four components ($t_1 = t_2 = 4$) were retained within each sample and transformed to an orthogonal varimax criterion. The resulting within-sample profiles Y_j are given in Table 2. For cosmetic reasons, the profiles have been scaled to T scores (mean of 50, standard deviation of 10).

A visual inspection of Table 2 suggests that the first two profiles across each sample are notably similar.

A valuable interpretative step is to compute intercorrelations among the sample profiles, that is, $R = Y'Y$. This multiprofile-multisample supermatrix is presented in Table 3 for the two adolescent samples. Since an orthogonal transformation was conducted in Equations 3 and 4, within-sample correlations form an identity matrix. Thus, an interesting feature of R in this case is that between-sample submatrices contain

Table 2
Within-Sample BPI Component Scores Y_j

BPI Scale	Sample I				Sample II			
	I	II	III	IV	I	II	III	IV
Hypochondriasis	59	49	41	27	60	46	50	44
Depression	56	42	59	44	52	40	53	55
Denial	61	76	46	59	62	76	59	49
Interpersonal Problems	35	54	59	53	33	51	56	25
Social Deviation	33	58	42	44	34	62	42	62
Persecutory Ideas	46	51	47	41	50	50	28	46
Anxiety	52	37	37	60	59	42	45	48
Thinking Disorder	54	42	37	65	53	48	41	55
Impulse Expression	37	45	57	55	37	46	61	61
Social Introversion	60	51	67	54	56	50	51	52
Self-Depreciation	57	45	59	49	53	40	64	51

Note. Each profile has been scaled to have a mean = 50, standard deviation = 10.

the multiple regression of each variable (within-sample profile) for one sample regressed on all variates of the other sample. The first two profiles demonstrate a high degree of correspondence (.95 and .94, respectively), and the third profile across each sample has moderate similarity (.51). Multiple correlations (R_m) may

be readily computed. That is, the multiple correlation for predicting the first profile in Sample I from profiles in Sample II is given by summing the square of entries in the first row of the submatrix R_{12} and taking the square root of this sum $((.95)^2 + (-.05)^2 + (.16)^2 + (.01)^2 = .93, R_m = .96)$.

Table 3
Multiprofile-Multisample Supermatrix R

	Sample I				Sample II			
	I	II	III	IV	I	II	III	IV
Sample I								
I	1.00	.00	.00	.00	.95	-.05	.16	.01
II	.00	1.00	.00	.00	.00	.94	.17	-.14
III	.00	.00	1.00	.00	-.22	-.23	.51	-.08
IV	.00	.00	.00	1.00	.01	.18	.15	.10
Sample II								
I	.95	.00	-.22	.01	1.00	.00	.00	.00
II	-.05	.94	-.23	.18	.00	1.00	.00	.00
III	.16	.17	.51	.15	.00	.00	1.00	.00
IV	.01	-.14	-.08	.10	.00	.00	.00	1.00

Table 4
Principal Components Solution

Principal Axes	Component Weightings A							
	1	2	3	4	5	6	7	8
Sample I								
I	.06	.98	.16	-.04	.04	.08	.10	.00
II	.96	-.10	.12	-.16	.14	.06	-.01	.04
III	-.19	-.17	.85	-.12	.11	.44	-.02	-.01
IV	.18	.01	.18	.73	-.63	.09	.00	.01
Sample II								
I	.10	.99	-.05	-.03	.02	.03	-.10	.00
II	.98	-.10	-.10	.07	.06	.05	.01	-.04
III	.10	.05	.87	.11	.09	-.45	.00	.00
IV	-.10	.04	-.12	.74	.65	.06	.00	.00
Eigenvalues	1.997	1.979	1.573	1.137	.863	.427	.021	.003

Next, the principal component weightings from Equation 7 are given in Table 4; Table 5 gives a varimax transformation of the first three components (Equation 9). For the special case of two samples ($m = 2$) along with an orthogonal transformation during the within-sample analyses, this between-sample stage is analogous to employing canonical correlation criteria for transforming both sets of component scores Y_j to maximal congruence (Kettenring, 1971). That

Table 5
Varimax Rotation of
the First Three Components

	1	2	3
Sample I			
I	.99	-.03	.10
II	.00	.97	.10
III	-.13	-.15	.86
IV	.03	.19	.17
Sample II			
I	.99	.01	-.12
II	-.02	.99	-.12
III	.12	.12	.86
IV	.02	-.11	-.12

is, for the first t_1 roots of R , the eigenvalues equal the respective canonical correlation plus unity. The remaining t_2 roots either correspond to a zero canonical correlation or provide redundant information. Thus, for $m = 2$ samples, eigenvalues > 1.0 provide an upper bound for the number of modal profiles i to retain. The magnitude of the eigenvalues Γ^2 (or respective canonical correlations) display the extent to which the sample component scores Y_j may be transformed to congruence with the $k \times i$ estimate U of the population component score matrix.

To continue the example, the canonical correlations from matching the two sets of component scores Y_j to maximal congruence were .997, .979, .573, and .137, respectively. This distribution suggests that three profiles should be retained in U . In the varimax solution, each component clearly isolated (replicated) one profile for each sample. Finally, Table 6 lists the three corresponding BPI modal profiles from Equation 10, scaled to resemble T scores. The first dimension contrasts Denial and Hypochondriasis with Social Deviation and Impulse Expression. Adolescents at the latter pole of this dimension would be rebellious, hold antisocial attitudes, and present problems with behavioral control.

Table 6
BPI Modal Profiles (P)

BPI Scale	Modal Profile		
	I	II	III
Hypochondriasis	60	46	43
Depression	54	40	56
Denial	62	77	54
Interpersonal Problems	34	54	60
Social Deviation	34	59	39
Persecutory Ideas	46	49	36
Anxiety	56	42	41
Thinking Disorder	54	47	39
Impulse Expression	37	46	59
Social Introversiion	57	49	60
Self-Depreciation	55	43	63

Note. The modal profiles are scaled to have a mean = 50, standard deviation = 10.

The second dimension, on the other hand, is marked by a tendency to minimize problem areas (Denial) versus a depression syndrome. The third dimension contrasts introverted adolescents who encounter interpersonal problems with other adolescents who are quite suspicious and untrusting of others. Thus, the modal profiles of Table 6 formed a three-dimensional model of psychopathology among delinquent adolescents.

Conclusion

In conclusion, the present data analysis strategy may be used to integrate classification methodology with the development of typological theories (cf. McQuitty, 1967). However, the choice of relevant samples, the reliable and valid assessment of important attributes, and the development of a cogent theoretical model integrating the ideal type constructs ultimately rest upon the individual researcher. Perhaps Johnson's (1968) penetrating remarks best capture the essence of classification research which he considers "an art, but a disciplined and respectable art whose aim is to serve science" (p. 235).

References

- Blashfield, R. K. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 1976, 83, 377-388.
- Carroll, J. D. Spatial, nonspatial, and hybrid models for scaling. *Psychometrika*, 1976, 41, 439-463.
- Cattell, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*, 1966, 1, 245-276.
- Cliff, N. Orthogonal rotation to congruence. *Psychometrika*, 1966, 31, 33-42.
- Cliff, N. The "idealized individual" interpretation of individual differences in multidimensional scaling. *Psychometrika*, 1968, 33, 225-232.
- Cronbach, L. J., & Gleser, G. C. Assessing similarity between profiles. *Psychological Bulletin*, 1953, 50, 456-473.
- Dahlstrom, W. G. *Personality systematics and the problem of types*. Morristown, NJ: General Learning Press, 1972.
- Degerman, R. L. The geometric representation of some simple structures. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling* (Vol. 1). New York: Seminar Press, 1972.
- Dobson, R., & Hepper S. J. SINGD: A PL/I program for Eckart-Young singular decomposition.

- Behavior Research Methods and Instrumentation*, 1974, 6, 60.
- Dwyer, P. S. The determination of factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 1937, 2, 173-178.
- Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211-218.
- Gabriel, K. R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 1971, 58, 453-467.
- Gleason, T. C., & Staelin, R. Improving the metric quality of questionnaire data. *Psychometrika*, 1973, 38, 393-410.
- Gollob, H. F. Confounding of sources of variation in factor-analytic techniques. *Psychological Bulletin*, 1968, 5, 330-344.
- Golub, G. H., & Reinsch, C. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 1970, 14, 403-420.
- Hakstian, A. R. The development of a class of oblique factor solutions. *British Journal of Mathematical and Statistical Psychology*, 1974, 27, 100-114.
- Harris, C. W., & Kaiser, H. F. Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 1964, 29, 347-362.
- Hartigan, J. A. *Clustering algorithms*. New York: Wiley, 1975.
- Hempel, C. G. *Aspects of scientific explanation*. New York: The Free Press, 1965.
- Horn, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 1965, 30, 179-185.
- Horst, P. *Factor analysis of data matrices*. New York: Holt, Rinehart, & Winston, 1965.
- Jackson, D. N. The dynamics of structured personality tests: 1971. *Psychological Review*, 1971, 78, 229-248.
- Jackson, D. N. *Basic Personality Inventory*. London, Ontario, Canada: Author, 1976.
- Jackson, D. N., & Messick, S. *Differential Personality Inventory*. London, Ontario, Canada: Authors, 1971.
- Johnson, L. A. S. Rainbow's end: The quest for an optimal taxonomy. *Systematic Zoology*, 1968, 119, 203-239.
- Katz, J. O., & Rohlf, F. J. Primary product function plane: An oblique rotation to simple structure. *Multivariate Behavioral Research*, 1975, 10, 219-231.
- Kendell, R. E. *The role of diagnosis in psychiatry*. Oxford: Blackwell Scientific Publications, 1975.
- Kendell, R. E. The classification of depressions: A review of contemporary confusion. *British Journal of Psychiatry*, 1976, 129, 15-28.
- Kettenring, J. R. Canonical analysis of several sets of variables. *Biometrika*, 1971, 58, 443-451.
- Kretschmer, E. *Physique and character* (Trans. from 2nd German ed. by W. J. H. Spratt). New York: Harcourt, Brace & Co., 1925.
- McQuitty, L. L. A mutual development of some typological theories and pattern analytic methods. *Educational and Psychological Measurement*, 1967, 27, 21-46.
- Meredith, W. Rotation to achieve factorial invariance. *Psychometrika*, 1964, 29, 189-206.
- Overall, J. E., & Klett, C. J. *Applied multivariate analysis*. New York: McGraw-Hill, 1972.
- Pinneau, S. R., & Newhouse, A. Measures of invariance and comparability in factor analysis for fixed variables. *Psychometrika*, 1964, 29, 271-281.
- Rohlf, F. J. An empirical comparison of three ordination techniques in numerical taxonomy. *Systematic Zoology*, 1972, 21, 271-280.
- Sandler, J. A technique for facilitating the rotation of factor axes, based on an equivalence between persons and tests. *Psychometrika*, 1952, 17, 223-229.
- Skinner, H. A. *Construct validation and psychiatric classification*. Paper presented at the Classification Society Meetings, Hanover, NH, June 1977. (a)
- Skinner, H. A. The eyes that fix you: A model for classification research. *Canadian Psychological Review*, 1977, 18, 142-151. (b)
- Skinner, H. A. Differentiating the contribution of elevation, scatter, and shape in profile similarity. *Educational and Psychological Measurement*, 1978, 38, 297-308.
- Skinner, H. A., & Jackson, D. N. A model of psychopathology based on an integration of MMPI actuarial systems. *Journal of Consulting and Clinical Psychology*, 1978, 46, 231-238.
- Skinner, H. A., Jackson, D. N., & Hoffmann, H. Alcoholic personality types: Identification and correlates. *Journal of Abnormal Psychology*, 1974, 83, 658-666.
- Skinner, H. A., Reed, P. L., & Jackson, D. N. Toward the objective diagnosis of psychopathology: Generalizability of modal personality profiles. *Journal of Consulting and Clinical Psychology*, 1976, 44, 111-117.
- Sneath, P. H. A. A vector model of disease for teaching and diagnosis. *Medical Hypotheses*, 1975, 1, 12-22.

- Sneath, P. H. A. A method for testing distinctness of clusters. *Mathematical Geology*, 1977, 9, 123-143.
- Sneath, P. H. A., & Sokal, R. R. *Numerical taxonomy*. San Francisco: Freeman, 1973.
- Spence, I., & Graef, J. The determination of the underlying dimensionality of an empirically obtained matrix of proximities. *Multivariate Behavioral Research*, 1974, 9, 331-341.
- Stenson, H. H., & Knoll, R. L. Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychological Bulletin*, 1969, 71, 122-126.
- Stewart, G. W. *Introduction to matrix computations*. New York: Academic Press, 1973.
- Strauss, J. S. Diagnostic models and the nature of psychiatric disorder. *Archives of General Psychiatry*, 1973, 29, 445-449.
- Torgerson, W. S. Multidimensional scaling of similarity. *Psychometrika*, 1965, 30, 379-393.
- Tucker, L. R. Learning theory and multivariate experiment: Illustration by determination of generalized learning curves. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966.
- Tucker, L. R., & Messick, S. An individual differences model for multidimensional scaling. *Psychometrika*, 1963, 28, 333-367.
- Wolfe, J. H. Comparative cluster analysis of patterns of vocational interest. *Multivariate Behavioral Research*, 1978, 13, 33-44.

Acknowledgements

The research for this paper was supported by the Addiction Research Foundation of Ontario. The author thanks Douglas N. Jackson and W. Carson Smiley for supplying the data used in the examples.

Author's Address

Send requests for reprints or further information to Harvey A. Skinner, Addiction Research Foundation, 33 Russell Street, Toronto, Ontario, M5S 2S1 Canada.