Some New Measures of Profile Dissimilarity

David V. Budescu
University of North Carolina at Chapel Hill

Four new measures of multidimensional profile dissimilarity are proposed that are (1) either symmetric or asymmetric and (2) either conditional or unconditional on profile shape. The four similarity indices are based on alternative normalizations of the regular distance (D) statistic of Cronbach and Gleser (1953), all taking values between 0 and 1. Methods of calculation and interpretations of the indices are demonstrated and discussed, and several generalizations are suggested.

Characterization of a Profile

In his seminal paper on profile analysis, Cattell (1949) proposed to partition the information contained in a profile into three possibly dependent components: level, scatter, and shape. He also noted that the notion of "profile similarity" may have several different meanings, which may lead to various similarity indices depending on the particular components of the profiles examined.

A profile $\mathbf{X}_i$ of individual i is a p dimensional vector, where p is the number of components (scales) of the profile. According to this notation, the score of individual i on scale j is $\mathbf{X}_{ij}$ (j=1 . . .p); and $\mathbf{X}_{.k}$, where k is an integer between 1 and p inclusive, is the general notation for the k^{th} scale.

The level of individual i is the mean of all his or her scores:

$$\bar{x}_i = \frac{1}{p} \sum_{k=1}^{p} x_{ik} \qquad [1]$$

The scatter (or dispersion) of a profile $\mathbf{X}_i$ is defined as

$$s_i^2 = \sum_{k=1}^{p} (x_{ik}-\bar{x}_i)^2 \qquad [2]$$

Nunnally (1967) provides a list of recommendations for interpretation and comparisons of the level and the scatter of profiles.

The remaining information in a profile is defined as the shape (Cronbach & Gleser, 1953). Operationally, the shape is the rank order of the scores of $\mathbf{X}_i$ (in this paper ordering refers always to weak

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 4, No. 2 Spring 1980 pp. 261–272

ordering). This definition of shape, suggested by Nunnally (1967), is an ordinal one. Two profiles, $\mathbf{X}_i$ and $\mathbf{X}_j$, are said to have the same shape if the rank correlation between their p scales is perfect. According to this definition a profile can have any one of $p!$ distinct shapes corresponding to the $p!$ possible orderings of p scales. There are two advantages in this approach to shape. It provides a systematic classification of the infinite number of shapes which can be observed empirically into a finite number of uniquely defined and well-understood classes. It also eliminates the dependence of the shape interpretation on the order in which the scales are listed. This dependence is inherent in any interpretation based on the profile's physical appearance (Nunnally, 1967).

Existing Measures of Dissimilarity

Two classes of indices of similarity between profiles have been proposed (Overall & Klett, 1972): vector product indices and distance indices.

The most popular vector product indices are the regular covariance and product-moment correlations. The former is independent of the profile's level and the latter is also independent of its scatter. This is attributable to the level being subtracted from each observation; and in the correlational techniques the covariance is normalized by the standard deviations. The intraclass correlation (Webster, 1952) is the only correlational index which is a function of all three components of similarity, but it is not very frequently used in this context. Therefore, these indices are mainly considered shape similarity measures. The correlational indices are bounded ($-1 \leqslant R \leqslant 1$), and therefore can be easily interpreted, but a clear weakness is their dependence on the spatial orientation of the scales. Since in many psychological fields the orientation is arbitrary, Cohen (1969) has developed a correlational profile similarity index (R_c) invariant under scale reflections.

The distance indices are all based on the representation of p dimensional profiles as points in a p dimensional space. The dissimilarity between two profiles can be described by the distance between them in this space (Cronbach & Gleser, 1953). The resulting distance statistic (D) is based on the assumption that the scales are uncorrelated or, in other words, that the axes of the space are orthogonal. Overall (1964) suggested use of the Mahalanobis distance function as an index of dissimilarity between profiles $\mathbf{X}_i$ and $\mathbf{X}_j$ in nonorthogonal spaces:

$$D' = (\underline{x}_i - \underline{x}_j)' S^{-1} (\underline{x}_i - \underline{x}_j) \qquad [3]$$

(Here **S** is the variance-covariance matrix of the variables.) Note that this statistic demands information about the intercorrelations between the scales. Both distance measures in their raw form are insensitive to the profile's shape; for any given profile, $\mathbf{X}_i$, there is an infinite number of equidistant profiles without any restriction on their shape. In fact, it can be shown that the raw distance measure is a function of dissimilarity in level, scatter, and shape (Skinner, 1978). Moreover, distance measures lack one of the most appealing features of the correlational measures because they can take any positive value and their interpretation and comparison are difficult.

Not all similarity measures can be strictly categorized into these two classes. Cattell (1949) proposed two indices of "pattern similarity":

$$R'_p = (2p - D^2)/(2p + D^2) \qquad [4]$$

$$R_p = (2k - D^2)/(2k + D^2) \qquad [5]$$

Here K is the median of a chi-square with $(p-1)$ *df*, and D is the regular distance between two profiles calculated on standardized scores. Both Cattell's measures are based on distances but are normalized to take values like a regular correlation $(-1 \leqslant R_p, R_p' \leqslant 1)$; and, indeed, Cattell (1949) recommends this interpretation. However, these indices are based on the assumption that all the scales are uncorrelated and normally distributed and, like all the other distance measures, are independent of a particular shape structure. Unlike the vector product and the distance statistics, R_p can not be factor analyzed because the matrix of R_p values may be non-Gramian (Nunnally, 1962).

Cronbach and Gleser (1953) have shown that the distance between two profiles is related to some of the correlational indices:

$$R_{\text{intraclass}} = 1 - \frac{D^2}{s_i^2 + s_j^2 - \frac{p}{2}(\bar{x}_i - \bar{x}_j)^2} \qquad [6]$$

$$2[1-R] = \frac{D^2 - p(\bar{x}_i - \bar{x}_j)^2 - (s_i - s_j)^2}{s_i s_j} \qquad [7]$$

These relations involve components of shape, scatter, and level. This fact has two important implications: First, the total dissimilarity as measured by the Euclidean distance can be decomposed into separate terms representing the three factors (Skinner, 1978). Second, if the distance is calculated from scores previously adjusted for their dissimilarities in level and scatter, the result is functionally related to some of the vector-product indices. In fact, there exists a one to one relationship between this particular version of the distance and the regular product-moment correlation. It follows that the two statistics contain identical information with regard to shape similarity (Overall & Klett, 1972). However, this relation does not hold for the raw distance measure and, in fact, one reason for the popularity of the D statistic is that it does reflect all the aspects of dissimilarity between vectors. Finally, note that all the measures reviewed above are symmetric, although the psychological notion of symmetric similarity (or dissimilarity) has been recently questioned on theoretical and empirical grounds (Tversky, 1977).

A New Class of Profile Dissimilarity Indices

The purpose of this paper is to propose a new class of dissimilarity indices. Four different indices are proposed which are (1) symmetric or asymmetric and (2) unconditional or conditional upon the shape of one or two profiles. All of them are based on the geometrical model that inspired the D statistic (Cronbach & Gleser, 1953). However, the D statistic is normalized by four different functions (Euclidean distances themselves) to provide ratios of distances, with values between 0 and 1. The novelty of this approach is in the introduction of these normalizing functions and their special nature. The general form of the indices denoted by $B_{k(ij)}$ $(k = 1,4)$, is given by:

$$B_{k(ij)} = \frac{D_{ij} - \min(D_{ij} \mid \text{restriction})}{\max(D_{ij} \mid \text{restriction}) - \min(D_{ij} \mid \text{restriction})} \qquad [8]$$

Here D_{ij} is the Euclidean distance between profiles $\mathbf{X}_i$ and $\mathbf{X}_j$. Each combination of the symmetry and conditionality classifications imposes a particular set of restrictions on the possible distances between a pair of profiles, and the measured distance (D_{ij}) is normalized by a function of the minimal and maximal distances that can be obtained under these restrictions to yield a dissimilarity index ($B_{k(ij)}$) with a value of 0 when the two profiles are as close as possible (under the restrictions) and a value of 1 when they are as distant as possible. All the intermediate values are ratios of distance: the ratio of the obtained distance to the maximal possible distance.

To summarize, like D (Cronbach & Gleser, 1953), the $B_{k(ij)}$ are indices of dissimilarity based on the level, scatter, and shape of the two profiles involved (Skinner, 1978), rescaled to values that allow for a more intuitive and meaningful interpretation. The conditional measures add a new dimension to this class, since they restrict the shape of the profile(s) and reflect only dissimilarities in level and scatter. The asymmetric indices provide new measures reflecting the particular characteristics of the standard points.

As Overall (1964) has shown, the case of orthogonal scales within a profile is only a special case of the general case in which the scales are allowed to have arbitrary correlations. The present discussion will be restricted to this special case, because of its simplicity, but will also include some possible generalizations to the nonorthogonal case.

The Euclidean Space and Distance

A Euclidean space of p dimensions (R_p) is the collection of all p component vectors for which the operations of vector addition and multiplication by a scalar are permissible. Moreover, for any two vectors, $\mathbf{X}_i$ and $\mathbf{X}_j$, in R_p there exist a nonnegative number, D_{ij}, called the Euclidean distance between the two vectors (Green & Carroll, 1976, p. 83). The function that produces the distance is called the Euclidean distance function and is defined as

$$D_{ij} = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \qquad [9]$$

The Euclidean distance function is positive and symmetric, satisfies the triangle inequality, and is invariant under a general class of similarity transformations (Green & Carroll, 1976, p. 288).

When applying the Euclidean model to measurement of psychological profiles dissimilarity, additional assumptions can be made in order to further simplify the model. For example, a bounded space is an Euclidean space in which the values that each dimension (scale) can take are bounded by minimal and maximal values. Therefore, in the bounded space a profile $\mathbf{X}_i$ is restricted such that

$$\underline{L} < \underline{x}_i < \underline{H} \qquad [10]$$

H is a vector including the maximal values that each scale can take, and **L** is the vector of the corresponding minimal values. Theoretically, $\mathbf{H}_k$ and $\mathbf{L}_k$ (k=1, p) can take any real value. However, it is useful to restrict them so that max $(\mathbf{L}_k) <$ min $(\mathbf{H}_k)$. This restriction assures that all $p!$ shapes can be obtained. It is also important to realize that in most psychological applications all p scales of a profile share the same lower and/or upper bounds. Such cases are easier to deal with, since a constant can be added to all p scales such that either **H** or **I** is set equal to the **O** vector. Substituting **O** for one of these two vectors largely simplifies most of the calculations involved in the computation of the dissimilarity indices.

It is possible to partition a p dimensional space into $p!$ isotonic subspaces. All the points in an isotonic space have the same rank ordering along the p dimensions. For example, the collection of all the points that satisfy $\mathbf{X}_{.3} > \mathbf{X}_{.1} > \mathbf{X}_{.2}$, in R_p form one isotonic region. In a bounded space there are only $p!$

isotonic regions and they correspond to the $p!$ possible shapes: All the points with the same shape are in the same subspace. Some points are located along the boundary hyperplanes. Since our interest is restricted to weak order, they can be considered members of more than one region. For example, the points along the line $\mathbf{X}_{.1} = \mathbf{X}_{.2} = \ldots = \mathbf{X}_{.p}$ are members of all the isotonic regions. All the isotonic spaces and all the distances within a bounded space are also bounded. Under these conditions it should be possible to calculate maximal and minimal distances within the space and within each isotonic subspace.

Before going into these calculations, which are the core of this paper and are necessary for the computation of the $B_{k(ij)}$ statistics, one general theorem can be stated: Under all restrictions the distance between two points will be maximal if and only if one (or both) is located at one of the vertices of the bounded space. (A vertex is a point whose coordinates are either members of the **L** vector or the **H** vector). The same theorem holds within each subspace with the restriction that the vertex should be a member of the subspace. The proof of the theorem is simple and intuitive and will not be presented here.

The Dissimilarity Indices

In this section the four dissimilarity indices will be introduced and the nature of the calculations involved will be demonstrated by a numerical example. A four-dimensional space example is sufficiently simple for hand calculations, yet sufficiently complex to demonstrate the variety of problems involved in a "profile analysis." Table 1 presents the location of four points—*a,b,c,* and *d*—in a four-dimensional space and also the minimal and maximal values of each of the dimensions of the space. The Euclidean distances between the points, calculated according to Equation 9, are presented in the first column of Table 2. Next, these distances will be normalized according to the different normalizing functions to yield the four dissimilarity indices.

Symmetric Unconditional Dissimilarity Index

$$B_{1(ij)} = D_{ij}/\max(D_{xy}) \qquad [11]$$

$$\max(D_{xy}) = \left[\sum_{k=1}^{p} (H_k - L_k)^2\right]^{\frac{1}{2}} = D(\underline{H},\underline{L}) \qquad [12]$$

This index compares the distance between a pair of points, $\mathbf{X}_i$ and $\mathbf{X}_j$, to the maximal distance between any two points in the space. This maximal value is calculated by the distance between two ex-

Table 1

A Numerical Example in a Four Dimensional Space

Dimension	Min.	Max.	a	b	c	d
$X_{.1}$	0	5	3	4	2	4
$X_{.2}$	0	6	2	3	5	2
$X_{.3}$	0	8	6	7	1	7
$X_{.4}$	0	7	1	2	6	3

Table 2
Values of the Four Statistics for
all Pairs of Profiles in the Examples

(i, j)	$D_{(ij)}$	$B_{1(ij)}$	$B_{2(ij)}$	$B_{3(ij)}$	$B_{4(ij)}$
a, b	2	.1516	.1696	.2030	.2828
a, c	7.745	.5872	.5872	.7864	.6556
a, d	2.449	.1850	.2077	.2487	.2737
b, a	2	.1516	.1696	.2010	.2264
b, c	7.745	.5872	.5872	.7784	.6844
b, d	1.414	.1070	.1199	.1421	.0870
c, a	7.745	.5872	.5872	.7100	.5339
c, b	7.745	.5872	.5872	.7100	.5339
c, d	7.615	.5773	.5773	.6980	.5145
d, a	2.449	.1850	.2077	.2487	.2144
d, b	1.414	.1070	.1199	.1435	.0870
d, c	7.615	.5773	.5773	.7732	.6343

treme points in the space: one with maximal coordinates along all p scales (**H**) and the second with minimal values along all scales (**L**). $B_{1(ij)}$ is closely related to D—in fact it is just a rescaling of D to a value bounded by 0 and 1. $B_{1(ij)}$ is symmetric; and since all the distances are normalized by the same factor regardless of their shape, the index is unconditional. B_1 reflects dissimilarities between the level, scatter, and shape of the two profiles (Skinner, 1978).

In the example all the distances are normalized by $D(\mathbf{H},\mathbf{L})=13.19$, and the results are presented in the second column of Table 2.

Symmetric Dissimilarity Index Conditional Upon the Shape of Two Profiles

$$B_{2(ij)} = D_{ij}/\max(D_{xy} \,|\, x \sim i; y \sim j) \qquad [13]$$

The notation ($x \sim i$) means "x has the same shape as i." This index compares the distance between two points, $\mathbf{X}_i$ and $\mathbf{X}_j$, from two different isotonic spaces labeled i and j, to the maximal distance that can be obtained between any two points taken from these particular subspaces. The maximal distance in the denominator of Equation 13 is the maximal distance between all pairs of vertices ($\mathbf{M}_i$, $\mathbf{N}_j$), where the first member of the pair is a member of subspace i and the second is a member of subspace j. The following algorithm can be used in order to compute this maximal distance:

1. Permute all the dimensions of the space so that they correspond to the ordering in subspace i. Define two new vectors $\mathbf{H}'$ and $\mathbf{L}'$ such that

$$H'_1 = H_1 \qquad [14a]$$

$$L'_p = L_p \qquad [14b]$$

$$H'_k = \min(H_k, H'_{k-1}) \qquad (k=2\ldots p) \qquad [14c]$$

$$L'_k = \max(L_k, L'_{k+1}) \qquad (k=(p-1)\ldots 1) \qquad [14d]$$

Calculate the values of the $(p+1)$ vertices of subspace i ($\mathbf{M}_i$) by combining the first r members of $\mathbf{H}'$ and the remaining $(p-r)$ members of $\mathbf{L}'$ ($r=0\ldots p$).

2. Permute all the dimensions of the space so that they correspond to the ordering in subspace j. Define $\mathbf{H}''$ and $\mathbf{L}''$ such that

$$H''_1 = H_1 \qquad [15a]$$

$$L''_p = L_p \qquad [15b]$$

$$H''_k = \min(H_k, H''_{k-1}) \qquad (k=2\ldots p) \qquad [15c]$$

$$L''_k = \max(L_k, L''_{k-1}) \qquad (k=(p-1)\ldots 1) \qquad [15d]$$

Calculate the values of the $(p+1)$ vertices of subspace j ($\mathbf{N}_j$) by combining the first r members of $\mathbf{H}''$ and the remaining $(p-r)$ members of $\mathbf{L}''$ ($r=0\ldots p$). The vectors obtained by these operations restrict the minimal values to be at least equal to the minimum of the lowest scale within the subspace of interest and the maximal values to be at most equal to the maximum of the highest scale within the isotonic subregion.

3. Permute the dimensions of $\mathbf{M}_i$ and $\mathbf{N}_j$ to a common order (possibly the original one) and compute the maximal distance by

$$\max(D_{xy} \mid x\sim i; y\sim j) = \max[D(\underline{M}_i, \underline{N}_j)] \qquad [16]$$

Since all subspaces have at least one point in common (along $\mathbf{X}_{.1}=\mathbf{X}_{.2}=\ldots=\mathbf{X}_{.p}$), the minimal distance is always 0. For each pair of points in any two subspaces the distance is normalized by the same factor. Therefore, B_2 is a symmetric index. Since the nature of the normalizing function depends on the particular pair of subspaces (shapes) involved, the index is conditional upon the two shapes. By imposing the shape restrictions, a statistic is obtained that measures dissimilarities only in the level and scatter of the two profiles.

The values of $B_{2(ij)}$ for the four points in the example are displayed in the third column of Table 2. Note that for several pairs in this example $\mathrm{B}_{1(ij)}=\mathrm{B}_{2(ij)}$, because the location of the points is such that the conditional maximal distance coincides with the unconditional maximal distance. However, usually $B_{2(ij)} > B_{1(ij)}$.

Asymmetric Unconditional Dissimilarity Index

$$B_{3(ij)} = D_{ij}/\max(D_{ix}) \qquad [17]$$

This index compares the distance between two points, $\mathbf{X}_i$ and $\mathbf{X}_j$, to the maximal distance between the first (call it the standard) and any other point in space. This maximal distance is computed by

$$\max(D_{ix}) = \{\sum_{k=1}^{p} \max[(x_{ik}-H_k),(x_{ik}-L_k)]^2\}^{\frac{1}{2}} \qquad [18]$$

This statistic measures the distance between $\mathbf{X}_i$ and the most distant vertex of the space. It is easy to verify that $B_{3(ij)} \neq B_{3(ji)}$, since B_3 depends on the location of $\mathbf{X}_i$ which is different from the location of $\mathbf{X}_j$. However, the index is defined as unconditional because the same normalizing function is used for all the distances from $\mathbf{X}_i$, regardless of the shape of $\mathbf{X}_j$.

Although the notion of asymmetric similarity might seem counterintuitive, it has some empirical support (Tversky, 1977) and is important in the areas in which profile analyses are usually conducted. In personnel testing the main motivation behind such analyses is "to determine how well an applicant's profile of test scores matches some standard profile, such as a representative pattern of scores in a group of superior employees" (Guion, 1965, p. 174). Similarly, in clinical applications a usual situation is one "where a standard or reference profile has been established and it is desired to match a given 'unknown' or referred profile against it" (Mosel & Roberts, 1954, p. 61). In both situations the asymmetry between the standard on one hand, and any particular individual on the other, is obvious. The standard profile is based on a large number of cases selected according to some prespecified criteria in order to represent a well-defined group or "type."

Methodologically, the coordinates of such a profile are more accurate and more reliable than those of any other "unknown" individual. Psychologically, the location of this profile has a special meaning which can be related to theoretically meaningful and empirically established concepts. The similarity index should reflect this asymmetry. It should be an extension of the form "*a* is like *b*." Such a statement is directional; it has a subject, *a*, and a referent, *b*, and it is not equivalent in general to the converse statement "*b* is like *a*" (Tversky, 1977, p. 328). In the geometrical model in this paper all the special characteristics of a standard profile are represented by its location within the bounded space which is reflected in the normalizing function of $B_{3(ij)}$.

The values of $B_{3(ij)}$ for the data in the numerical example are presented in the fourth column of Table 2. The asymmetric characteristic of B_3 is clearly demonstrated. However, since max (D_{ax}) = max (D_{dx}), $B_{3(ad)} = B_{3(da)}$, but this is only a special case.

Asymmetric Dissimilarity Index Conditional Upon the Shape of One Profile

This index compares the distance between two points, $\mathbf{X}_i$ and $\mathbf{X}_j$, to the maximal distance between the first and all the points in the subspace with ordering *j*, from which the second profile is taken. The magnitude of this index depends on the shape of $\mathbf{X}_j$ and therefore is stated to be conditional upon *j*. Since $\mathbf{X}_i$ and $\mathbf{X}_j$ have different locations, the index is asymmetric. The computation of the maximal distance for this case is very similar to the procedure used in the computation of $B_{2(ij)}$:

$$B_{4(ij)} = \frac{D_{ij} - \min(D_{ix}|x\sim j)}{[\max(D_{ix}|x\sim j) - \min(D_{ix}|x\sim j)]} \qquad [19]$$

1. Without affecting the distances, the dimensions are permuted to the order prescribed by the shape of $\mathbf{X}_j$.
2. Two new vectors ($\mathbf{H}'$ and $\mathbf{L}'$), as described in Equations 14a to 14d, are calculated.

3. From these two vectors the $(p+1)$ vertices of the subspace j are obtained by combining the first r members of $\mathbf{H}'$ and the remaining $(p-r)$ members of $\mathbf{L}'$ $(r=0..p)$. The maximal distance is obtained by comparing the distances from $\mathbf{X}_i$ to each of these $(p+1)$ vertices:

$$\max(D_{ix} | x \sim j) = \max[D(\underline{x}_i, \underline{N}_j)] \qquad [20]$$

In all the other statistics $\min(D_{ij})$ was always 0. In this case, this is generally not true (unless $\mathbf{X}_i$ and $\mathbf{X}_j$ are in the same subspace). The minimal distance is the length of a perpendicular line from $\mathbf{X}_i$ to the boundary hyperplane separating regions i and j. The characteristics of this hyperplane can be found by transforming contradictory inequalities in the two orderings ($\mathbf{X}_{.1}<\mathbf{X}_{.m}$ for $\mathbf{X}_i$ but $\mathbf{X}_{.m}<\mathbf{X}_{.1}$ for $\mathbf{X}_j$), to equalities ($\mathbf{X}_{.1}=\mathbf{X}_{.m}$). The desired hyperplane is the one which satisfies both orderings following a minimal number of such changes. The coordinates of the desired point (and its distance from $\mathbf{X}_i$) can be obtained by replacing those coordinates of $\mathbf{X}_i$ whose order was changed by a constant satisfying the equalities derived and by leaving the other coordinates unchanged. It is possible to show that a constant satisfying the orthogonality requirement is the average of the coordinates of $\mathbf{X}_i$ whose order was altered.

The new index measures the dissimilarity in the level and scatter of a profile with a prespecified shape and a fixed "standard." An interesting special case is the one in which the two profiles have identical shapes. This situation can be conceptualized as a two-stage process, similar to the one suggested by Skinner (1978). First, only profiles satisfying some shape restriction are identified, and in the second stage their similarity to the standard is evaluated in terms of the differences in level and shape alone. The computations in this case are easier to handle because the minimal distance between $\mathbf{X}_i$ and $\mathbf{X}_j$ is 0.

The values of $B_{4(ij)}$ for the data in the example are presented in the last column of Table 2. Again, as in the case of B_3, the asymmetry of the dissimilarity indices is obvious and is violated only by the special case, $B_{4(bd)}=B_{4(db)}$, in which the minimal and maximal distances from b and d to the boundary hyperplane are equal. It is important to realize that $B_{4(ij)}=0$ does not imply (as it does for B_1, B_2, and B_3) that the two points coincide, but rather that $\mathbf{X}_j$ is located on the boundary hyperplane and that of all the points on this hyperplane it is the closest to $\mathbf{X}_i$.

Interpretation of the Dissimilarity Indices

It is convenient to discuss separately two ways of interpreting the $B_{k(ij)}$ statistics: interpretations based on their actual values and on their distributions.

By the nature of their scale the $B_{k(ij)}$ statistics introduce meaningful reference points. Since they are calculated as ratios of distances, the indices are on a ratio scale and therefore allow stronger inferences. These two features, combined together, offer a more meaningful interpretation of the profile space. Usually, the interpretation of profiles ignores the fact that the space is bounded and the number of values each scale can take is finite and countable. These features are explicitly included in the normalizing functions and add new information. The conditional index, $B_{2(ij)}$, is restricted only to a prespecified region of interest in the space. Its computation and interpretation is independent of similarities and distances outside this subspace. Note that if for some reason one is interested only in profiles with two prespecified shapes, a correlational index is useless (all pairs of points will correlate equally) and an unconditional index of similarity (like $B_{1(ij)}$) takes care of only part of the problem, since it depends on other distances outside the domain of interest. A normalization with respect to the restricted regions of interest allows for finer differentiations related exclusively to the domain of interest and independent of all other points in the space.

$B_{3(ij)}$ and $B_{4(ij)}$ are special cases of $B_{1(ij)}$ and $B_{2(ij)}$, respectively, in which one of the points is fixed. For all practical purposes and without any loss of generality, the fixed point can be referred to as a "standard." It may be a "typical" psychotic in a MMPI, a "typical" engineer in Strong's SVIB, or an "ideal self" in a self-reporting questionnaire. Because the coordinates of the standard profile are fixed, and because its location has a specific psychological meaning, the range and distribution of similarities to it (in the whole space as well as in any subspace) can differ drastically from the range and distribution of unrestricted similarities (as reflected in $B_{1(ij)}$ or $B_{2(ij)}$). The asymmetric indices capture these particular characteristics of the standard point in their normalizing functions and allow inferences to be made about the relative similarity of points to the standard or about the relative similarity of points within a subspace to the standard. No other measure of profile similarity allows a researcher to make similar judgments.

The isomorphism between distance measures adjusted for level and scatter on one hand and the product-moment correlation on the other was already mentioned. The implication of this relation is that any $B_{k(ij)}$ calculated on standardized scores can also be expressed as a function of the correlations between the profiles. The relation can be expressed as

$$B_{k(ij)} = \left[\frac{1 - R_{ij}}{1 - R_{min}}\right] = \left[1 - \frac{R_{ij} - R_{min}}{1 - R_{min}}\right] \qquad [21]$$

The numerator is a function of the correlation between $\mathbf{X}_i$ and $\mathbf{X}_j$ and the denominator is a function of the minimal correlation that can be obtained under the restrictions of the model. Under these conditions the $B_{k(ij)}$ become pure shape similarity indices and can be interpreted as normalized differences between the observed correlations and their lower bounds. Note that in this context, shape is reflected by the regular correlation coefficient. This correlation reflects not only dissimilarity between the rank order of the scales of the two profiles but also some characteristics of the underlying distribution (Stuart, 1954). If the "standardized" scores are replaced by their ranks, the values of $B_{k(ij)}$ are related to the Spearman rank correlation and can be considered a measure of "pure" shape dissimilarity.

It is possible to calculate under very general and nonrestrictive assumptions the probability distribution functions of each of the $B_{k(ij)}$ statistics. Since the space of interest was already assumed to be bounded and to contain a finite number of points, only a small number of assumptions need to be added. For example, assume that along each dimension there is only a finite known number of values that the scale can take and that each of these values is equally likely. In the example here, these values are the integers between $\mathbf{L}_k$ and $\mathbf{H}_2$ (k=1. . .p). Since the dimensions were assumed to be independent, each point can be located in any of *NP* ways (with equal probability) where:

$$NP = \prod_{k=1}^{p} (H_k - L_k + 1) \qquad [22]$$

If $\mathbf{X}_i$ and $\mathbf{X}_j$ are a pair of arbitrary profiles sampled at random from the space and if they are independent, then they can be located in the space in *NP* × *NP* different ways. Each of these locations is related with a $B_{1(ij)}$ value in a unique way. The probability function of $B_{1(ij)}$, for any given space, can be calculated and is a function of the number of dimensions *(p)* and the minimal and maximal values (**L** and **H**).

For the distribution of $B_{2(ij)}$ all the points with shape like $\mathbf{X}_i$ (say there are *NPI* such points) and all the points with shape like $\mathbf{X}_j$ (there are *NPJ*) need to be singled out. Under the independence assumption there are *NPI* × *NPJ* possible pairs of points in the two subspaces, each of them related to a $B_{2(ij)}$ value. The same reasoning can be applied in order to derive the probability functions for $B_{3(ij)}$

and $B_{4(ij)}$. Since in these cases, point $\mathbf{X}_i$ is fixed, there are $(1 \times NP)$ $B_{3(ij)}$'s and $(1 \times NPJ)$ $B_{4(ij)}$'s, depending on the specified $\mathbf{X}_j$. On the basis of these distributions, the probability of obtaining the results by chance can be calculated; and according to the level of confidence considered necessary, this hypothesis can be rejected or accepted.

The $B_{k(ij)}$ statistics should be interpreted as proportions. Each $B_{k(ij)}$ reflects the proportion of dissimilarity observed, with respect to the maximal and minimal possible dissimilarities under the restrictions of the index. Since the normalizations are conditional upon a given domain of interest (defined by shape or location of the fixed point), any comparison of dissimilarity indices should be limited to indices calculated under identical restrictions.

With the exception of $B_{1(ij)}$, the indices do not necessarily form Gramian matrices and can not be factor analyzed. However, all four statistics can be analyzed by the ALSCAL multidimensional scaling model which can handle asymmetric and conditional similarity measures (Young & Lewyckyj, 1979).

Some Possible Generalizations

All four indices are based on different normalizations of D, the Euclidean distance. As long as interpretation of D is thus restricted, the use of these statistics is conditional upon the assumption of orthogonality of the scales. However, D and its different normalizations can be computed and used as indices of dissimilarity without using the Euclidean model: D can be regarded as a function of the differences between the scores on the two profiles along all p scales. In particular, it is a function in which all these p differences are equally weighted, regardless of the nature of the correlations between the scales. If this interpretation of D (suggested by E. M. Cramer) is accepted, there is no need to develop special solutions for the nonorthogonal case.

If the Euclidean model is retained, the unconditional indices ($B_{1(ij)}$ and $B_{3(ij)}$) can be directly generalized to any pattern of correlations among the variables. This can be best understood from the fact that a principal components decomposition, which transforms the space into a new p dimensional space with uncorrelated axes, leaves the distances among the points unchanged. The conditional indices are not directly generalizable, since the notion of "shape" does not have as clear a meaning in this context as in the orthogonal case: The shape of the profiles is completely distorted when they are represented in a new orthogonal space following a principal components or any other orthogonalization technique. However, if the scales of the profile have a known factorial structure, the four indices can be calculated on the estimated factor scores. The calculations and interpretations are identical, except that they refer to a new set of underlying factors.

Other possible generalizations consist of the development of parallel measures of dissimilarity based on the squared distance between any pair of points or on alternative distance functions (e.g., "the city-block" distance) and of the development of distribution functions for the $B_{k(ij)}$ statistics based on other underlying distributions. Of particular interest seems to be the normal distribution, which was already suggested by Cattell (1949).

References

Cattell, R. B. R_p and other coefficients of pattern similarity. *Psychometrika*, 1949, *14*, 279-298.

Cohen, J. R_c: A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin*, 1969, *71*, 281-284.

Cronbach, L. J., & Gleser, G. C. Assessing similarity between profiles. *Psychological Bulletin*, 1953, *50*, 456-73.

Green, P. E., & Carroll, D. J. *Mathematical tools for applied multivariate analysis*. New York: Academic Press, 1976.

Guion, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.

Mosel, J. N., & Roberts, J. B. The comparability of measures of profile similarity: An empirical study. *Journal of Consulting Psychology*, 1954, *18*, 61-66.

Nunnally, J. The analysis of profile data. *Psychological Bulletin*, 1962, *59*, 311-319.

Nunnally, J. *Psychometric theory*. New York: McGraw-Hill, 1967.

Overall, J. E. Note on multivariate methods for profile analysis. *Psychological Bulletin*, 1964, *61*, 195-198.

Overall, J. E., & Klett, C. J. *Applied multivariate analysis*. New York: McGraw-Hill, 1972.

Skinner, H. A. Differentiating the contribution of elevation, scatter, and shape in profile similarity. *Educational and Psychological Measurement*, 1978, *38*, 297-308.

Stuart, A. The correlation between variate-values and ranks in samples from a continuous distribution. *British Journal of Statistical Psychology*, 1954, *7*, 37-44.

Tversky, A. Features of similarity. *Psychological Review*, 1977, *84*, 327-352.

Webster, H. A note on profile similarity. *Psychological Bulletin*, 1952, *49*, 538-540.

Young, F. W., & Lewyckyj, R. *ALSCAL-4 User's Guide*. Carrboro, NC: Data Analysis and Theory Associates, 1979.

Acknowledgments

The author thanks M. I. Appelbaum, E. M. Cramer, A. Rapoport, T. S. Wallsten, and a reviewer for their valuable comments on earlier drafts of this paper.

A FORTRAN program that calculates the values of the four $B_{k(ij)}$ statistics, as well as their probability functions, is available upon request from the author.

Author's Address

Send requests for reprints or further information to David V. Budescu, Psychometric Laboratory, Davie Hall 013-A, University of North Carolina, Chapel Hill, NC 27514.