MULTIDIMENSIONAL SCALING: I. THEORY AND METHOD*

WARREN S. TORGERSON

SOCIAL SCIENCE RESEARCH COUNCIL

Multidimensional scaling can be considered as involving three basic steps. In the first step, a scale of comparative distances between all pairs of stimuli is obtained. This scale is analogous to the scale of stimuli obtained in the traditional paired comparisons methods. In this scale, however, instead of locating each stimulus-object on a given continuum, the distances between each pair of stimuli are located on a distance continuum. As in paired comparisons, the procedures for obtaining a scale of comparative distances leave the true zero point undetermined. Hence, a comparative distance is not a distance in the usual sense of the term, but is a distance minus an unknown constant. The second step involves estimating this unknown constant. When the unknown constant is obtained, the comparative distances can be converted into absolute distances. In the third step, the dimensionality of the psychological space necessary to account for these absolute distances is determined, and the projections of stimuli on axes of this space are obtained. A set of analytical procedures was developed for each of the three steps given above, including a least-squares solution for obtaining comparative distances by the complete method of triads, two practical methods for estimating the additive constant, and an extension of Young and Householder's Euclidean model to include procedures for obtaining the projections of stimuli on axes from fallible absolute distances.

Introduction

The traditional methods of psychophysical scaling presuppose knowledge of the dimensions of the area being investigated. The methods require judgments along a particular defined dimension, i.e., A is brighter, twice as loud, more conservative, or heavier than B. The observer, of course, must know what the experimenter means by brightness, loudness, etc. In many stimulus domains, however, the dimensions themselves, or even the number of relevant dimensions, are not known. What might appear intuitively to be a single dimension may in fact be a complex of several. Some of the intuitively given dimensions may not be necessary—it may be that they can be accounted for by linear combinations of others. Other dimensions of importance may be completely overlooked. In such areas the traditional approach is inadequate.

Richardson, in 1938 (3; see also Gulliksen, 1) proposed a model for multidimensional scaling that would appear to be applicable to a number of

these more complex areas. This model differs from the traditional scaling methods in two important respects. First, it does not require judgments along a given dimension, but utilizes, instead, judgments of similarity between the stimuli. Second, the dimensionality, as well as the scale values, of the stimuli is determined from the data themselves.

Multidimensional scaling may perhaps best be considered as involving three basic steps. In the first step, a scale of comparative distances between all pairs of stimuli is obtained. The second step involves estimating an additive constant and using this estimate to convert the comparative distances into absolute distances. In the third step, the dimensionality of the psychological space necessary to account for these absolute distances is determined, and the projections of the stimuli on axes of this space are obtained.
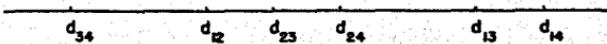
The Scale of Comparative Distances

The scale of comparative distances obtained in the multidimensional methods is analogous to the one-dimensional scale of stimulus-objects obtained in the traditional paired comparison type methods.

In the one-dimensional methods, the obtained scale locates the stimulus-objects with respect to one another on the given continuum. For example, given four stimulus-objects designated S_1, S_2, S_3, and S_4, the one-dimensional procedure might yield the following scale:

$$\overline{\quad\quad S_1 \quad\quad\quad\quad\quad S_2 \quad\quad\quad\quad\quad\quad\quad\quad S_3 \quad S_4 \quad\quad}$$

In this scale, the locations of the stimuli relative to one another only are determined from the data. The zero point of the scale is arbitrary. While the usual procedure is to locate the zero point so as to coincide with the stimulus having the lowest scale value, any other finite location on the continuum would serve equally well.

In the analogous scale of comparative distances obtained in the multi-dimensional procedures, the element, instead of being a stimulus-object, is a distance between two stimuli. Thus, given the same four stimulus-objects, the scale of comparative distances locates, with respect to one another on a distance continuum, the six inter-stimulus distances, d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, and d_{34}:

$$\overline{\quad\quad d_{34} \quad\quad\quad d_{12} \quad d_{23} \quad d_{24} \quad\quad\quad d_{13} \quad d_{14} \quad\quad}$$

The locations of the inter-stimulus distances relative to one another only are determined from the data. The zero point is again arbitrarily selected. It is important to note, however, that a comparative distance is not a "distance" in the usual sense of the term, but is a distance minus an unknown constant. In order to obtain absolute distances between stimuli, it is necessary to

estimate this constant. This is equivalent to estimating the true zero point of the scale of comparative distances. Thus, a comparative distance h_{jk} plus an unknown additive constant C gives the corresponding absolute distance d_{jk} :

$$h_{jk} + C = d_{jk} .$$

The Additive Constant for Converting Comparative Distances into Absolute Distances

In estimating the additive constant, it is assumed that that value which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality is the value wanted. Consider, for example, five points having the following comparative interpoint distances h_{jk} $(j, k = 1, 2, \ldots , 5; j \neq k)$:

$$h_{12} = 1, \quad h_{14} = \quad 1, \quad h_{23} = 1, \quad h_{25} = 0, \quad h_{35} = -1,$$
$$h_{13} = 2, \quad h_{15} = -1, \quad h_{24} = 4, \quad h_{34} = 1, \quad h_{45} = \quad 0.$$

With these comparative distances the value of the additive constant which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality is 4. If we add 4 to each of the comparative distances to convert them into absolute distances we obtain

$$d_{12} = 5, \quad d_{14} = 5, \quad d_{25} = 4, \quad d_{15} = 3, \quad d_{35} = 3,$$
$$d_{13} = 6, \quad d_{24} = 8, \quad d_{34} = 5, \quad d_{23} = 5, \quad d_{45} = 4.$$

The five stimuli can be plotted in a two-dimensional space:
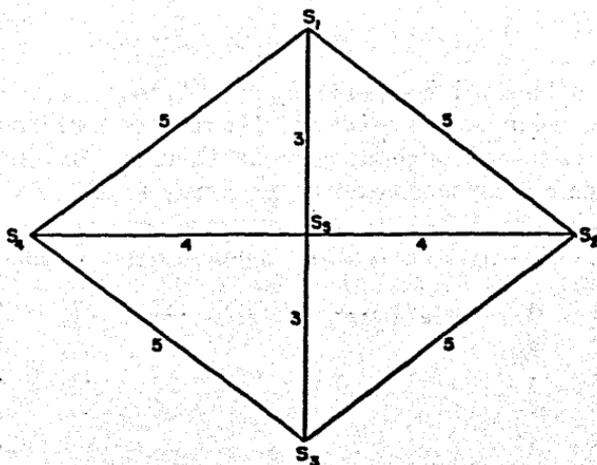
Note that for any smaller value of the additive constant the points do not exist in a real Euclidean space. For example, if 1, 2, or 3 is added, then $d_{45} + d_{25} < d_{24}$, an impossible relationship in real Euclidean space. Also, for any larger value of the additive constant, the points lie in a real space of dimensionality greater than two.

Determination of the Dimensionality of the Psychological Space and the Projections of the Stimuli on Axes of the Space from the Absolute Distances Between the Stimuli

Young and Householder (5) have given a method for determining whether a set of absolute interpoint distances can be considered to be the distances between points lying in a real Euclidean space. They also have given, provided that the distances can so be considered, methods for determining the dimensionality of the space, and the projections of the points on a set of orthogonal axes of the space. Their theorems involve two basic matrices, B_i and F.

If we let

i, j, and k be alternate subscripts for n points $(i, j, k = 1, 2, \ldots, n)$ and d_{ij} , d_{ik} , and d_{jk} be the distances between the points, then B_i is an $(n - 1) \times (n - 1)$ symmetric matrix with elements

$$b_{jk} = \tfrac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2). \tag{1}$$

The element b_{jk} may be considered to be the scalar product of vectors from point i to points j and k. This follows directly from the cosine law. That is, given the three points i, j, and k,

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2d_{ij}d_{ik} \cos \theta_{jik} ,$$

which rearranged becomes

$$d_{ij}d_{ik} \cos \theta_{jik} = \tfrac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2). \tag{3}$$

From Equations 1 and 3, it is seen that $b_{jk} = d_{ij}d_{ik} \cos \theta_{jik}$, the scalar product of vectors from point i to points j and k. Matrix B_i is thus a matrix of scalar products of vectors with origin at point i. There are, of course, n possible B_i matrices, since i may assume any value from 1 to n.

Matrix F is an $(n + 1) \times (n + 1)$ symmetric matrix of squares of interpoint distances bordered by a row and column of ones as follows:

$$
F = \begin{matrix}
0 & d_{12}^2 & \cdots & d_{1k}^2 & \cdots & d_{1n}^2 & 1 \\
d_{21}^2 & 0 & & & & d_{2n}^2 & 1 \\
\cdot & & & & & \cdot & \cdot \\
\cdot & & & & & \cdot & \cdot \\
\cdot & & & & & \cdot & \cdot \\
d_{j1}^2 & & & & & d_{jn}^2 & 1 \\
\cdot & & & & & \cdot & \cdot \\
\cdot & & & & & \cdot & \cdot \\
\cdot & & & & & \cdot & \cdot \\
d_{n1}^2 & d_{n2}^2 & \cdots & d_{nk}^2 & \cdots & 0 & 1 \\
1 & 1 & \cdots & 1 & \cdots & 1 & 0
\end{matrix}
\tag{4}
$$

Young and Householder have shown that:

1. If any matrix B_i is positive semidefinite, the distances may be considered to be the distances between points lying in a real Euclidean space.

2. The rank of any positive semidefinite matrix B_i is equal to the dimensionality of the set of points.

3. The rank of matrix F is two greater than the dimensionality of the set of points.

4. Any positive semidefinite matrix B_i may be factored to obtain a matrix A_i such that

$$B_i = A_i A'_i . \tag{5}$$

If the rank of B_i is r, where $r \leq (n-1)$, then matrix A_i is an $(n-1) \times r$ matrix of projections of points on r orthogonal axes with origin at the ith point of the r-dimensional, real Euclidean space.

It is interesting to note that except for Richardson's original experiment (only an abstract of which has been published) only one person, Klingberg, (2) has used the model. It may well be that one of the reasons for the lack of experimental investigation in this area is that no clear statement of analytical procedure has been published. The problem of precisely how to proceed in obtaining comparative distances from proportions of judgments has not been adequately answered for either Richardson's method of triadic combinations or Klingberg's method of multidimensional rank order. While the analogy between the logic of paired comparisons and both of these methods is clear, the procedures cannot be directly applied in obtaining an efficient estimate. The least-squares solution for paired comparisons scales cannot be used because the analogous proportion matrix contains a rather large number of vacant cells—neither multidimensional method obtains judgments of the differences in distance between *all* possible pairs of distances, but only between pairs having one stimulus in common. Furthermore, in reducing the matrix of distance-differences between pairs to a scale of comparative distances, one is almost overwhelmed by the great number of possible modes of attack—each likely to give a somewhat different answer due to error in the observed data.

The problem of how to obtain a best estimate of the unknown additive constant has not been answered. The method used by Klingberg is quite tedious (it involved obtaining two tenth-order polynomials from the fifth-order minors and then solving for the unknowns) and does not insure that the answer obtained is a best estimate, or that it even approximates the value desired.

Similarly, while Young and Householder give adequate procedures for obtaining projections of points on axes from distances when the data are infallible, a number of difficulties arise when fallible data are employed.

The purpose of the present paper is to present a set of analytical procedures for multidimensional scaling, including, as far as possible, routine procedures for obtaining comparative distances, for estimating the additive constant, and for obtaining projections of stimuli on axes when fallible absolute distances are given. We shall first consider the complete method of triads for obtaining comparative distances between the stimuli. Following this, the problem of obtaining projections of stimuli on axes from fallible absolute distances will be discussed. Finally, we shall consider various methods for estimating the unknown additive constant.

The Complete Method of Triads for Obtaining Comparative Distances Between Stimuli

The stimuli are presented to the subject in triads. The judgment required of the subject is of the form: "Stimulus k is more similar to stimulus j than to stimulus i." With n stimuli, there are $n(n-1)(n-2)/6$ triads. In each triad, each stimulus is compared with each other pair, making a total of $n(n-1)(n-2)/2$ judgments for each subject. From these judgments we obtain the proportion of times any stimulus k is judged more similar to stimulus j than to i.

These proportions can be arranged in the n matrices $_kP_{ij}$ where k, i, and j are alternate subscripts for the stimuli. k gives the number of the matrix, i is a row index, and j is a column index. The element $_kp_{ij}$ is the proportion of times stimulus k is judged closer to stimulus j than to i. The matrices $_kP_{ij}$ have vacant cells in the principal diagonal, and in the kth row and column.* The matrices are such that the sum of symmetric elements is unity—e.g., $_kp_{ji} + _kp_{ij} = 1$. For example, given four stimuli, 1, 2, 3, and 4, there are four $_kP_{ij}$ matrices. The second matrix ($k = 2$) is illustrated below:

$$_2P_{ij}$$

	1	2	3	4
1	.	.	$_2p_{13}$	$_2p_{14}$
2	.	.	.	.
3	$_2p_{31}$	.	.	$_2p_{34}$
4	$_2p_{41}$	.	$_2p_{43}$	.

The first problem is to transform the proportions $_kp_{ij}$ into differences in distances $_kx_{ij}$. We shall assume that the proportion of times stimulus k is judged closer to stimulus j than to i is a function of the difference in the

---

*It might be noted that the elements in the kth row and column *could* be obtained experimentally. However, since the method would ordinarily be used in connection with supraliminal distances, the experimentally determined proportions would be either .00 or 1.00. As in paired comparisons, proportions of .00 and 1.00 cannot be utilized.

distances, $d_{ki} - d_{kj} = {}_k x_{ij}$, the function being

$$_k p_{ij} = \int_{-\infty}^{kx_{ij}} \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}x^2} \, dx. \tag{6}$$

$_k x_{ij}$ is thus $\sqrt{2}$ times the deviate of the unit normal curve measured in σ units from the mean. This is analogous to Thurstone's Case V in paired comparisons (4) and is the same assumption used previously by Richardson (3).

Making this transformation we obtain the n matrices $_k X_{ij}$. These matrices are skew symmetric ($_k x_{gh} + {}_k x_{hg} = 0$), have zero diagonal elements, and have vacant cells in the kth row and column.

We have $n(n-1)(n-2)/2$ independent observations of differences in distances $_k x_{ij}$ from which we wish to determine $n(n-1)/2$ comparative distances h_{ik} . Since $\frac{1}{2}n(n-1) - 1$ differences in distances are sufficient to determine a matrix of comparative distances, it is apparent that the data are considerably overdetermined. There are, of course, a large number of sets of $\frac{1}{2}n(n-1) - 1$ differences in distances which could be used. Also, there are many different ways of obtaining the comparative distances from each set. With fallible data, the matrices could be expected to differ somewhat from each other.

The first problem, then, is to find a best estimate, in a least-squares sense, of the matrix of comparative distances h_{ik} in terms of the available data.

The element $_k x_{ij} = d_{ik} - d_{kj} + {}_k e_{ij}$, where d_{ik} and d_{kj} are *absolute* distances between stimuli k and i, and k and j, respectively, and $_k e_{ij}$ is an error.* It would seem that we want that set of interpoint distances which minimizes the sum of squares of the errors $_k e_{ij}$. For a least-squares solution, then, we wish to select the distances to minimize the following function:

$$2F = \sum_k^n \sum_{\substack{j \\ j \neq k}}^n \sum_{\substack{i \\ i \neq j \\ i \neq k}}^n [_k x_{ij} - (d_{ik} - d_{kj})]^2. \tag{7}$$

If we define a set of matrices $_k E_{ij}$ with elements $(_k x_{ij} - d_{ik} + d_{kj})$, it is seen that $2F$ is equal to the sum of squares of elements of the matrices $_k E_{ij}$.

Let g and h correspond to two particular stimuli with $d_{gh} = d_{hg}$, the distance between them. The term d_{gh} (or d_{hg}) occurs only in the error matrices $_g E_{ij}$ and $_h E_{ij}$ as follows:

in $_g E_{ij}$ the hth column contains the elements $_g x_{ih} - d_{ig} + d_{gh}$,
   the hth row contains the elements $_g x_{hj} - d_{hg} + d_{gj}$;
in $_h E_{ij}$ the gth column contains the elements $_h x_{ig} - d_{ih} + d_{hg}$,
   the gth row contains the elements $_h x_{gj} - d_{gh} + d_{hj}$.

---

*$_k x_{ij}$ is also equal to the difference between the comparative distances, since the difference in absolute distances $d_{ik} - d_{kj}$ is identical with the difference in comparative distances $(d_{ik} - C) - (d_{kj} - C)$.

To minimize $2F$, we first take the derivative of F with respect to d_{gh}. We shall designate this derivative as F'. It is apparent that the derivatives of all terms of F except those containing the element d_{gh} vanish. Therefore,

$$F' = \sum_{i \neq g,h}^{n} ({}_g x_{ih} - d_{ig} + d_{gh}) - \sum_{j \neq g,h}^{n} ({}_g x_{hj} - d_{hg} + d_{gj})$$

$$+ \sum_{i \neq g,h}^{n} ({}_h x_{ig} - d_{ih} + d_{hg}) - \sum_{j \neq g,h}^{n} ({}_h x_{gj} - d_{gh} + d_{hj}). \qquad (8)$$

But, since matrices ${}_g E_{ij}$ and ${}_h E_{ij}$ are skew symmetric,

$$\sum_{i \neq g,h} ({}_g x_{ih} - d_{ig} + d_{gh}) = - \sum_{j \neq g,h} ({}_g x_{hj} - d_{hg} + d_{gj}), \qquad (9)$$

and

$$\sum_{i \neq g,h} ({}_h x_{ig} - d_{ih} + d_{hg}) = - \sum_{j \neq g,h} ({}_h x_{gj} - d_{gh} + d_{hj}). \qquad (10)$$

Therefore, we may write

$$F' = 2 \sum_{i \neq g,h} ({}_g x_{ih} - d_{ig} + d_{gh}) + 2 \sum_{i \neq g,h} ({}_h x_{ig} - d_{ih} + d_{hg}). \qquad (11)$$

Setting F' equal to zero, and summing over each term, we find

$$\sum_{i \neq g,h} {}_g x_{ih} - \sum_{i \neq g,h} d_{ig} + (n - 2)d_{gh} + \sum_{i \neq g,h} {}_h x_{ig} - \sum_{i \neq g,h} d_{ih}$$
$$+ (n - 2)d_{hg} = 0. \qquad (12)$$

Remembering that $d_{gg} = d_{hh} = 0$, and that the diagonals and kth row and column of all ${}_k X_{ij}$ matrices are vacant, we can write:

$$\sum_{i} {}_g x_{ih} - \sum_{i \neq h} d_{ig} + (n - 2)d_{gh} + \sum_{i} {}_h x_{ig} - \sum_{i \neq g} d_{ih}$$
$$+ (n - 2)d_{hg} = 0. \qquad (13)$$

Subtracting d_{gh} from $- \sum_{i \neq h} d_{ig}$ and from $- \sum_{i \neq g} d_{ih}$, adding d_{gh} to $(n - 2)d_{gh}$ and $(n - 2)d_{hg}$, and remembering that $d_{gh} = d_{hg}$, we have

$$\sum_{i} {}_g x_{ih} - \sum_{i} d_{ig} + (n - 1)d_{gh} + \sum_{i} {}_h x_{ig} - \sum_{i} d_{ih}$$
$$+ (n - 1)d_{hg} = 0. \qquad (14)$$

Summing over g, $g \neq h$, we have

$$\sum_{g \neq h} \sum_{i} {}_g x_{ih} - \sum_{g \neq h} \sum_{i} d_{ig} + (n - 1) \sum_{g \neq h} d_{gh} + \sum_{g \neq h} \sum_{i} {}_h x_{ig}$$
$$- (n - 1) \sum_{i} d_{ih} + (n - 1) \sum_{g \neq h} d_{hg} = 0. \qquad (15)$$

But

$$\sum_{\substack{\sigma \\ \sigma \neq h}} \sum_i {}_h x_{i\sigma} = 0,\tag{16}$$

and, since $d_{hh} = 0$,

$$\sum_i d_{ih} = \sum_{\substack{\sigma \\ \sigma \neq h}} d_{h\sigma} \; ;\tag{17}$$

therefore,

$$\sum_{\substack{\sigma \\ \sigma \neq h}} \sum_i {}_\sigma x_{ih} - \sum_{\substack{\sigma \\ \sigma \neq h}} \sum_i d_{i\sigma} + (n-1) \sum_\sigma d_{\sigma h} = 0.\tag{18}$$

Subtracting $\sum_i d_{ih}$ from $\sum_{\substack{\sigma \\ \sigma \neq h}} \sum_i d_{i\sigma}$ and adding $\sum_\sigma d_{\sigma h}$ to $(n-1) \sum_\sigma d_{\sigma h}$, we

see, from Equation (17), that

$$\sum_{\substack{\sigma \\ \sigma \neq h}} \sum_i {}_\sigma x_{ih} - \sum_\sigma \sum_i d_{i\sigma} + n \sum_\sigma d_{\sigma h} = 0.\tag{19}$$

Rearranging, dividing by $n(n-1)$, and remembering that cells ${}_\sigma x_{i\sigma}$ are vacant, we find

$$\frac{1}{n(n-1)} \sum_\sigma \sum_i {}_\sigma x_{ih} = \frac{1}{n(n-1)} \sum_\sigma \sum_i d_{i\sigma} - \frac{1}{(n-1)} \sum_\sigma d_{\sigma h} .\tag{20}$$

Also, if we divide Equation (14) by $2(n-1)$ and rearrange, we obtain

$$\frac{1}{2(n-1)} \left[\sum_i d_{i\sigma} + \sum_i d_{ih} \right] - d_{\sigma h} = \frac{1}{2(n-1)} \left[\sum_i {}_\sigma x_{ih} + \sum_i {}_h x_{i\sigma} \right].\tag{21}$$

It will be convenient to define the averages in Equations (20) and (21) as follows:

$$d_{\cdot h} \equiv \frac{1}{(n-1)} \sum_i d_{ih} = \frac{1}{(n-1)} \sum_\sigma d_{\sigma h} ,\tag{22}$$

$$d_{\cdot \sigma} \equiv \frac{1}{(n-1)} \sum_i d_{i\sigma} ,\tag{23}$$

$$d_{\cdot \cdot} \equiv \frac{1}{n(n-1)} \sum_i \sum_\sigma d_{i\sigma} ,\tag{24}$$

$${}_\sigma x_{\cdot h} \equiv \frac{1}{(n-1)} \sum_i {}_\sigma x_{ih} ,\tag{25}$$

$${}_h x_{\cdot \sigma} \equiv \frac{1}{(n-1)} \sum_i {}_h x_{i\sigma} ,\tag{26}$$

$${}_\sigma x_{\sigma h} \equiv \frac{1}{n(n-1)} \sum_i \sum_\sigma {}_\sigma x_{ih} .\tag{27}$$

After substitutions have been made for the appropriate terms, Equation (21) becomes,

$$\tfrac{1}{2}(d_{.g} + d_{.h}) - d_{gh} = \tfrac{1}{2}(_gx_{.h} + _hx_{.g}),\tag{28}$$

and Equation (20) becomes

$$_.x_{.h} = d_{..} - d_{.h},\tag{29}$$

and, when $h = g$,

$$_.x_{.g} = d_{..} - d_{.g}.$$

Substituting for $d_{.g}$ and $d_{.h}$ in Equation (28), we have

$$\tfrac{1}{2}(d_{..} - {}_.x_{.g} + d_{..} - {}_.x_{.h}) - \omega_{gh} = \tfrac{1}{2}(_gx_{.h} + _hx_{.g}),\tag{30}$$

which rearranged becomes

$$d_{..} - d_{gh} = \tfrac{1}{2}(_gx_{.h} + {}_.x_{.g} + _hx_{.g} + {}_.x_{.h}).\tag{31}$$

When $g = j$, $h = k$, the comparative distance $h_{jk} = d_{gh} - d_{..}$. Since the x-values are functions of the observed proportions (Equation (6)), Equation (31) gives the comparative distances as functions of the observed data. Equation (31), then, gives a rather straightforward method for obtaining the best estimate, in a least-squares sense, of the matrix of comparative distances.

Obtaining Projections of Stimuli on Axes from Fallible Absolute Distances

For a situation in which the data are not fallible and in which absolute distances are given, Young and Householder have shown (a) how to determine if the stimuli lie in a real Euclidean space, (b) if they do, how to determine the dimensionality of the set of points, and (c) how to obtain the projections of the points on an arbitrary orthogonal reference system. This reference system may then be rotated to the "most meaningful" dimensions, if criteria for such are available.

We saw that if matrix B_i (Equation 5) is positive semidefinite, the stimuli lie in real Euclidean space. The rank of B_i (or two less than the rank of matrix F) is then equal to the dimensionality of the set of stimuli. Matrix B_i can be factored to obtain projections of the stimuli on an arbitrary set of orthogonal axes.

Matrix B_i, however, is constructed by placing the origin arbitrarily at one of the stimuli. With errorless data, the results will be identical (except for the orientation of axes and location of the origin) for each of the n possible matrices $B_i(i = 1, 2, \ldots, n)$. With fallible data, however, each point is somewhat in error. Assuming a true rank considerably less than the number of points, each matrix B_i will yield different results. We would then have the problem of deciding which B_i matrix gives the best solution.

One solution to this problem would be to place the origin at the centroid of the stimuli. This procedure would give a unique solution and would tend to allow the errors in the individual points to cancel each other. An origin at the centroid would, on the average, be less in error than an origin at any arbitrary stimulus. The problem would seem to be to find a convenient method of obtaining a matrix B^* with origin at the centroid of the stimuli instead of at one of the stimulus points.

We shall use the following notation:

m = axes $(m = 1, 2, \ldots, r)$,
j, k = points $(j, k = 1, 2, \ldots, n)$,
i = point taken as the origin,
a_{im} = projection of point j on axis m, and
d_{ik} = distance between points j and k;

and take as given Equations (1) and (5):

$B_i = AA'$

$b_{ik} = \frac{1}{2} (d_{ij}^2 + d_{ik}^2 - d_{jk}^2)$, where point i is taken as the origin. From Equation (5) it is seen that

$$b_{ik} = \sum_{m}^{r} a_{im} a_{km} . \tag{32}$$

We shall, however, consider B_i to be an $n \times n$ matrix with the ith row and column composed of zero elements. In like manner A is $n \times r$ with the ith row composed of zero elements.

We wish to translate the axes from an origin at point i to an origin at the centroid of all n points.

Let $A^* = \| a_{im^*} \|$ be the desired matrix of projections of points j on axis m^* of the new coordinate system with origin at the centroid of the n points.
Then

$$a_{im^*} = a_{im} - c_m , \tag{33}$$

where

$$c_m = \frac{1}{n} \sum_{j=1}^{n} a_{im} = \text{the average projection of points on} \tag{34}$$

axis m = projection of centroid on axis m.

$$B^* = A^*A^{*'} = \| b_{ik}^* \|, \tag{35}$$

and

$$b_{ik}^* = \sum_{m}^{r} a_{im^*} a_{m^*k} . \tag{36}$$

Substituting, we have

$$b^*_{ik} = \sum_m^r (a_{im} - c_m)(a_{km} - c_m)$$

$$= \sum_m^r a_{im}a_{km} - \sum_m^r a_{im}c_m - \sum_m^r a_{km}c_m + \sum_m^r c_m c_m . \qquad (37)$$

From Equation (34) it is seen that

$$b^*_{ik} = \sum_m^r a_{im}a_{km} - \frac{1}{n}\sum_m^r a_{im} \sum_j^n a_{im} - \frac{1}{n}\sum_m^r a_{km} \sum_k^n a_{km}$$

$$+ \frac{1}{n^2}\sum_m^r \left[\sum_j^n a_{jm} \right]\left[\sum_k^n a_{km} \right]. \qquad (38)$$

But

$$\sum_i^n b_{ik} = \sum_m^r a_{km} \sum_i^n a_{im} \qquad (39)$$

and

$$\sum_k^n \sum_i^n b_{ik} = \sum_m^r \left[\sum_j^n a_{jm} \right]^2 . \qquad (40)$$

Substituting, we have

$$b^*_{ik} = b_{ik} - \frac{1}{n}\sum_i^n b_{ik} - \frac{1}{n}\sum_k^n b_{ik} + \frac{1}{n^2}\sum_i \sum_k b_{ik} . \qquad (41)$$

Equation (41) gives a routine method of translating a matrix B_i with origin at point i to an equivalent matrix B^* with an origin at the centroid of the points. It makes no difference, of course, which of the n matrices B_i is used in obtaining matrix B^*.

Matrix B^*, then, is the B-matrix we wish to factor to obtain projections of stimuli on axes.

Estimating $d_{..}$, the Unknown Additive Constant

The procedures of obtaining dimensionality and projections on axes discussed in the preceding section require absolute distances as given data. When the given data are comparative distances ($\hat{h}_{ik} = d_{..} - d_{ik}$)†, the unknown constant must be estimated to convert the comparative distances into absolute distances. We shall first consider the case where the data are not fallible, after which we shall discuss procedures for fallible data.

1. Estimating $d_{..}$ from errorless comparative distances

With errorless data, in order that the stimuli be considered as lying in a real Euclidean space of r dimensions, the B_i matrix must be positive semi-

†Comparative distances with signs reversed, actually ($h_{ik} = -\hat{h}_{ik}$).

definite and have a rank equal to r. This is equivalent to the statement that r latent roots of B_i must be positive and the remaining $(n - r)$ equal to zero.

The value of $d_{..}$ desired is the value which will permit the location of the stimuli in a real, Euclidean space of the smallest possible dimensionality. In terms of the matrix B^*, the value of $d_{..}$ desired is that value which results in the positive semidefinite B^* with the lowest rank. In terms of the latent roots, this becomes that value which results in a matrix B^* with the largest number of zero roots under the condition that the remaining non-zero roots are all positive.

This value can be determined, although it involves a tremendous amount of labor. The straightforward solution would be as follows:

1. Construct matrix B^* from the given data $(d_{ik} = d_{..} - \hat{h}_{ik})$.
2. Obtain the characteristic equation:

$$| B^* - \lambda I | = 0.$$

3. Set the last term equal to zero and solve for the real, positive values of $d_{..}$. This term will be an $(n - 1)$th degree polynomial in $d_{..}$. One of these values is the value desired.
4. Substitute each of the values for $d_{..}$ in the complete characteristic equation. Inspection of these equations shows which value of $d_{..}$ yields the largest number of zero roots.
5. The value which yields the largest number of zero roots with the remaining roots all positive is the value desired.

A "short-cut" procedure would be to evaluate the determinant of B^* directly. This determinant is the last term of the characteristic equation. One could then obtain the real, positive values of $d_{..}$ as in (3) above. Each value could then be substituted for $d_{..}$ in B^*. The latent roots of B^* could then be computed for each real positive value of $d_{..}$. One would be the desired value. This method would also involve a prohibitive amount of labor.

A third method would be to first estimate the dimensionality of the set of stimuli. To check the estimate, one could obtain an estimate of $d_{..}$ by evaluating one (or more) of the principal minors of B^* having an order equal to one greater than the estimated dimensionality. This estimate could then be substituted into B^* and the latent roots calculated.

There are a number of other methods possible involving the principal minors of the B^* matrix.† In general, they would all hinge on the fact that the correct estimate of $d_{..}$, if used in B^* results in (a) all principal minors of

---

†One could also use the matrix F (Equation 4) to obtain the value of $d_{..}$ which would minimize the dimensionality of the set of stimuli since it is known that the rank of F is two greater than the rank of B_i. There would seem to be little point in this, however, since F is a larger matrix, therefore involving more tedious procedures in evaluating the determinants, and since no properties of F have been given from which to determine whether the value of $d_{..}$ obtained will allow the stimuli to be placed in a *real* space.

order greater than the dimensionality of the stimuli being null, and (b) all principal minors of order equal to or less than the dimensionality being non-negative.

It is interesting to note that unless all points are equidistant from each other (in which case things collapse down to zero dimensions) it is always possible to obtain an estimate of $d_{..}$ in which the rank of B^* is at least two less than the number of stimuli. Thus before one could place any confidence in the obtained dimensionality of the stimuli, the rank of B^* would have to be smaller by three than the number of stimuli and preferably considerably smaller.

All of these methods require a great deal of labor. In addition, when fallible data are used, it is doubted whether the methods would give the solution we wish. *We would probably never obtain a positive semidefinite B^* matrix with a rank less than the number of stimuli minus two from fallible data.*

2. *Estimating $d_{..}$ from fallible comparative distances*

If we could obtain a positive semidefinite B^* whose non-zero roots consisted of a few large positive roots and a number of small positive roots by the methods outlined in the previous section, we could probably discard the small roots and conclude that the true dimensionality is equal to the number of large positive roots. Even in this case, however, there would probably be a better estimate of $d_{..}$. In the above example we have essentially assumed that any error must be such as to change the zero roots to positive values. It would be more reasonable to assume that errors would tend to change some zero roots in the positive direction and some in the negative direction. If we think of 3 points lying in a line so that $d_{12} + d_{23} = d_{13}$, the former would hold that any error would tend to make $(d_{12} + e_{12}) + (d_{23} + e_{23}) > (d_{13} + e_{13})$, whereas the latter would hold that $(d_{12} + e_{12}) + (d_{23} + e_{23}) < (d_{13} + e_{13})$ is equally likely.

This means that with fallible data the condition that B^* be positive semidefinite as a criterion for the points' existence in real space is not to be taken too seriously. What we would like to obtain is a B^*-matrix whose latent roots consist of

1. A few large positive values (the "true" dimensions of the system), and
2. The remaining values small and distributed about zero (the "error" dimensions).

It may be that for fallible data we are asking the wrong question. Consider the question, "For what value of $d_{..}$ will the points be most nearly (in a least-squares sense) in a space of a given dimensionality?" When one is interested in, or has reason to suspect, a one-dimensional case, the best $d_{..}$ in a least-squares sense is rather easy to obtain. In a one-dimensional set of points, $d_{ij} + d_{jk} = d_{ik} + e$ where j is between k and i, and e is an error.

In terms of available data $(d_{..} - d_{ik} = \hat{h}_{ik})$, this becomes

$$d_{..} - \hat{h}_{ij} + d_{..} - \hat{h}_{ik} = d_{..} - \hat{h}_{ik} + e$$

or

$$d_{..} + \hat{h}_{ik} - \hat{h}_{ij} - \hat{h}_{ik} = e.$$

The $d_{..}$ which will minimize the sums of squares of all of the e's would seem to be the $d_{..}$ desired. If we let

$$2F = \sum_{k>i>i} (d_{..} + \hat{h}_{ik} - \hat{h}_{ik} - \hat{h}_{ij})^2, \qquad (42)$$

then, to minimize $2F$, we take the derivative of F with respect to $d_{..}$ and set it equal to zero. Designating this derivative as F', we have

$$F' = \sum_{k>i>i} (\hat{h}_{ik} - \hat{h}_{ik} - \hat{h}_{ij}) + \sum_{k>i>i} d_{..} = 0, \qquad (43)$$

which rearranged becomes

$$\sum_{k>i>i} d_{..} = \sum_{k>i>i} (\hat{h}_{ij} + \hat{h}_{ik} - \hat{h}_{ik}). \qquad (44)$$

Dividing by $\dfrac{6}{n(n-1)(n-2)}$, we find

$$d_{..} = \frac{6}{n(n-1)(n-2)} \sum_{k>i>i} (\hat{h}_{ij} + \hat{h}_{ik} - \hat{h}_{ik}). \qquad (45)$$

In the one-dimensional case, it will ordinarily be possible to obtain the order of the n stimuli. If we define a symmetric matrix H_{ik} (j designates rows, k designates columns, $j, k = 1, 2, \cdots, n$) composed of elements $\hat{h}_{ik}$, the sum of the columns of H_{ik} divided by $(n-1)$ gives the average distance of all points from each other minus the average distance from point k to all other points. Small values of $\hat{h}_{.k}$ indicate that k is near one end of the continuum, and large values indicate that k is near the center. Inspection of H_{ik} will ordinarily suffice to determine on which side of the continuum the particular stimulus is located. Given matrix H_{ik} with rows and columns in correct order, a shortcut method of obtaining

$$\sum_{k>i>i} (\hat{h}_{ij} + \hat{h}_{ik} - \hat{h}_{ik}) = L$$

is to

1. Obtain the diagonal sums S_c of elements above the principal diagonal:

$$S_c = \sum_{j=1}^{n-c} \hat{h}_{j(j+c)} \qquad (c = 1, 2, \cdots, n-1). \qquad (46)$$

2. Multiply S_c by $(n - 2c)$. The sum is equal to L—i.e., if we let $t_c = (n - 2c)$,

$$L = \sum_{c=1}^{n-1} S_c t_c . \qquad (47)$$

For the case where the dimensionality is expected to be greater than one, this general approach does not seem to be very practical. While one could think of finding that $d_{..}$ which will minimize the sums of squares of volumes of all possible tetrahedrons for the two-dimensional case, and the corresponding hyper-volumes for the higher-dimensional cases, it would seem that the labor involved would again be prohibitive.

There is another procedure which might serve to give a fair estimate of $d_{..}$ for cases where the expected dimensionality is greater than one. If a one-dimensional subspace of four or more points exists in the data, that subspace could be used to estimate $d_{..}$. While this procedure does not give a "best fit" in the least-squares sense, it does appear to be the most practical method suggested thus far. The method has been applied to actual data and was found to work quite well. The existence of such a subspace is relatively easy to determine. One can compute, for each set of three stimuli, the value of $d_{..}$ which would be required to locate the set of three along a straight line. There are $n(n - 1)(n - 2)/6$ of these "estimates," one for each set of three different stimuli; and they will be designated as $\tilde{d}_{..}$. The values of $\tilde{d}_{..}$ may be obtained from the following equation:

$$\tilde{d}_{..} = \hat{h}_{ij} + \hat{h}_{jk} - \hat{h}_{ik} , \qquad \text{where} \qquad \hat{h}_{ik} < \hat{h}_{ij} , \hat{h}_{jk} . \qquad (48)$$

Given the $n(n - 1)(n - 2)/6$ values of $\tilde{d}_{..}$, the following points can be noted:

a. Except for error, points most nearly in a straight line will give the largest value of $\tilde{d}_{..}$.

b. If the four sets of three of any four points give about the same "highest" value of $\tilde{d}_{..}$ in a consistent manner, we can conclude that the four points are in a one-dimensional subspace. This value of $\tilde{d}_{..}$ would then be the estimate of $d_{..}$ wanted.

c. If such a set is not found, the largest value of $\tilde{d}_{..}$ might still be worth trying as an estimate of $d_{..}$. Using this value is equivalent to assuming that of the set of points at least one group of three is approximately linear. If one constructs a B-matrix with one of the points at the origin using this estimate and then finds that the third-order principal minors involving these three points vanish (approximately) this value of $\tilde{d}_{..}$ is probably a good estimate. The entire B-matrix need not, of course, be constructed. One would need to evaluate only $(n - 3)$ third-order principal minors involving only 3 $(n - 2)$ distinct elements instead of the $(n - 1)(n - 2)/2$ elements in the complete matrix.

Summary

A set of methods for multidimensional scaling based on Richardson's original model (3) have been developed, including a least-squares solution for obtaining comparative distances, and routine procedures for estimating the

additive constant necessary to convert comparative distances to absolute distances and for obtaining projections of stimuli on axes when fallible absolute distances are given. An outline of the procedures developed is given below.

A Routine Procedure for Multidimensional Scaling

A. To obtain comparative distances by the complete method of triads.

1. Construct the n matrices ${}_kP_{\cdot j}$ from the raw data.
2. Construct the corresponding matrices ${}_kX_{ij}$.
3. Obtain a row vector of averages of columns for each of the n matrices ${}_kX_{ij}$.

$$ {}_kx_{\cdot j} = \frac{1}{n-1} \sum_i {}_kx_{ij} \cdot $$

4. Construct matrix ${}_kX_{\cdot j}$ composed of these row vectors (k designates row, j designates columns).
5. Obtain a row vector of averages of columns of ${}_kX_{\cdot j}$.

$$ {}_{\cdot}x_{\cdot j} = \frac{1}{n} \sum_k {}_kx_{\cdot j} \cdot $$

6. Add the gth element of ${}_{\cdot}X_{\cdot j}$ to each element in the gth row of ${}_kX_{\cdot j}$. Call this new Matrix G_{kj}. Matrix G_{kj} thus contains the elements $({}_gx_{\cdot j} + {}_{\cdot}x_{\cdot g})$.
7. Average the symmetric elements of G_{kj} to obtain the symmetric matrix H_{jk}. Matrix H_{jk} is composed of the elements $\hat{h}_{jk} = d_{\cdot\cdot} - d_{jk}$, the comparative distances (with a negative sign) between stimuli.

$$ \hat{h}_{gk} = \hat{h}_{kg} = \tfrac{1}{2}(g_{gk} + g_{kg}). $$

B. To obtain an estimate of $d_{\cdot\cdot}$.

1. If the hypothesis of unidimensionality of stimuli seems reasonable:

   a. Arrange rows and columns of H_{jk} in order of magnitude of the stimuli by
      (1) Noting magnitudes of sums of columns of H_{jk}, and
      (2) Examining elements of H_{jk}.

   b. Obtain diagonal sums of H_{jk} above principal diagonal.

$$ S_c = \sum_j^{n-c} \hat{h}_{j(j+c)} \cdot $$

   c. Multiply each S_c by $(n - 2_c)$ and sum the products to obtain L.

$$ L = \sum_{c=1}^{n-1} S_c \, (n - 2c). $$

d. Divide L by $n\,(n-1)\,(n-2)/6$ to obtain $d_{..}$

$$d_{..} = \frac{L}{n(n-1)(n-2)/6}.$$

2. If it is reasonable to assume dimensionality greater than one with at least one set of four stimuli in a one-dimensional subspace:

a. Obtain the $n(n-1)\,(n-2)/6$ values of $\tilde{d}_{..}$ assuming in turn that each set of three stimuli lie in a line.

$$\tilde{d}_{..} = \hat{h}_{ij} + \hat{h}_{jk} - \hat{h}_{ik} ,$$

$$\hat{h}_{ik} < \hat{h}_{ij} , \hat{h}_{jk} .$$

b. The four sets of three of any four points lying in a line will give the same "highest" value of $\tilde{d}_{..}$ (except for error) in a consistent manner. If such a set is found, this value of $\tilde{d}_{..}$ will be a good estimate of $d_{..}$

c. If no such set is found, use the highest value of $\tilde{d}_{..}$ obtained as an estimate of $d_{..}$. Compute the necessary elements of a matrix B_i with one of the three points as the origin. Evaluate the $(n-3)$ third-order principal minors of B_i . If these minors all vanish (approximately) this $d_{..}$ is probably a good estimate.

C. To obtain projections of stimuli on axes.

1. Construct D_{ik} .

$$d_{ik} = d_{..} - \hat{h}_{ik} .$$

2. Construct B_i with origin at any stimulus i.

$$b_{ik} = \tfrac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2).$$

3. Obtain from B_i averages of

a. Columns,

$$b_{.k} = \frac{1}{n} \sum_j b_{jk} ,$$

b. Rows,

$$b_{j.} = \frac{1}{n} \sum_k b_{jk} ,$$

c. And all elements,

$$b_{..} = \frac{1}{n^2} \sum_k \sum_j b_{jk} .$$

4. Construct matrix B^* with origin at the centroid of stimuli.

$$b_{jk}^* = b_{jk} - b_{.k} - b_{j.} + b_{..}$$

5. Factor B^*, obtaining A^*_{jm}, the matrix of projections of stimuli j on arbitrary axes m.

6. Rotate and translate matrix A_{jm} to a meaningful set of dimensions if criteria for such are available.

REFERENCES

1. Gulliksen, Harold. Paired comparisons and the logic of measurement. *Psychol. Rev.* 1946, **53**, 199–213.

2. Klingberg, F. L. Studies in measurement of the relations among sovereign states. *Psychometrika*, 1941, **6**, 335–352.

3. Richardson, M. W. Multidimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659. (Abstract).

4. Thurstone, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389.

5. Young, G., and Householder, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1938, **3**, 19–22.