DOCUMENT RESUME

ED 078 058 TM 002 875

AUTHOR Subkoviak, Michael J.
TITLE The Effect of Nonnormality in Torgerson's Multidimensional Scaling Model.
PUB DATE 73
NOTE 15p.; Paper presented at annual meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Comparative Analysis; Correlation; *Error Patterns; *Models; *Multidimensional Scaling; *Statistical Analysis; Technical Reports
IDENTIFIERS *Torgerson Multidimensional Scaling Model

ABSTRACT

When Torgerson's multidimensional scaling model is used in conjunction with the method of tetrads, derived coordinates are based on data which is assumed to be distributed normally. The object of this study was to determine the amount of error contained in derived coordinates when the normality assumption is violated. Torgerson coordinates were derived from various cases of nonnormally distributed data. Derived coordinates were then compared for accuracy to true coordinate values, which were known in each case. The Torgerson model produced highly accurate coordinates in all cases. Since a wide range of distributions was considered, the results appear to be quite generalizable. (Author)

The Effect of Nonnormality in Torgerson's Multidimensional Scaling Model

Michael J. Subkoviak
University of Wisconsin - Madison

Introduction and Purpose

Given measures of dissimilarity between each pair of n objects $O_1, O_2, \ldots, O_n$, Torgerson's multidimensional scaling model determines: (a) the number of dimensions r needed to account for dissimilarities among the n objects and (b) the projection of each object on each dimension. The dimensions obtained may be identified with properties of the objects which account for input dissimilarities among objects. Also, the recovered coordinates can be used to construct an r-dimensional representation or picture which shows the relations among the objects in a more visualizable form than that provided by the complex set of dissimilarity measures. For example, Torgerson (1958, pp. 280-290) used perceived dissimilarities among n = 9 objects to recover an r = 2 dimensional picture which showed quite clearly dissimilarities among the objects. The two dimensions corresponded to color brightness and color saturation of the objects.

The method of tetrads (Torgerson, 1958, pp. 261-262) is a technique employed to obtain the dissimilarity measures used as input to Torgerson's multidimensional scaling model. The method of tetrads assumes that perceived dissimilarity between a pair of objects is normally distributed across a population of subjects. To the extent that this assumption is violated with real data, error is introduced into resulting dissimilarity values (Subkoviak, in press). These

erroneous input dissimilarities in turn affect the accuracy of the numerical coordinates output by the Torgerson model. Finally, inaccurate coordinates lead to fallacious conclusions regarding the relationships among the objects. Since nonnormal distributions of perceived differences between objects do occur in practice (Jones & Thurstone, 1955; Mosier, 1941; Rambo, 1963), the purpose of the study was to determine the amount of error introduced into derived coordinates when the normality assumption of the method of tetrads is violated.

The Mathematical Model

The Torgerson multidimensional scaling model is a two stage procedure (Torgerson, 1958, pp. 247-280). In the first stage a unidimensional scaling method such as the method of tetrads is used to obtain the perceived dissimilarity s_{jk} between each pair of objects (O_j,O_k). A constant is then added to convert each dissimilarity s_{jk} to a distance d_{jk} between pair (O_j,O_k) (Messick & Abelson, 1956). In the second stage, distances d_{jk} are converted to a matrix of scalar products b_{jk} between objects O_j and O_k, and the matrix is decomposed by factor analysis to obtain coordinates $(a_{j1},a_{j2},\ldots,a_{jr})$ for each object O_j $(j = 1,2,\ldots,n)$. Violations of the normality assumption in the first stage introduces error into the coordinates generated in the second stage. The sequence of steps involved in the two stages is outlined below.

First Stage

(1) In the method of tetrads all possible pairs of object-pairs $\{(O_h,O_i), (O_j,O_k)\}$ are presented to a population of subjects who report which object-pair is more dissimilar. Symbol $p_{hi,jk}$ represents the proportion of times (O_j,O_k) is reported more dissimilar than (O_h,O_i); $p_{jk,hi} = 1 - p_{hi,jk}$ is the proportion of times (O_h,O_i) is reported more dissimilar than (O_j,O_k).

(2) When pair $\{(O_h,O_i), (O_j,O_k)\}$ is presented to a subject he perceives the dissimilarity between O_h and O_i to have some magnitude S_{hi}; the perceived dissimilarity between O_j and O_k has magnitude S_{jk}. If $S_{jk} - S_{hi} > 0$, the subject reports pair (O_j,O_k) to be more dissimilar than pair (O_h,O_i); if $S_{jk} - S_{hi} < 0$, he reports pair (O_h,O_i) to be more dissimilar. Ties $S_{jk} - S_{hi} = 0$ are not allowed.

(3) The process in (2) gives rise to unseen distributions across the entire population of subjects: (a) a distribution of S_{jk} values with mean s_{jk} for each pair (O_j,O_k) and (b) a distribution of $S_{jk} - S_{hi}$ values with mean $s_{jk} - s_{hi}$ for each pair of object-pairs $\{(O_h,O_i), (O_j,O_k)\}$. Furthermore, reported proportions $p_{hi,jk}$ and

$p_{jk,hi}$ equal the areas to the right and left of zero (0) under the density function of $S_{jk} - S_{hi}$.

(4) The density of $S_{jk} - S_{hi}$ is assumed to be normal for each pair of object-pairs.

(5) In the most popular case, all $S_{jk} - S_{hi}$ distributions are also assumed to have the same variance $\sigma^2 = \text{Var}\ (S_{jk} - S_{hi})$.

(6) Under the assumptions of (4) and (5), the dissimilarity s_{jk} of pair (O_j, O_k) is given by:

$$s_{jk} = \sum_{hi} x_{hi,jk} / [n(n-1)/2] \qquad (1)$$

where $x_{hi,jk}$ is the normal deviate at the $p_{hi,jk}^{th}$ percentile of the standardized normal distribution and $p_{hi,jk}$ is the reported proportion associated with $\{(O_h, O_i), (O_j, O_k)\}$. The summation in Equation 1 is across the $n(n-1)/2$ values $x_{hi,jk}$ corresponding to all $\{(O_h, O_i), (O_j, O_k)\}$ which involve (O_j, O_k).

(7) Finally values s_{jk} given by Equation 1 are used to compute a constant c which is added to each s_{jk} to obtain a distance $d_{jk} = s_{jk} + c$ between O_j and O_k (Messick & Abelson, 1956).

Second Stage

(8) The distances d_{jk} obtained in (7) are used to compute an n x n matrix of scalar products b_{jk} between each pair (O_j, O_k) $(j,k = 1,2,\ldots,n)$.

(9) The scalar product matrix in (8) is decomposed by factor methods to obtain an n x r matrix of factor loadings. The j^{th} row of the factor loading matrix gives the desired coordinates $(a_{j1}, a_{j2}, \ldots, a_{jr})$ of object O_j. The accuracy of these coordinates is a function of the validity of the normality assumption in Step (4) of the first stage.

Method

The procedure used in the study was as follows:

(1) Coordinates were randomly specified for n = 8 hypothetical objects in r = 2 dimensional Euclidean space. The coordinates and their locations in Euclidean space are shown in Figure 1. The object of the study was to determine the extent to which this known (true) configuration of coordinates could be recovered by Torgerson multidimensional scaling.

(2) The 8(8-1)/2 = 28 true Euclidean distances d_{jk} between pairs of coordinate points in (1) were computed.

(3) The (28)(28-1)/2 = 378 differences $d_{jk} - d_{hi}$ between pairs of distances in (2) were computed. These differences $d_{jk} - d_{hi} = (s_{jk}+c) - (s_{hi}+c) = s_{jk} - s_{hi}$ are the means of the distributions $S_{jk} - S_{hi}$ which are assumed to be normal by the method of tetrads.

(4) A specific form of density function was specified for each $S_{jk} - S_{hi}$ distribution with the mean $s_{jk} - s_{hi}$ computed in (3). In some cases the 378 specified densities were normal as assumed by the method of tetrads, and in other cases the densities were nonnormal. Some of the nonnormal densities used in the study are compared to the assumed normal in Figures 2-5. All densities specified had the same variance $\sigma^2 = \text{Var}\ (S_{jk} - S_{hi}) = (115)^2$. The rationale for choosing this value of σ^2 is indicated in (5) below.

(5) The density functions $S_{jk} - S_{hi}$ specified in (4) were integrated to the right and left of zero (0) to obtain proportions $p_{hi,jk}$ and $p_{jk,hi}$ (areas under the density curves) which are required to compute s_{jk} in Equation 1. The choice of $\sigma^2 = \text{Var}\ (S_{jk} - S_{hi}) = (115)^2$ avoided the occurrence of extreme proportions $p_{hi,jk} = 1$ or 0 which result in normal deviates $x_{hi,jk} = \pm \infty$ in Equation 1. When

such proportions arise in actual practice they are ignored, and s_{jk} values are obtained in another way.

(6) The proportions $p_{hi,jk}$ from (5) were used in conjunction with Equation 1 to obtain $n(n-1)/2$ dissimilarities s_{jk}. Since the computational method assumes that all proportions $p_{hi,jk}$ arise from normal $S_{jk} - S_{hi}$ densities, error was introduced into s_{jk} values which is propagated in all subsequent computations to the coordinates derived by the model.

(7) The s_{jk} values from (6) were used to compute constant c (Messick & Abelson) which was added to each s_{jk} value to obtain distances $d_{jk} = s_{jk}+c$. The distances were converted to scalar products b_{jk}, and the matrix of scalar products was factored to obtain coordinates which contained error intorduced by the nonnormal densities specified in (5).

(8) Finally, the derived coordinates (a_{j1},a_{j2}) $(j = 1,2,\ldots 8)$ from (7) were compared for accuracy to the known true coordinates of Figure 1. Accuracy was measured by correlating the 28 distances $\hat{d}_{jk}$ between derived coordinates with the corresponding true distances d_{jk} between coordinates in Figure 1.

Results and Discussion

Thirty-six different cases were considered by specifying different density functions in Step (4) of the Method Section. The densities varied in skewness and kurtosis and ranged in shape from normal to extremely nonnormal. Coordinates (a_{j1},a_{j2}) $(j = 1,2,\ldots,n)$ were derived in each of the thirty-six cases, and a correlation coefficient was computed as described in Step (8) of the Method Section. The

obtained correlation coefficients are shown in Table 1. Each coefficient in Table 1 is a measure of the accuracy of the coordinates derived for that case. In Cases 1-22 of Table 1, the 378 $S_{jk} - S_{hi}$ densities specified in Step (4) were identical in shape, although the shape differed in each case. In Cases 23-36 the 376 densities specified in Step (4) were not identical in shape, i.e. a variety of shapes was considered simultaneously.

The primary result was that the Torgerson model produced highly accurate coordinates for all thirty-six cases. In no instance did the measure of correlation in Table 1 fall below .999. Considering the variety of densities employed in the study, it appears that the Torgerson model is quite robust with respect to violations of the normality assumption inherent to the method of tetrads. The following results of theoretical interest were also noted: (a) the accuracy of coordinates tended to decrease slightly as densities departed from normality and (b) the accuracy of coordinates tended to be greater when densities were identical in shape as in Cases 1-22 than when densities varied in shape as in Cases 23-36.

These results are consistent with other findings. Specifically, it has been demonstrated that Thurstone's comparative judgment method, of which the method of tetrads is a derivative, produces reasonably accurate estimates s_{jk} as defined by Equation 1 regardless of the density functions from which proportions $p_{hi,jk}$ arise (Subkoviak, in press). Thus, given reasonably accurate s_{jk} values it is not surprising that accurate coordinates result. Furthermore, increasing error is introduced into s_{jk} values as densities depart from assumed normality or become more heterogeneous in shape (Subkoviak, in press).

Thus, results (a) and (b) noted above can also be accounted for.

While it appears that the researcher need not be overly concerned about violations of the normality assumption in the Torgerson model, certain limitations of the study should be born in mind. First, only a single random arrangement of eight points in two dimensional space was considered. Different configurations of points or different numbers of dimensions might produce other results. Second, the effect of extreme proportions such as $p_{hi,jk} = 1$ or 0 on resulting coordinates was not considered. Finally, all $S_{jk} - S_{hi}$ densities considered had the same variance as assumed by the model. Violations of this assumption could result in less accurate coordinates than those derived in the present study. For further details regarding violation of this homogeneity of variance assumption consult Mosteller (1951).

References

Jones, L.V. & Thurstone, L.L. The psychophysics of semantics: An empirical investigation. Journal of Applied Psychology, 1955, 39, 31-36.

Messick, S.J. & Abelson, R.P. The additive constant problem in multidimensional scaling. Psychometrika, 1956, 21, 1-17.

Mosier, C.G. A psychometric study of meaning. Journal of Social Psychology, 1941, 13, 123-140.

Mosteller, F. Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. Psychometrika, 1951, 16, 203-206.

Rambo, W.W. The distribution of successive interval judgments of attitude statements: A note. Journal of Social Psychology, 1963, 60, 251-254.

Subkoviak, M.J. Remarks on the method of paired comparisons: The effect of nonnormality in Thurstone's comparative judgment model. Educational and Psychological Measurement, in press.

Torgerson, W.A. Theory and Methods of Scaling. New York: Wiley, 1958.

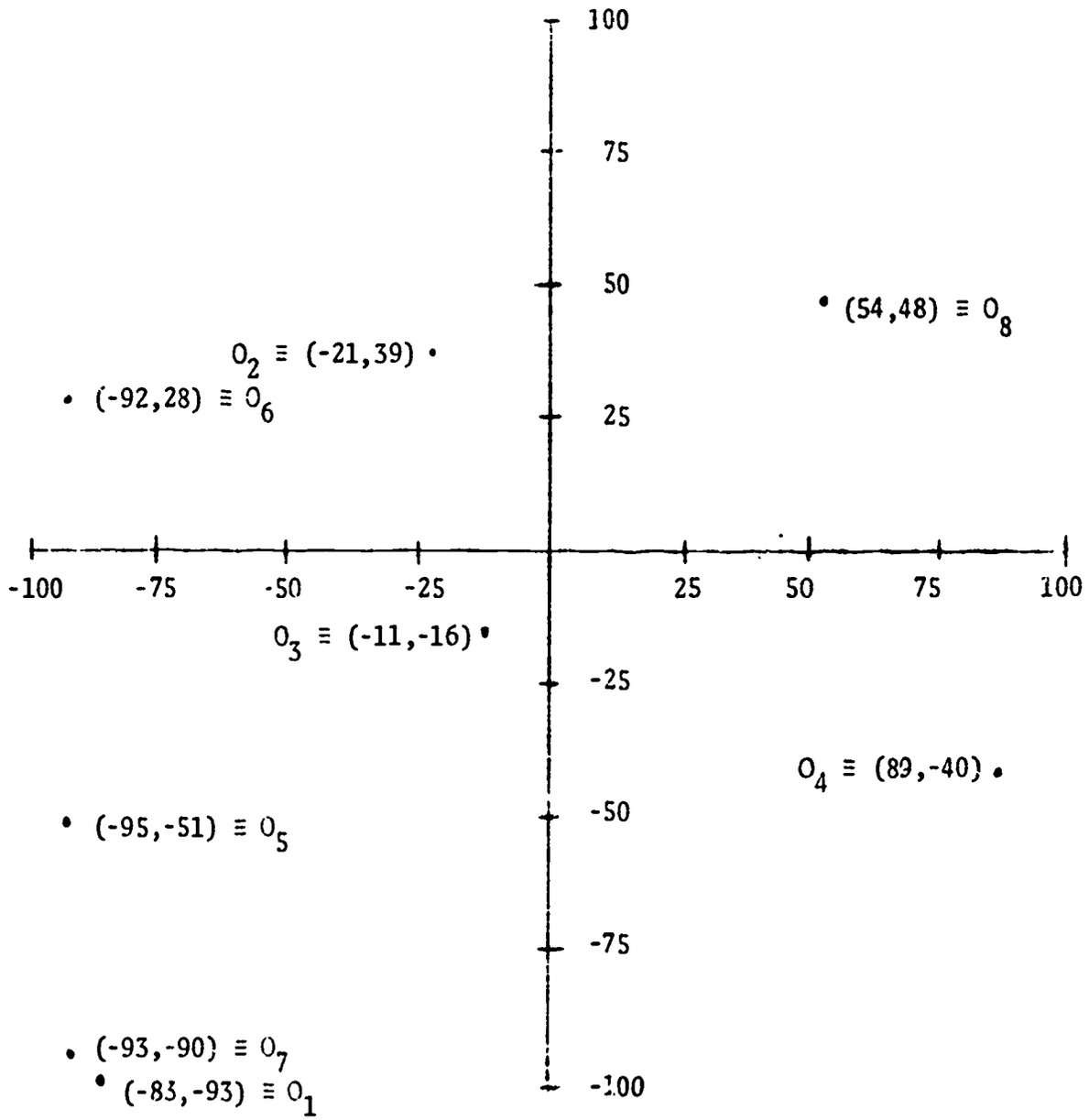
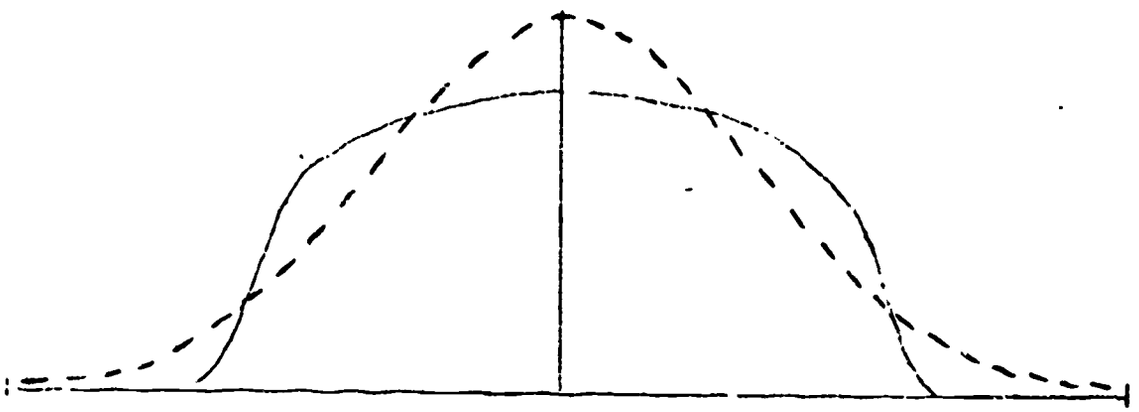


Figure 1

True Configuration of Eight Points in Two-Space

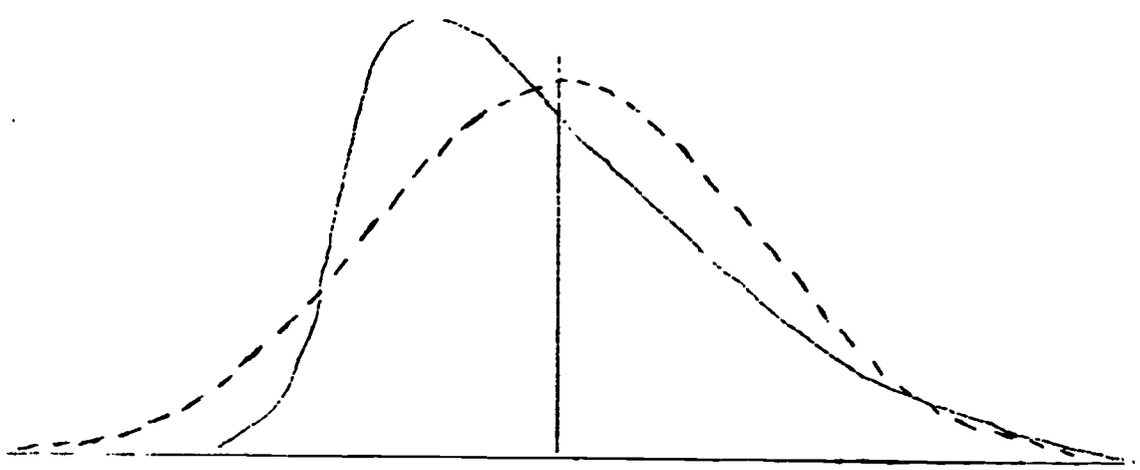


$s_{jk} - s_{hi}$

Figure 2

——— Nonnormal Density

----- Normal Density

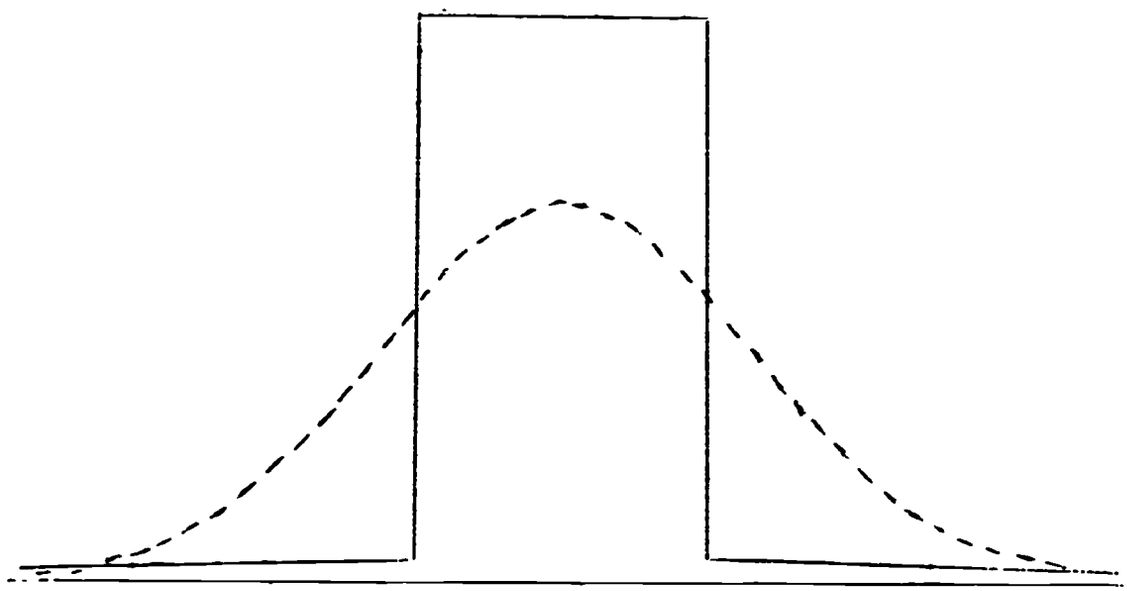


$s_{jk} - s_{hi}$

Figure 3

——— Nonnormal Density

----- Normal Density

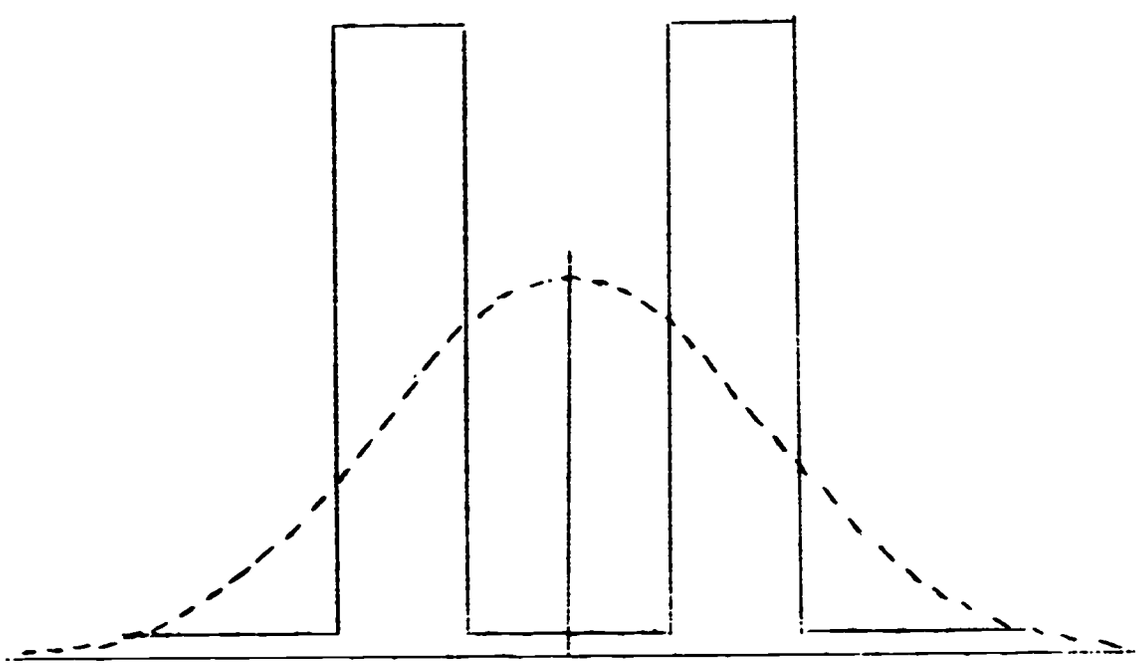


$s_{jk} - s_{hi}$

Figure 4

——— Nonnormal Density

----- Normal Density



$s_{jk} - s_{hi}$

Figure 5

——— Nonnormal Density

----- Normal Density

TABLE 1

Correlation of True Distances d_{jk} With Derived Distances $\hat{d}_{jk}$

Case	Correlation
1	.999999
2	.999999
3	.999999
4	.999998
5	.999989
6	.999984
7	.999982
8	.999980
9	.999976
10	.999972
11	.999970
12	.999967
13	.999966
14	.999962
15	.999960
16	.999959
17	.999954
18	.999954
19	.999950
20	.999948
21	.999938
22	.999124
23	.999979
24	.999977
25	.999976
26	.999974
27	.999974
28	.999974
29	.999962
30	.999962
31	.999962
32	.999957
33	.999957
34	.999929
35	.999805
36	.999465