AN ALTERNATIVE APPROACH TO CONFIRMATORY INFERENCE AND GEOMETRIC MODELS

INGWER BORG

Rheinisch–Westfälische Technische Hochschule

JAMES C. LINGOES

The University of Michigan

Introduction

Hubert and Subkoviak (1979) have recently proposed a confirmatory technique which allows testing whether a given proximity matrix or some geometrical representation of it can be said to possess certain predicted structural properties. Their approach is essentially very simple. They compute the correlation between the set of values $q(o_i, o_j)$, which are either the proximity values for objects o_i and o_j or a distance value for the respective point-representations, and the collection of $c(o_i, o_j)$, the numerical assignments given to these pairs on the basis of some theory. Thus, if an observation of $q(o_i, o_j) > q(o_k, o_l)$ were hypothesized, then, for example, $c(o_i, o_j) = 2 > 1 = c(o_k, o_l)$ might be defined. Having set up a "structure" function c in this way, a (linear or rank-order) correlation between c and q is then an index of how well the actual data or distances correspond (linearly or ordinally) to this particular choice of numerical predictions. That is, "the theory used to generate the function $c(o_i, o_j)$ is given empirical support if the two sets of elements, $c(o_i, o_j)$ and $q(o_i, o_j)$, have a similar patterning of high and low entries" (Hubert and Subkoviak, 1979, p. 363).

In order to be able to evaluate the probability with which such a correlation can be expected to occur for random data, it is possible simply to compute the correlations for all — or a reasonably large sample — of the permutations of the sets of $c(o_i, o_j)$ or $q(o_i, o_j)$ values, respectively. If the observed correlation then exceeds some sufficiently high percentage point of the cumulative distribution of the generated set of

coefficients, it is possible to reject the null hypothesis of randomness.

Although this approach is, in principle, a useful one, it may often not be very interesting to test such a relatively weak hypothesis. It seems that a more common type of research question would involve testing two competing hypotheses rather than just one hypothesis versus pure noise. In the following, we illustrate such a test by using one of the data sets that was also employed by Hubert and Subkoviak to illustrate their procedure. This test, which does not require the generation of empirical test distributions, is particularly interesting in the context of confirmatory multidimensional scaling analyses, as is shown below.

A Test for Predictor Effectiveness

Hotelling (1940) described a statistical method for distinguishing which of two or more predictors is most "effective" in explaining a common dependent variable y. For simplicity, let there be just two predictors, x_1 and x_2, with fixed values, and a dependent variable y consisting of fallible data. The null hypothesis that we are concerned with in this paper asserts that the residual variances about the regression lines of y on x_1 and y on x_2, respectively, do not differ statistically. Note that this hypothesis is not the same as the one in the usual test for comparing two correlation coefficients, insofar as (a) it involves a common dependent variable y, and (b) the independent variables x_1 and x_2 are not subject to random variation but are fixed — just as in tests for the significance of regression coefficients.

Assume, then, that each observed value y_i consists of a true component η_i and an error term ϵ_i, i.e., $y_i = \eta_i + \epsilon_i$, where ϵ_i is taken randomly from a normal distribution with variance σ^2. Moreover, $E(y_i) = \eta_i = \alpha_i + \beta_1 x_{1i} + \beta_2 x_{2i}$, i.e., the expected value of y_i is a linear combination of x_1 and x_2. For convenience, the predictor variables x_1 and x_2 can be standardized to z-score form. This makes $\Sigma z_{1i} = 0 = \Sigma z_{2i}$ and $\Sigma z_{1i}^2 = 1 = \Sigma z_{2i}^2$, where z_k is the standardized x_k, $k = 1, 2$. The regression weights for predicting y from z_1 and z_2, respectively, are then

$$b_k = \Sigma_i y_i z_{ki} , \quad k = 1, 2 \tag{1}$$

The unexplained variance, u_k^2, is equal to $\Sigma_i(y_i - \hat{y}_i)^2 = \Sigma [y_i - (a_k + b_k z_{ki})]^2$, with $a_k = \bar{y}$. This simplifies to

$$\Sigma_i(y_i - \bar{y})^2 - (\Sigma_i y_i z_{ki})^2 = u_k^2 \tag{2}$$

Now, evaluation of the difference between u_1^2 and u_2^2 leads to

$$u_1^2 - u_2^2 = b_1^2 - b_2^2 \tag{3}$$

or

$$b_1 = b_2 \tag{4}$$

as our H_0. Thus, $b_1 - b_2 = \Sigma_i y_i z_{1i} - \Sigma_i y_i z_{2i} = \Sigma_i y_i (z_{1i} - z_{2i})$. Replacing y_i by $\eta_i + \epsilon_i$ in the last term, and computing its variance, yields $\Sigma_i \sigma^2 (z_{1i} - z_{2i})^2$. To estimate σ^2, y is regressed on z_1 and z_2 simultaneously and the residual variance, s^2, is set equal to σ^2. The significance of the difference $b_1 - b_2$ can then be tested by an ordinary t-test, $t = (b_1 - b_2)/\sqrt{\Sigma_i s^2 (z_{1i} - z_{2i})^2}$, with $n - 3$ degrees of freedom (see Healey, 1955; Williams, 1959; Lingoes and Borg, 1980).

An equivalent, but more convenient formula is provided by Hotelling (1940). To use it, only the following correlations are required: $r_1 = r(y, x_1)$, $r_2 = r(y, x_2)$, and $r_0 = r(x_1, x_2)$. The statistic

$$(r_1 - r_2)[(n - 3)(1 + r_0)/D]^{1/2} \tag{5}$$

where

$$D = \begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_0 \\ r_2 & r_0 & 1 \end{vmatrix}$$

is then distributed as t with $n - 3$ degrees of freedom. If the probability for a given t-value "is sufficiently small . . . we have a corresponding degree of confidence that the variate chosen because of higher correlation in the sample has actually a higher correlation than the other in the population" (Hotelling, 1940, p. 278).

The statistic (5) is distributed as t only if the assumptions discussed above are met. That is, the values of y have to be independently and normally distributed with equal variance, and the expected values of y must be a linear function of x_1 and x_2. In the context in which we want to use the test, the dependent variable y will be a set of dissimilarity scores or distance estimates. (The predictors x_1 and x_2 are some fixed numerical vectors and no assumptions are involved on their side.) For such a y-variable, it is not always possible simply to assume independency and normality. The first problem, i.e. that errors be uncorrelated, does not appear to be a very serious one. Yet, in any case, the issue may be eliminated entirely by substituting the values in y by their respective ranking numbers. This weakens the test somewhat in cases where y is a metric variable, but has other advantages (see below). As to the distribution assumption, it has been shown, for example, that for ratio-scaled distance estimates the non-central χ^2 or the log-normal distributions are more reasonable models than the normal curve (Ramsay, 1969, 1977).

Although, except for small values, the former distributions do not differ greatly from the normal one, it is quite unnecessary to worry about such distributional properties at all, since Lingoes and Borg (1980) have demonstrated that the test is extremely robust with respect even to serious violations of the normality assumption. Finally, the linearity assumption requires also that the values in x_1 and x_2 be substituted by their respective ranking numbers to allow for perfect correlations. Moreover, if x_1 and x_2 are distances of non-metric MDS solutions, the ranking-number substitutions have the advantage of linearizing the relationships between y and x_k, $k = 1, 2$ (Weeks and Bentler, 1979). Lingoes and Borg (1980) have theoretically and empirically investigated a great variety of different assumptions and transformations and have come to the conclusion that — with the ranking-number substitutions — "one need but assume that y is a random variable or a random permutation for the statistical inference to be valid" (Lingoes and Borg, 1980, p. 15).

EXAMPLE 1: A DIRECT COMPARISON OF THE EFFECTIVENESS OF TWO STRUCTURE MATRICES

In the following illustrations, we use Glushko's data on pattern goodness and redundancy (Glushko, 1975). Since these data were also used by Hubert and Subkoviak (1979), the reader can easily see where and how our and their approaches differ from each other.

Glushko was interested in testing the hypothesis that "a pattern's goodness is related to its redundancy as measured by the number of patterns inferred from or equivalent to it" (Glushko, 1975, p. 158). He constructed seventeen dot-patterns (Fig. 1), and set up 136 cards each representing a different pair of patterns. The subjects were then asked to indicate which pattern on each card was the "better" one. Summing the preferences over all twenty subjects, a 17 × 17 preference matrix was obtained. Proximity measures were then derived, based on the following logic: "Since dissimilar goodness between two patterns is implied by frequent choice of either over the other, the absolute value of the difference between the observed and the expected frequency of a goodness preference represents the similarity of the pattern goodness of the two patterns . . ." (Glushko, 1975, p. 159). Since there were twenty subjects, the expected (chance) preference value for each pair was 10. Thus, subtracting 10 from each entry of the preference matrix and taking the absolute value, the proximity matrix in the lower half of Table I is generated.

It may be observed in Fig. 1 that the set of patterns is partitioned

Fig. 1. The Seventeen Patterns used in Glushko's Experiment. Equivalence Sets correspond to the Number of Rotations in 90° Increments and Reflections that Generate the Same Dot Pattern.

into three groups on the basis of their "redundancy", i.e. the number of different configurations resulting from reflections or rotations of a given pattern in 90° increments. Pattern 3, for example, thus generates four different "T" configurations, whereas pattern 1 looks the same under all of these transformations, or, in other words, "it suggests only itself".

Hubert and Subkoviak (1979) used these group measures directly as a structural hypothesis. They therefore defined the upper half matrix in Table I as their "structure matrix". Correlating all corresponding values in both halves of this table yields $r = 0.64$, which is higher than any value of the empirical test distribution. It is thus concluded that the equivalence-set hypothesis is significantly better than chance.

As Hubert and Subkoviak pointed out themselves, other monotonic transformations of this structure matrix might also be considered. Indeed, the theory says that the goodness of a pattern is simply "related to its redundancy as measured by the number of patterns inferred from or equivalent to it" (Glushko, 1975, p. 158). Specifying what "related" means, would, so it seems, require that the correspondence be at least weakly monotonic. If the structure matrix in Table I is weakened such that its elements are required to express only such a monotonic hypothesis, it has to be decided whether, by so doing, the percentage of variance explained in the data can be significantly increased.

TABLE I

Glushko's Proximity Matrix (lower half) and Hubert and Subkoviak's Structure Matrix (upper half)

Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	x	3	3	3	3	3	3	3	3	3	7	7	7	7	7	7	7
2	1	x	3	3	3	3	3	3	3	3	7	7	7	7	7	7	7
3	1	2	x	0	0	0	0	0	0	0	4	4	4	4	4	4	4
4	2	4	0	x	0	0	0	0	0	0	4	4	4	4	4	4	4
5	3	3	1	1	x	0	0	0	0	0	4	4	4	4	4	4	4
6	2	4	1	1	1	x	0	0	0	0	4	4	4	4	4	4	4
7	2	4	1	2	1	2	x	0	0	0	4	4	4	4	4	4	4
8	3	5	3	1	1	1	0	x	0	0	4	4	4	4	4	4	4
9	4	4	2	1	2	3	3	4	x	0	4	4	4	4	4	4	4
10	4	5	2	1	5	3	3	5	4	x	4	4	4	4	4	4	4
11	5	5	4	4	3	0	2	3	1	1	x	0	0	0	0	0	0
12	5	6	3	6	3	1	5	5	2	1	3	x	0	0	0	0	0
13	6	7	4	6	4	4	5	6	5	1	4	1	x	0	0	0	0
14	7	6	7	4	5	4	6	5	4	2	4	1	1	x	0	0	0
15	6	7	5	7	4	5	5	4	5	0	3	0	0	0	x	0	0
16	7	8	5	5	6	4	4	3	4	1	4	2	2	0	1	x	0
17	7	7	5	5	5	6	5	4	6	3	6	2	3	1	1	1	x

Although .many monotonic transformations might be considered here, we choose a particularly simple one which is likely to increment the correlation between the structure values and the data. This transformation is a mapping of the proximities in Table I into the structure matrix such that the rank order of the image values corresponds to the rank order of the structure coefficients. Such a mapping is known as a "rank image transformation" (Guttman, 1968). The structure matrix will then contain all the small-proximity data in those blocks that previously contained the zeros; the block consisting of 3's in Table 1 will then absorb the smallest remaining proximities, etc. Thus a structure matrix is obtained that is monotonic with Hubert and Subkoviak's. (Strictly speaking, it is only weakly monotonic since, for example, some of the largest elements in the "zero"-block are equal to the smallest values in the "three"-block, etc. There are, however, only few such "weak" links.) This structure matrix can be brought into optimal correspondence to the proximities by permuting the elements within each block such that the sum of the squared differences is minimized. Lingoes and Roskam (1973) term this method the "primary approach to ties".

The correlation between this monotonic structure matrix and the data is computed as $r = 0.69$, which is only slightly higher than the measure for the values in Table I ($r = 0.64$). However, applying Hotelling's test to these results, a t-value of 2.13 with 133 degrees of freedom is found, which indicates that the difference is significant. It can therefore safely be concluded that the redundancy hypothesis in its strong, metric form makes better than chance predictions, but some monotonic version of the theory is even better and further theory development in this direction may prove fruitful.

EXAMPLE 2: A CONFIRMATORY MULTIDIMENSIONAL SCALING APPROACH TO THE REDUNDANCY HYPOTHESIS

Glushko (1975) analyzed his data not in the confirmatory fashion that Hubert and Subkoviak (1979) proposed. Rather, he scaled the proximity matrix via an ordinal MDS procedure and interpreted the resulting two-dimensional configuration. He found that one dimension allowed a partitioning of the space into three regions which contained points belonging to just one redundancy class. (There were just two errors.) Moreover, these regions were ordered along this dimension as predicted. A hierarchical cluster analysis induced the same three "cuts" into the space. In Fig. 2, a similar scaling configuration obtained via SSA-I (Lingoes, 1973) is represented. Its stress is $S = 0.11$.
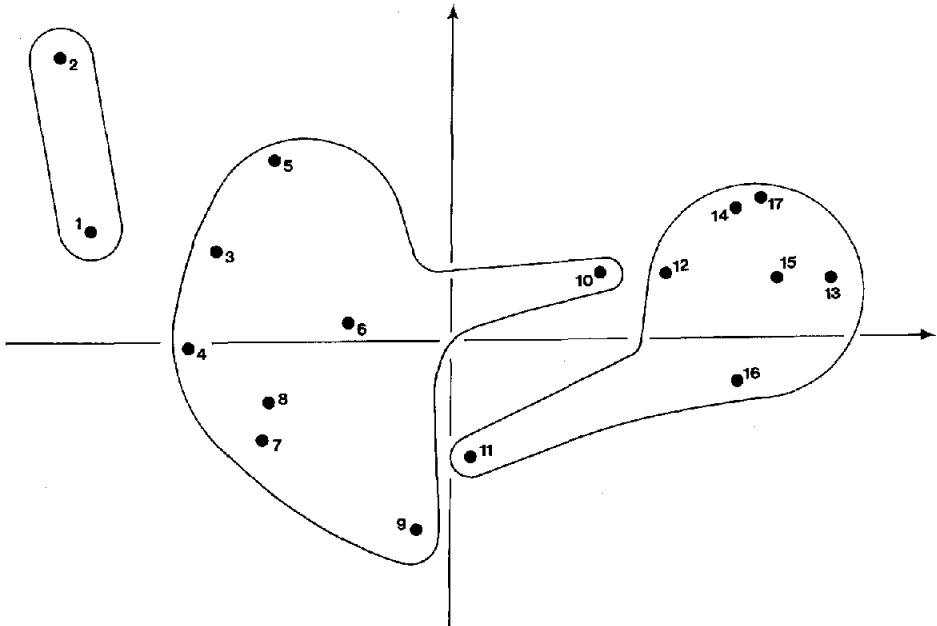
Fig. 2. SSA-I Representation of Proximity Matrix in Table I (lower half).

The horizontal axis in Fig. 2 reflects the group-membership informa-
tion. It can be seen, however, that points 10 and 11 are not in their
appropriate regions: 10 should be located more to the left, 11 to the
right.

It can be enforced, of course, that these points move such that the
regionalization hypothesis is perfectly expressed by the configuration.
For that purpose, we use the CMDA procedure recently developed by
Borg and Lingoes (1979, 1980) and Lingoes and Borg (1978). CMDA
allows specification of certain side constraints on the distances of the
MDS configurations. In this case, we impose some inequalities on the
distances which will make point 10 closer to any point in its appropri-
ate group {3, ..., 9} than to any point in {11, ..., 17}; the opposite con-
straints are imposed on point 11. With these restrictions, CMDA pro-
duces the configuration in Fig. 3, which perfectly satisfies the redun-
dancy-class conditions. Except for points 10 and 11, Fig. 3 is not greatly
different from Fig. 2. However, the fit of these two points is now much
worse and, consequently, the total stress is increased by some 5%, to
$S = 0.16$.

The question to be asked now is, of course, how such an increment
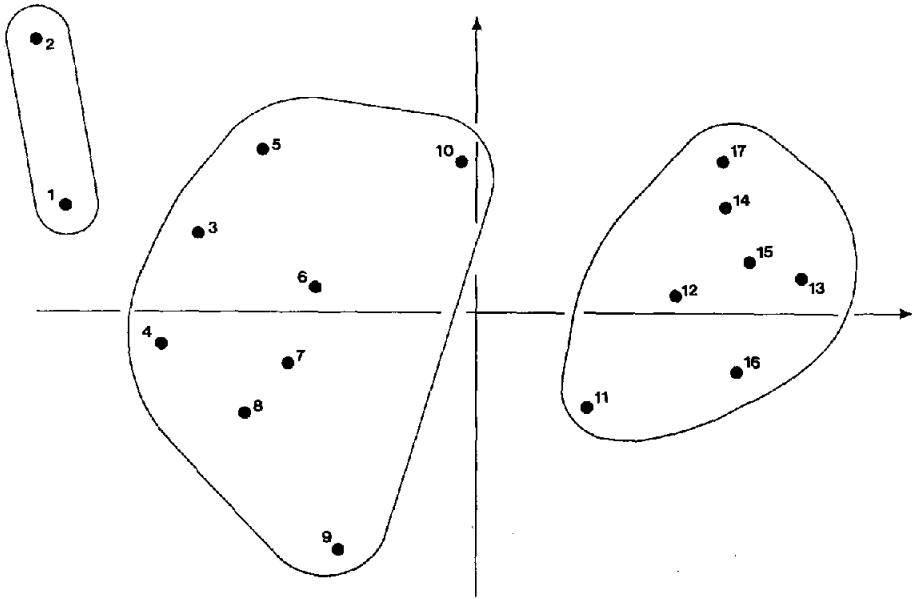in stress should be evaluated. We give here a statistical answer by using

Fig. 3. CMDA Configuration of Glushko's Proximity Data with Weak Contiguity Constraints on Points 10 and 11.

Hotelling's test again. For that purpose, the distances computed from the SSA and the CMDA solutions are used as the two predictor sets x_1 and x_2 for the dependent variable y consisting of the proximity values. (Remember though that ranking numbers are substituted for the values in all three variables.) A t-statistic of 3.149 with 133 degrees of freedom is obtained in this comparison. This value is significant, indicating that the unconstrained and the constrained MDS solutions are indeed reasonably different in explaining the proximities. It is probably worthwhile, therefore, to take a closer look at the patterns 10 and 11 since they are, in fact, responsible for this difference. It may be that the subjects make use of other redundancy features besides the indicated motions.

In the preceding scaling analysis, however, only the weakest of a hierarchy of "contiguity" hypotheses was used. Lingoes (1979) has recently described in quite some detail how such predictions may gradually be strengthened. The strongest-contiguity type, in his system, is a clustering which requires each point in a given region to be closer to every other point within this region than to any outside point. Yet, that is just what Hubert and Subkoviak's structure matrix implies if it is interpreted ordinally under the primary approach to ties. (The
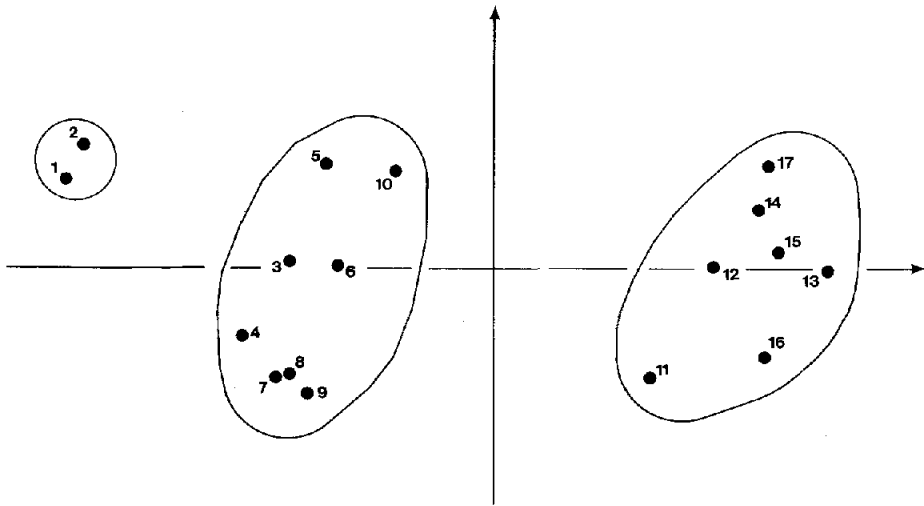
Fig. 4. CMDA Configuration of Glushko's Proximity Data with Strong Clustering Constraints on the Equivalence Sets.

primary approach to ties means that it is not insisted that all distances in a given group should be equal — as the structure matrix in Table I indicates — but that they should simply be smaller/greater than those in the respective adjacent blocks.) Since the constraints for a CMDA scaling analysis also are formulated by setting up a structure matrix, Table I may be taken directly in order to generate a strong-contiguity solution for these data. The result is shown in Fig. 4.

The effects of the clustering constraints are obvious from visual inspection of the configuration in Fig. 4. The stress is increased to $S = 0.23$, which is more than twice that for the SSA solution ($S = 0.11$) and also substantially higher than that obtained for the weak-contiguity hypothesis ($S = 0.16$).

It is, of course, clear that this solution is significantly worse than the SSA result. It is also not surprising that it differs significantly from the weak-contiguity solution ($t = 3.45$ with 133 degrees of freedom).

Discussion

The test described above is not intended to be a substitute for the procedure proposed by Hubert and Subkoviak (1979), although it represents an alternative if it is desired to answer such a null hypothesis

as used by them. The necessary comparison correlation would then be one taken from the empirically generated test distribution. We feel, however, that the researcher is typically not interested in testing a structural hypothesis against pure noise. Under such conditions, Monte Carlo techniques would not only be more complicated and expensive but also unnecessary: Hotelling's test does not require such computations.

The test has been generalized to deciding among more than two predictors. We have not discussed this problem here since it, too, does not seem to be the typical question a researcher is asking. The interested reader may here consult Hotelling (1940) or Williams (1959).

In concluding, we would like to mention an important comment made by an anonymous reviewer. S/he pointed out that "the test does not take into account the amount of prior information used in generalizing a structured model. Suppose a highly structured model based on theoretically important hypotheses fits slightly but significantly worse than a completely exploratory model. The structured model would still be preferred". Exactly the same point is also made by Lingoes and Borg (1980). We must admit that we have no simple solution for this problem, nor do we know anyone who does. Therefore, we can only recommend that the statistical results of our tests are considered as just another piece of information in the process of theory construction in which the ultimate statistical test is provided by invariance of findings over a large number of replications.

References

Borg, I. and Lingoes, J.C. (1979). "Multidimensional scaling with side constraints on the distances", pp. 753–790 in J.C. Lingoes, E.E. Roskam and I. Borg, eds., *Geometric Representations of Relational Data.* Ann Arbor, MI: Mathesis Press.

Borg, I. and Lingoes, J.C. (1980). "A model and algorithm for multidimensional scaling with external restrictions on the distances", *Psychometrika* 45: 25–38.

Glushko, R.J. (1975). "Pattern goodness and redundancy revisited: multidimensional scaling and hierarchical clustering analyses", *Perception and Psychophysics* 17: 158–162.

Guttman, L. (1968). "A general nonmetric technique for finding the smallest coordinate space for a configuration of points", *Psychometrika* 33: 469–506.

Healey, M.J.R. (1955). "A significance test for the difference in efficiency between two predictors", *Journal of the Royal Statistical Society B* 17: 266–268.

Hotelling, H. (1940). "The selection of variates for use in prediction with some comments on the general problem of nuisance parameters", *Annals of Mathematical Statistics* 11: 271–283.

Hubert, L.J. and Subkoviak, M.J. (1979). "Confirmatory inference and geometric models", *Psychological Bulletin* 86: 361–370.

Lingoes, J.C. (1973). *The Guttman–Lingoes Nonmetric Program Series*. Ann Arbor, MI: Mathesis Press.

Lingoes, J.C. (1979). "Testing regional hypothesis in multidimensional scaling", paper presented at the Second International Symposium on Data Analysis and Informatics, Versailles, France, October 17–19, 1979. Reprinted in I. Borg, ed. (1981). *Multidimensional Data Representations: When and Why*. Ann Arbor, MI: Mathesis Press.

Lingoes, J.C. and Borg, I. (1978). "CMDA-U: confirmatory monotone distance analysis – unconditional", *Journal of Marketing Research* 15: 610–611.

Lingoes, J.C. and Borg, I. (1980). "An exact significance test for dimensionality and for choosing between rival hypotheses in confirmatory multidimensional scaling", *Michigan Computing Technical Reports* 1.

Lingoes, J.C. and Roskam, E.E. (1973). "A mathematical and empirical analysis of two multidimensional scaling algorithms", *Psychometrika* 38 (*Monograph Supplement*). Reprinted in J.C. Lingoes, E.E. Roskam and I. Borg, eds. (1979). *Geometric Representations of Relational Data*. Ann Arbor, MI: Mathesis Press.

Ramsay, J.O. (1969). "Some statistical considerations in multidimensional scaling", *Psychometrika* 34: 167–182.

Ramsay, J.O. (1977). "Maximum likelihood estimation in multidimensional scaling", *Psychometrika* 42: 241–266.

Weeks, D.G. and Bentler, P.M. (1979). "A comparison of linear and monotone multidimensional scaling models", *Psychological Bulletin* 86: 349–354.

Williams, E.J. (1959). "The comparison of regression variables", *Journal of the Royal Statistical Society* B21: 396–399.