On evaluating the equivalency
of alternative MDS representations

JAMES C. LINGOES [1] & INGWER BORG [2]
[1] *The University of Michigan;* [2] *Justus Liebig Universität*

Abstract. A typical question in MDS is whether two alternative configurations that are both acceptable in terms of data fit may be considered "practically the same". To answer such questions on the equivalency of MDS solutions, Lingoes & Borg (1983) have recently proposed a quasistatistical decision strategy that allows one to take various features of the situation into account. This paper adds another important piece of information to this approach: for the Lingoes-Borg decision criterion R, we compute what proportion of R-values is greater/less than the observed coefficient if one were to consider all possible alternative distance sets within certain bounds defined by the observed fit coefficients for two alternative MDS solutions, what are the limits of acceptability for such fit coefficients, and how are the observed MDS configurations interrelated.

1. Introduction

In the context of multidimensional scaling (MDS), in particular in its con-firmatory variety, one often faces the question whether two alternative repre-sentations should be considered 'equivalent'. E.g., given a set of proximity data, we could represent them optimally in some geometry via a standard procedure like SSA-I (Lingoes, 1973), and then again in the same geometry with additional external constraints on the distances by using CMDA (Borg & Lingoes, 1980). Or, in another context, two MDS solutions differing only in their dimensionality may be computed. In either case, one would first compare the respective fit coefficients. If they are both within the limits of acceptabil-ity, one proceeds to evaluate whether the more demanding mapping conditions of the more restrictive scaling approach had an appreciable effect.

Lingoes & Borg (1983a) describe two coefficients that are useful in such evaluations. Let y the vector of proximity data, and x and z the representing distances in two MDS representations, where z is generated under the more restrictive conditions. We then compute the product-moment correlation coef-ficients $\rho_1 = \rho_{xy}$, $\rho_2 = \rho_{zy}$, and $\rho_3 = \rho_{xz}$. (In ordinal MDS, due allowance is made for the weaker mapping constraints by substituting the original x, y, and z values with their ranks and by an appropriate coefficient of monotonic-

ity. These values are combined into the partial or conditional correlation coefficient:

$$\rho_c = \rho(xz \cdot y) = (\rho_3 - \rho\rho_2)/k_1 k_2,$$ [1.1]

$$k_i = \left(1 - \rho_i^2\right)^{1/2}.$$ [1.2]

A simple absolute benchmark for evaluating the size of ρ_c would be to check whether $\rho_c^2 > 0.5$. If so, the distances in the two MDS configurations share more variance, independent of their common data source y, than they do not due to the additional constraints on z. Naturally, as $\rho_c \rightarrow 1$, one should be more and more inclined to conclude that these additional constraints are essentially irrelevant.

But rather than evaluating the absolute size of ρ_c, one can compare it to another measure, $1 - \rho_3^2$, the decrement in common variance due to the imposition of the additional constraints onto z, i.e., in other words, the loss from satisfying the constraints. If $\rho_c^2 \gg 1 - \rho_3^2$, one should feel very confident about the essential equivalency of the solutions x and z. In the context of a decision approach, where the equivalency is the H_1, Lingoes & Borg (1983a) suggest that:

$$R = \rho_c/\left(1 - \rho_3^2\right)^{1/2}$$ [1.3]

should be greater than 3. For R-values less than 3, additional criteria may be considered. Seven such factors are discussed by Lingoes & Borg (1983a). E.g., if the number of points, n, is small, one should expect that additional constraints can generally be satisfied without moving the points around very much. Hence, one should require that R be greater for a small n than for a large n. So, e.g., with $w = 3$ for $n < 9$, $w = 2$ for $9 \leqslant n \leqslant 15$, and $w = 1$ for $n > 15$, R should be greater than w to accept H_1. Similar weights result from taking into account the sample size for the proximities, the dimensionality, the ratio of additional constraints on z relative to those on x, etc. Each of these conditions gives rise to a weight w of 1, 2, or 3, and then the average of various weights defines the final decision hurdle w.

The coefficient R is apparently purely descriptive and involves only the given vectors x, y, and z. However, in setting w criteria, one does actually go beyond the observed values and expresses an expectancy that, e.g., R should generally be greater when n decreases. Some bounds for such expectancies will now be derived.

2. Populations for ρ_c and R

With x, y, and z as given eliminate vectors, the three bivariate correlations ρ_1, ρ_2, and ρ_3, and, consequently, also ρ_c and R are fixed values. Yet, this does

not correspond to the situation in which the researcher finds him/herself in practice. Even if we assume that x and z are indeed optimal and unique representations for y – and so there would be no higher values for ρ_1 and ρ_2 –, it would be unreasonable to argue that the given x and z are the only acceptable solutions. Rather, if x and z are acceptable, then other solutions with somewhat lower fit coefficients ρ_1 and ρ_2 could certainly not be completely unacceptable representations of y either.

Assume now we define two lower bounds for ρ_1 and ρ_2 so that distance sets with fit values at least as great as these would be considered acceptably precise representations of y. Without any considerations as to what realizations of x and z are feasible in a particular distance context, it is clear that by going over all numerically feasible vectors x and z (i.e., all unit length vectors in case of interval MDS or all permissible permutations of n ranking numbers in case of ordinal MDS) one will obtain a certain proportion of conditional correlation values greater and less than the empirically observed value. Analogous to the lower bounds for the fit values of x and z, we could define a lower bound for the conditional correlations below which the question as to the equivalency of any two distance sets is automatically answered in the negative. Setting this bound to 0 would certainly be 'low enough'. So, for conditional correlations lying between 0, say, and 1, what proportion of them is greater than the observed ρ_c if all x and z with fit values exceeding certain minima are considered?

We can add another condition here, i.e., the requirement that only those x and z are taken into account whose difference in representational goodness is not greater than the observed difference $\rho_1 - \rho_2$ or some other fixed difference.

The various coefficients and constraints are best described by the diagram in Fig. 1 (suggested by a reviewer). The observed coefficients are denoted by Greek letters. The values resulting from considering all possible x and z are denoted by corresponding Roman letters. Fig. 1 then shows a region in r_1, r_2, r_3 space. The region ABCDEF consists of all admissible triples (r_1, r_2, r_3) which result as a consequence of the constraints:

$$0 \leqslant r_1 \leqslant \rho_1; \tag{2.1}$$

$$0 \leqslant r_2 \leqslant \rho_2; \tag{2.2}$$

$$0 \leqslant r_1 - r_2 \leqslant \rho_1 - \rho_2; \tag{2.3}$$

$$r_1 r_2 \leqslant r_1 r_2 + \left[\left(1 - r_1^2\right)\left(1 - r_2^2\right)\right]^{1/2}; \tag{2.4}$$

where the latter constraints follow from $0 \leqslant r_c \leqslant 1$. The surface ABC is the level surface for $r_c = 0$; the surface UVW is defined by the condition $r_c = \rho_c$; and the surface DEF is the set of all triples (r_1, r_2, r_3) satisfying $r_c = 1$.

What we are interested in is the proportion of the volume of the region
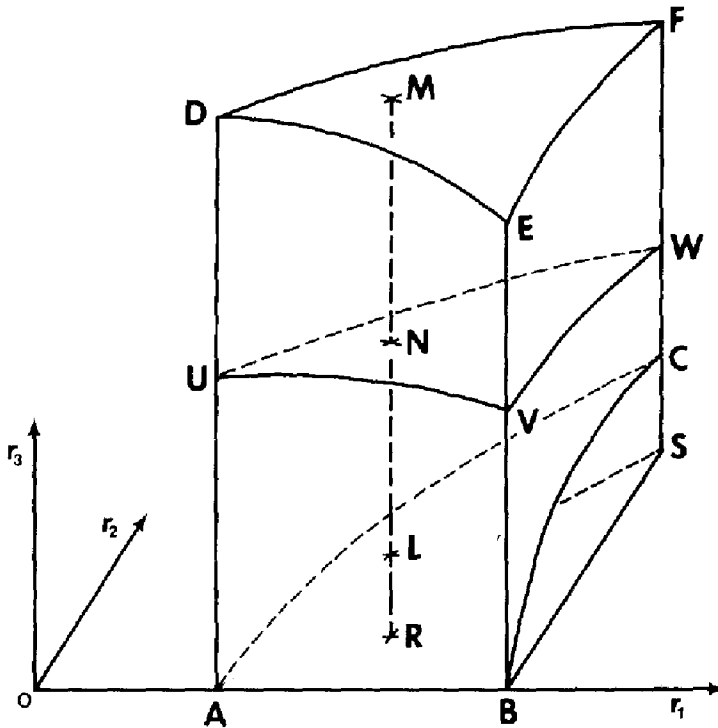
Fig. 1. Integration in (r_1, r_2, r_3)-space. Line AS is defined by $r_1 - r_2 = \rho_1\rho_2$. Point coordinates: $A = (\rho_1 - \rho_2, 0, 0)$, $B = (\rho_1, 0, 0)$, $S = (\rho_1, \rho_2, 0)$, $C = (\rho_1, \rho_2, \rho_1\rho_2)$, $W = (\rho_1, \rho_2, \rho_3)$, $M = (r_1, r_2, r_1r_2 + k_1k_2)$, $N = (r_1, r_2, r_3)$, $L = (r_1, r_2, r_1r_2)$, $R = (r_1, r_2, 0)$.

UVWDEF to the volume of the entire feasible region ABCDEF. This proportion, $P(r_c > \rho_c)$, can be computed by some method of numerical integration.

The diagram in Fig. 1 can also be used to illustrate the criterion R. In that case, the surface UVW defines the correlation triples with an R value equal to the one for the three observed correlations, i.e., R_e. The surface DEF is the level for $R = \infty$. The proportion $P(R > R_e)$ is derived by the same methods as for $P(r_c > \rho_c)$.

3. Some remarks on computing the proportions

Many procedures exist for computing the volume of three dimensional regions with curved surfaces. One fairly simple but relatively expensive approach would be to randomly choose the coordinates from a rectangular distribution of a large number of points lying within the feasible region and then check what proportion of them falls into the region UVWDEF. More and more points could be added until the proportions converged within reasonable bounds.

Another method is to define a sufficiently fine rectilinear three-dimensional lattice and count over all its points, whether or not $r_c > \rho_c$ and $R > R_e$, respectively. Lingoes & Borg (1983b) find that a relatively coarse lattice will yield sufficiently precise estimates of the desired proportions.

However, some results or short-cuts can be arrived at analytically. The most important one is $P(r_c > \rho_c)$, which can be found directly. Consider Fig. 1. Given some r_1 and r_2 and the observed value ρ_c, we can find the r_3-coordinate of the points L, N, and M by using formula [1.1]: $L = (r_1, r_2, r_1 r_2)$, $N = (r_1, r_2, r_1 r_2 + \rho_c k_1 k_2)$, and $M = (r_1, r_2, r_1 r_2 + k_1 k_2)$. Hence we obtain the distance $d(N, M) = k_1 k_2 (1 - \rho_c)$ and $d(L, M) = k_1 k_2$, and their ratio $d(N, M)/d(L, M) = 1 - \rho_c$. But this ratio is constant over the (r_1, r_2) region, and, thus, is just the proportion of the volumes of UVWDEF and ABCEDEF. So, $P(r_c > \rho_c) = 1 - \rho_c$.

No such direct solutions exists for R. However, given some r_1, r_2, and R_e, we can compute $d(N, M) = \{(a + b) - [a + b R_e (b^2 R_e^2 + 1 - a^2)^{1/2}/(b^2 R_c^2 + 1)]\}$, with $a = r_1 r_2$ and $b = k_1 k_2$, for $k_i = (1 - r_i^2)^{1/2}$ and hence the ratio $d(N, M)/d(R, M) = 1 - r_c$, which when integrated over all (r_1, r_2) pairs yields $P(R > R_e) = 1 - \bar{r}_c$ for a sufficiently fine grid in the feasible region ABS. A computer program, GUIDER which proceeds in this manner is described by Lingoes & Borg (1983b).

4. Evaluating the proportion $P(R > R_e)$

Having obtained the proportion $P(R > R_e)$ we now ask how is it to be evaluated. It seems unproblematic to say that if one finds a very small proportion $P(R > R_e)$ in a particular situation (e.g., $P(R > R_e)$ less than 0.05), then the observed R value, R_c, can at least be called "remarkably high", because within the constraints defined by our choice of $r_1(\min)$, $r_2(\min)$, and the maximal difference of r_1 and r_2, only a very small proportion of x and z would lead to higher conditional correlations.

Obviously, there is no need to limit our use of $P(R > R_e)$ as an efficacy coefficient. Rather, it could be embedded into a decision strategy if one felt that this would be useful. In that case one could proceed lexicographically and first test whether $\rho_c^2 > 0.5$, and, if so, whether $P(R < R_e) < 0.05$, say. If both decisions yield positive answers, then the equivalence hypothesis should be accepted. It should be noted, however, that our choice of $P < 0.05$ is merely illustrative of what might be considered 'low enough' and, depending upon the scientific context, some other choice might well be made.

It is of immediate concern, of course, to ask about the stability of the obtained proportions when the distance context is re-introduced. In the above, we required only that x and z satisfy certain numerical constraints. But if x and z have to be distances, and even Euclidean distances in a fixed dimensionality, then they have to satisfy more constraints. What effects would such additional constraints have? Starting with point W in Fig. 1 we see that it is

no doubt possible to arrive at distance sets x and z correlating with $r_c = \rho_c$ if the fit values r_1 and r_2 were to be lowered. However, to move from W towards F or S, i.e., to find distance sets with the same fit to the data y but with any r_c from $[\rho_c, 1]$ or $[0, \rho_c]$, respectively, is not possible in general. This is only possible at some cost for r_1 and/or r_2 (see, e.g., McGee, 1968). This means that we should consider the proportion of the regions UVWED to DEWBA – or an appropriate subregion with bounds $r_1(\min)$ and $r_2(\min)$ – rather than the one computed above. Yet, to determine the shape of the surfaces DEW and ABW would be very demanding indeed. In any case, we see what this problem implies: it is obvious that the resulting proportions $P(r_c > \rho_c)$ and $P(R > R_c)$ can only be smaller under the additional constraints; hence, the proportions computed by the methods above are upper bounds, so that a $P(R > R_e) < 0.05$, say, should be taken as a strong indicator of configurational equivalency.

5. Applications

We now apply the evaluation procedure to three sets of data that have been analyzed in different ways in the literature. Of course, only those data were selected where the test described in Section 1 does not immediately lead to an acceptance of the equivalency hypothesis. All data sets are presented and discussed in Borg (1981).

Glushko (1975) describes a set of $\binom{17}{2}$ data, y, on the similarity of different dot patterns vis-à-vis their 'patterns goodness". He presents unconstrained MDS solutions which showed roughly the predicted point groupings. The corresponding distances define the vector x. Borg & Lingoes (198) reanalyzed these data and enforced certain regional constraints consistent with the predicted groupings onto the point configurations. Their "weak contiguity test" (Lingoes, 1981) leads to the distance set z. We find $\rho(x, y) = 0.873$, $\rho(z, y) = 0.810$, $\rho(x, z) = 0.886$, and $\rho(xz \cdot y) = 0.626$. This yield $R_e = 1.352$. If we choose the constraints (2.1) + (2.4), $P(R > R_e)$ could be computed with one of the described methods. In practice, however, it turns out that it is usually sufficient to compute $P(R > R_e)$ for two simple versions of (2.3) only, i.e., for $r_1 - r_2 = \rho_1 - \rho_2 = D$ and for $r_1 - r_2 = 0$, and then $P(R > R_e)$ must lie between the values resulting under these conditions. We find $P(R > R_e \mid D) = 0.227$ and $P(R > R_e/0) = 0.230$, and thus conclude that there is some reason to question the equivalency of the two alternative MDS configurations.

Ekman (1954) presents a set of similarity data on 14 colors ranging from 434 mμ to 674 mμ. Using ordinal MDS, one obtains an almost perfect 2-dimensional point configuration that looks roughly like a circle (Shepard, 1962). One might ask whether it is possible to force the points onto a perfect circle without much loss (Borg & Lingoes, 1980). This question leads to MDS solutions x and z with coefficients $\rho(x, y) = 0.985$, $\rho(z, y) = 0.968$, $\rho(x, z) = 0.972$, $\rho(xz. y) = 0.429$, and $R_e = 1.824$. Using the GUIDER program as

before, we find $P(R > R_e | D) = 0.155$ and $P(R > R_e | 0) = 0.155$, which suggests that the two alternative representations differ more than might be expected from looking at their very similar fit values, i.e., approximately 16% of all possible configuration pairs would be found to be more similar than the pair under consideration.

Levelt et al. (1966) discussed the similarity of 15 tonial intervals. By metric MDS methods, they found that these data lead to a 'horseshoe-structure' in two dimensions. If three dimensions are used, this horseshoe exhibits three bending points on the third dimension. Shepard (1974) felt, however, that the data showed roughly a simplex gradient. Borg & Lingoes (1979) investigated the loss incurred from actually bending the horseshoe structure into a simplicial manifold. This yields $\rho(xy) = 0.911$, $\rho(zy) = 0.866$, $\rho(xz) = 0.927$, $\rho(xz \cdot y) = 0.671$, and $R_e = 1.795$. Proceeding as before, we find $P(R > R_e | D) = 0.154$ and $P(R > R_e | 0) = 0.156$, and so Shepard's proposition should probably be rejected.

6. Discussion

In the applications, it was conspicuous that $P(R > R_e | D) \approx P(R > R_e | 0)$. This relation of being almost equal was found in some three dozen empirical data sets that were investigated. Only when the difference in fit values, $\rho_1 - \rho_2$, becomes very large (> 0.5), then the two proportions tend to differ as ρ_c increases. However, the proportions are always ordered as $P(R > R_e | \rho_1 - \rho_2) < P(R < R_e | 0)$. So, $P(R < R_e | 0)$ can be considered an upper bound for $P(R < R_e | D)$. It seems more realistic though to go even further: since "large" differences in the fit values of x and z would prevent one from asking the equivalency question in the first place, one can conclude that $d(N, M)/d(R, M)$ remains almost equal over different choices of lines RM in the feasible region of Fig. 1. But then we could simply use $P(R > R_e | 0)$ as an estimate for $P(R > R_e)$, which can be computed directly, at a computational savings of roughly one half.

Apart from considerable savings in computation time, the approximate invariance of the $d(N, M)/d(R, M)$ ratio over different lines leads to another benefit. We started out be noting that if the observed fit values ρ_1 and ρ_2 were deemed "acceptably high", then somewhat lower fit values should also be acceptable. However, it is incompatible with this argumentation – at least if one does not adhere to a scientifically questionable accept/reject decision logic – to set fixed lower bounds $r_1(\min)$ and $r_2(\min)$ below which the fit values become unacceptable. Rather, the lower these fit values, the less acceptable they get. The decrement in acceptability is thus continuous. Yet, it seems impossible to characterize this function except, perhaps, by such general statements that it is positively accelerated. But this is, of course, not sufficient to incorporate it into the procedure to compute efficacy coefficients on the equivalency of x and z. However, if $d(N, M)/d(R, M)$ is approximately

constant over all lines RM, then it is not relevant how $r_1(\min)$ and $r_2(\min)$ are defined – just as $P(r_c > \rho_c) = 1 - \rho_c =$ constant, whatever $r_1(\min)$ and $r_2(\min)$.

Finally, we should point out that evaluating $P(R > R_e)$ will lead to the same conclusions whether: (a) the numerical sets that could have resulted in the various r's have much or no error, or (b) the N on which the r's could have been based is large or small, or (c) the true nature of the univariate, bivariate, or trivariate distributions for the quantities is normal or otherwise, or (d) the r's were computed on continuous or discrete vectors. The method is independently efficacious without the usual statistical considerations. We do not imply, of course, that such factors are unimportant or scientifically irrelevant, but only that for the restricted interpretations that are to be made with this model, knowledge of such factors is not helpful nor, in any way, determinative of the conclusions reached.

References

Borg, I. Anwendungsorientierte Multidimensionale Skalierung. Heidelberg, Germany: Springer, 1981.

Borg, I. & Lingoes, J.C. Multidimensional scaling with external constraints on the distances. In J.C. Lingoes, E.E. Roskam, & I. Borg (eds.) Geometric Representations of Relational Data. Ann Arbor, MI: Mathesis Press, 1979, 2nd edition.

Borg, I. & Lingoes, J.C. A model and algorithm for multidimensional scaling with external constraints on the distances. Psychometrika, 1980, 45, 25–38.

Borg, I. & Lingoes, J.C. An alternative approach to confirmatory inference and geometric models. Quality and Quantity, 1981, 15, 541–552.

Ekman, G. Dimensions of color vision. Journal of Psychology, 1954, 38, 467–474.

Levelt, W.J.M., van de Geer, J.P., & Plomp, R. Triadic comparisons of tonal intervals. British Journal of Mathematical and Statistical Psychology, 1966, 19, 163–179.

Lingoes, J.C. The Guttman-Lingoes Nonmetric Program Series. Ann Arbor, MI: Mathesis Press, 1973.

Lingoes, J.C. Testing regional hypotheses in multidimensional scaling. In I. Borg (ed.) Multidimensional Data Representations: When and Why. Ann Arbor, MI: Mathesis Press, 1981.

Lingoes, J.C. & Borg, I. A quasi-statistical model for choosing between alternative configurations derived from ordinally constrained data. British Journal of Mathematical and Statistical Psychology, 1983a, 36, 36–53.

Lingoes, J.C. & Borg, I. A mathematical decision model for the Lingoes-Borg criteria in testing the equivalence hypothesis. University of Michigan Computing Technical Report, 1983b, 1, 1–20.

McGee, V. Multidimensional scaling of N sets of similarity measures: A nonmetric individual differences approach. Multivariate Behavioral Research, 1968, 3, 233–248.

Shepard, R.N. Analysis of proximities: Multidimensional scaling with an unknown distance function - II. Psychometrika, 1962, 27, 219–246.

Shepard, R.N. Representations of structure in similarity data: Problems and Prospects. Psychometrika, 1974, 39, 373–421.