

Matrix and Stimulus Sample Sizes in the Weighted MDS Model: Empirical Metric Recovery Functions

Joseph Lee Rodgers
University of Oklahoma

The only guidelines for sample size that exist in the multidimensional scaling (MDS) literature are a set of heuristic “rules-of-thumb” that have failed to live up to Young’s (1970) goal of finding functional relationships between sample size and metric recovery. This paper develops answers to two important sample-size questions in nonmetric weighted MDS settings, both of which are extensions of work reported in MacCallum and Cornelius (1977): (1) are the sample size requirements for number of stimuli and number of matrices compensatory? and (2) what type of functional relationships exist between the number of matrices and metric recovery? The graphs developed to answer the second question illustrate how such functional relationships can be defined empirically in a wide range of MDS and other complicated nonlinear models.

Index terms: metric recovery, monte carlo study, multidimensional scaling, sample size, weighted multidimensional scaling.

A research problem of considerable practical interest is the choice of sample size. In settings based on parametric statistical models (e.g., ANOVA), the relationship between sample size and power has been carefully studied and charted (e.g., Pearson & Hartley, 1951). The present research was concerned with the relationship between sample size and recovery of metric information in the nonmetric weighted multidimensional scaling (WMDS) model.

Nonmetric MDS is a nonparametric method used to assign scale values to stimuli. Raw data are measures of relationship or proximity (e.g., similarities, confusions, transitions) between stimuli; the Euclidean distance model is used to

define scale values. MDS is not usually used for hypothesis testing (but see Ramsey, 1977). Rather, its most common application is to extract interpretable quantitative dimensions from qualitative data. In this sense, MDS is more similar to factor analysis than to ANOVA or related parametric techniques.

However, selecting the proper number of matrices and stimuli is an important MDS consideration for several reasons. First, the ability of a nonmetric procedure to recover quantitative information from qualitative data can depend on the number of matrices and stimuli used. Other factors related to sample size include stability and external validity of scale values, interpretability of the optimization function, and the cost of obtaining MDS data. To psychometricians, the first reason—metric recovery—is probably the most important. In fact, recovery in MDS research is a near counterpart to power in a traditional parametric test. Power measures the probability of correctly detecting an existing alternative hypothesis; metric recovery measures the ability of an MDS model to correctly detect the quantitative information in qualitative data. High metric recovery will lead naturally to some of the other criteria above, particularly to stability and goodness of fit.

Several monte carlo studies have been concerned with the problem of matrix (frequently referred to as “subject” in the MDS literature) and stimulus sample size in the WMDS model. However, unresolved or contradictory issues still exist. Two questions were addressed in the current study. First, are matrix and stimulus sample sizes compensatory in the WMDS model?

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 15, No. 1, March 1991, pp. 71-77

© Copyright 1991 Applied Psychological Measurement Inc.
0146-6216/91/010071-07\$1.60

Second, what is the functional relationship between matrix sample size and metric recovery in the WMDS model? Research is reviewed that has addressed the issue of sample size within the WMDS model. This includes a reanalysis of data from MacCallum and Cornelius (1977) to resolve the issue of whether matrix and stimulus sample sizes compensate for each other. Then monte carlo results are presented that extend the results of MacCallum and Cornelius; the design of the present study permitted graphical portrayal of the functional relationship between matrix sample size and several measures of metric recovery.

Previous Research

Shiffman, Reynolds, and Young (1981) distinguished between classical MDS (CMDS), replicated MDS (RMDS), and WMDS. The CMDS model assumes a single $I \times I$ square symmetric matrix of proximities between all pairs of stimuli (one-mode, two-way data; see Carroll & Arabie, 1980). Output is a single space with I stimuli in K dimensions. The RMDS model assumes multiple $I \times I$ proximity matrices (two-mode, three-way data), but produces a single space representing the stimuli (as in CMDS). The WMDS model requires the same input data as RMDS (two-mode, three-way), and also produces a stimulus space like CMDS and RMDS. In addition, the separate matrices (e.g., persons, companies, points in time) are assigned scale values in a second space (usually called the "subject space" in the MDS literature; here called the "matrix space") indicating how they differentially weight the axes of the stimulus space.

Early MDS sample size research studied stimulus sample size and metric recovery. Young (1970) considered the relationship between number of stimuli and recovery in nonmetric CMDS settings, and concluded that "it is, apparently, still too early to determine just what the functional relation is" (p. 472). Once stated, the goal of finding such functional relationships has never been met; instead, the literature contains suggestions and heuristic "rules-of-thumb" (e.g., Davison, 1983; Kruskal & Wish, 1978; Spence & Domoney,

1974; Young, 1970). The only theoretically based criterion to be carefully studied was Shepard's (1966) "degrees-of-freedom ratio" (e.g., Spence & Domoney, 1974; Young, 1970). However, no general organizing principles concerning sample size have resulted from these considerations.

Davison (1983, pp. 41, 48) discussed incomplete designs, in which multiple (N) respondents use different subsets of the $I(I-1)/2$ possible paired comparisons, and then data are averaged for every pair before applying the CMDS model. He studied patterns in the monte carlo literature (e.g., Spence & Domoney, 1974), and developed a formula that insures at least 20 data points for each estimated parameter: Select N matrices and I stimuli so that $N > 40K/(I-1)$.

The Davison formula implies a trade off between the number of matrices and stimuli. Multiplying through by $(I-1)$ gives the inequality $N(I-1) > 40K$. Clearly, the formula treats number of matrices and stimuli as compensatory for a given dimensionality (i.e., increasing the number of matrices reduces the need for stimuli, and vice versa). This relationship obviously must be restricted to settings with more than a few stimuli, because a scaling study with fewer than five stimuli, for example, is not very sensible (e.g., Shepard, 1966). Davison's (1983) formula was based on intuition obtained from inspection of previous studies.

MacCallum (1979) presented simulation results from nonmetric WMDS settings. In his study, number of stimuli was fixed at $I = 30$, and number of matrices was $N = 10$ or $N = 20$. Further, error level and amount of missing data were also manipulated. Results suggested that quality of recovery decreased with increasing error and increasing level of missing data, but was relatively insensitive to matrix sample size. Because stimulus size was fixed, neither the number of stimuli nor the compensatory relationship between I and N could be assessed.

MacCallum and Cornelius (1977) varied both number of stimuli (I) and number of matrices (N), as well as error level and dimensionality. They defined four different measures of recovery,

and complete three-way data were used. Their findings replicated previous two-way results (e.g., Isaac & Poor, 1974; Young, 1970); true overall recovery improved with increasing number of stimuli, as did recovery of stimulus and matrix spaces. The STRESS measure used in this study (SSTRESS from ALSICAL), a measure of *apparent* recovery derived from the data (but which contains no information about the *true* underlying structure of the data), actually gave misleading information and suggested a better fit of the model for decreasing numbers of stimuli. (Sherman, 1972, used a different routine to generate data, in which each configuration was truly independent, and found that STRESS was a fairly accurate index of recovery for many stimuli and low levels of error.) Despite this one anomaly, MacCallum and Cornelius (1977) suggested that recovery of true structure increases monotonically with increasing number of stimuli in both two- and three-way settings.

The interaction between matrix and stimulus sample size was assessed in MacCallum and Cornelius (1977). Number of matrices and number of stimuli showed a significant interaction for two of the four dependent variables measuring recovery—recovery of the true stimulus configuration, and recovery of the true matrix weights. No interaction was observed for the two most important measures of metric recovery, SSTRESS and *M* (the index of overall true metric recovery). An interaction does not necessarily imply that the number of matrices and stimuli compensate for each other in producing quality of recovery (e.g., the effect of one might be increasingly enhanced by increases in the other). However, the absence of an interaction shows definitively that such a compensatory relationship does not exist. MacCallum and Cornelius (1977) did not plot cell mean patterns for the two dependent variables for which the interaction occurred, which leaves doubts about the type of interaction that occurred. This issue was resolved through a reanalysis of their original data.

R. C. MacCallum (personal communication, 1988) provided the cell means from MacCallum

and Cornelius (1977). Plots of these means (available from the author) showed clearly the absence of any type of compensatory relationship, even for the two indices showing significant interactions. All four measures were sensitive to changes in number of stimuli, and virtually insensitive to changes in number of matrices. The interactions were caused by small, nonsystematic crossovers in otherwise very consistent patterns. The implication of this reanalysis is that the number of matrices and number of stimuli are not compensatory in three-way WMDS settings. Rather, selecting these sample sizes should be treated as separate problems.

Next, matrix sample size was considered. When MacCallum and Cornelius (1977) tested the main effect for number of matrices, the relationship of recovery to *N* was the same as in MacCallum (1979, p. 421): “the number of individuals, within the range studied, did not have a substantial effect on any of the dependent variables.”

The levels of number of matrices in MacCallum and Cornelius (1977) were 15, 25, 35, and 50, and the authors concluded that “empirical investigators . . . need make no effort to obtain such (large) samples.” The authors, however, did not resolve the point at which recovery is degraded by small sample size (if, indeed, such a point occurs at all), or provide any indications of the functional relationship between *N* and metric recovery.

The early goal that Young (1970) established has still not been met. Sample size recommendations found in the literature for all of the MDS models are qualitative statements that are heavily qualified. Functional quantitative relationships cannot be found. However, they can be specified empirically. For example, MacCallum and Cornelius (1977) stopped just short of such a specification. They could have plotted their four measures of recovery against the sample sizes they investigated ($N = 15, 25, 35, \text{ and } 50$), but their functions would have been flat, because their work showed that asymptotic recovery is achieved at levels below their smallest sample size. The

range of the function that is not flat is the range of particular interest to applied researchers. The work of MacCallum and Cornelius (1977) was extended in the present study to specify the quantitative relationship between sample sizes and metric recovery in WMDS models.

Method

This study replicated MacCallum and Cornelius (1977), with adjustments reflecting methods in Rodgers (1985). Three-way data were generated and submitted to PROC ALSCAL in SAS (SAS, 1983). Stimulus size was fixed at 12, and error level at moderate, including both medium range levels from the MacCallum and Cornelius study. Nonmetric analyses were performed using the INDSCAL option (for WMDS) in ALSCAL.

The 12 stimuli were randomly sampled from the 50 real data points presented in Girard and Cliff (1976) and used in MacCallum and Cornelius (1977). These were four-dimensional data; for a given dimensionality K , the first K dimensions were used. The design varied number of matrices across eight levels ($N = 2, 4, 6, 8, 10, 12, 15, \text{ and } 25$), which included two sample sizes studied by MacCallum and Cornelius.

Dimensionality was varied across three levels ($K = 2, 3, \text{ and } 4$). The matrix sample was randomly selected from 200 matrices constructed before the simulation was run. Matrix weight vectors for these 200 simulated individuals were constructed to have unit length in a specified dimensionality (thus, three different matrix sets were generated), and to be Von Mises distributed with a mean of 45° and a concentration parameter of 8 (equivalent to a standard deviation of 20°). The Von Mises distribution gives the closest counterpart on the unit circle to the normal distribution on the number line (Mardia, 1972). When matrix vectors were occasionally generated outside of the first quadrant, which would have produced negative matrix weights and imaginary distances, they were set back to the boundary. This generation procedure was tested and presented in Rodgers (1985).

Once simulated individuals and stimuli were

randomly selected, a matrix of true distances was generated for each individual. Next, error generation routines presented in MacCallum and Cornelius (1977) were used. Their random error generation routine was adapted from Girard and Cliff (1976), and involved a hyperbolic tangent transformation of the distances with random normal error added. This transformation achieves the goal of making underestimates more likely at the upper extreme of the scale, and overestimates more likely at the lower extreme, which are properties characteristic of actual empirical proximity data.

Within each of the 24 cells of the 8×3 design, 15 replications were run (compared to five in MacCallum & Cornelius, 1977). Because the present study was concerned with observing functional relationships rather than testing for cell differences, the added stability in the results appeared to be worth the cost of obtaining additional data.

The same four dependent variables used by MacCallum and Cornelius (1977)—SSTRESS, M , delta, and phi—were computed for each replication (algebraic expressions for each index may be found in their article). SSTRESS—squared STRESS, the optimization criterion in ALSCAL—was used as an index of recovery of distorted distances. This is the only goodness-of-fit index available in actual research. In monte carlo studies, however, the true structure of the data is known, and this structure was used to construct the additional three dependent variables. M is a measure of metric recovery—the recovery of the true distances for each matrix; it is the normalized sum of squared differences between the derived and true distances. M was the index used in Young's (1970) original study investigating recovery of metric information in nonmetric CMDS. Delta is an index of recovery of the true stimulus configuration. MacCallum and Cornelius (1977, p. 410) describe delta as the "square root of the mean squared differences between the true and recovered projections." Finally, phi is the index of recovery of the true matrix weights. This index was computed by rearranging the derived

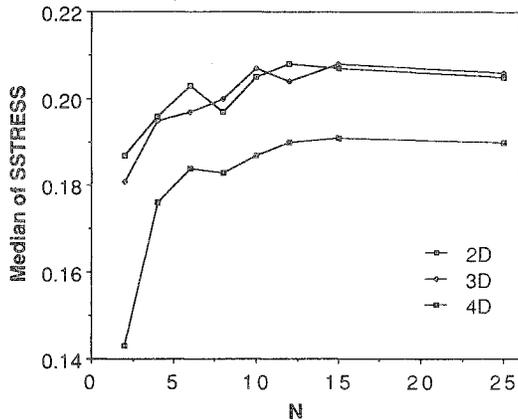
dimensions to most closely match the original dimensions, and is computed as 1 minus the average cosine of the angles between the original and derived matrix vectors.

Results

The median values across the 15 replications in each cell are presented and discussed separately for the four dependent variables in Figures 1 through 4. Of particular interest in each case is the point of asymptote, beyond which increasing the matrix sample size does not increase recovery.

Figure 1

Empirical Relationship Between WMDS Matrix Sample Size and SSTRESS in ALSICAL for Two, Three, and Four Dimensions



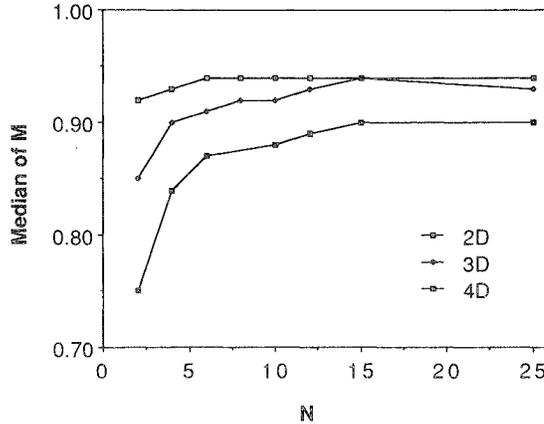
For SSTRESS (Figure 1), higher values imply lower recovery of the distorted distances. The apparent asymptote for each dimensionality occurs somewhere around 6 matrices. As expected, recovery improves for higher dimensionality, although two and three dimensions are difficult to distinguish.

For M (Figure 2), higher values imply better recovery of the true distances. There is little sensitivity to matrix sample size for two dimensions. The asymptote for three and four dimensions occurs around $N = 6$. Ability to recover the true distances decreases with increasing dimensionality.

For δ (Figure 3), higher values imply lower

Figure 2

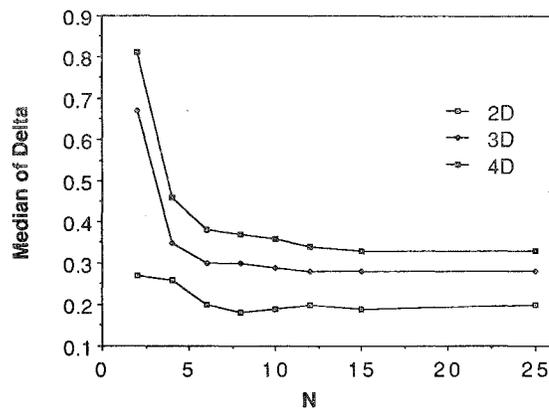
Empirical Relationship Between WMDS Matrix Sample Size and Metric Recovery (M) in ALSICAL for Two, Three, and Four Dimensions



recovery of the true stimulus space ($\delta = 0$ implies perfect recovery). For $K = 2$, there is little sensitivity to increasing N . For $K = 3$ and 4, there appears to be an asymptote, again around 6. Recovery is poorer with increasing dimensionality.

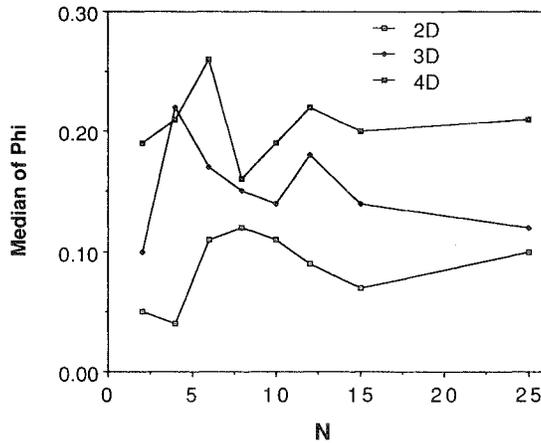
Figure 3

Empirical Relationship Between WMDS Matrix Sample Size and Stimulus Space Recovery (Δ) in ALSICAL for Two, Three, and Four Dimensions



Finally, for ϕ (Figure 4), higher values imply lower recovery. Both within and collapsing across dimensionality, there appears to be little relationship of ϕ to sample size, even for very small

Figure 4
 Empirical Relationship Between WMDS Matrix Sample Size and Matrix Space Recovery (Phi) in ALSICAL for Two, Three, and Four Dimensions



samples. Again, recovery is degraded for increasing dimensionality.

Discussion

Results of this monte carlo study are consistent with those found by MacCallum and Cornelius (1977). SSTRESS is the only index for which recovery improved with dimensionality, and there was virtually no sensitivity to the difference between sample sizes of 15 and 25. The present study extended these results to the range below $N = 15$.

For three of the four indices (those reflecting characteristics of the stimulus space), a sample size of approximately 6 was the point below which recovery dropped off. This drop-off was greater for four dimensions than for three, and somewhat greater for three than for two. Thus, if a two-dimensional solution is desired, a sample as small as 1 or 2 can provide high-quality recovery. For three or four dimensions, samples larger than 6 provide little improvement.

The insensitivity of phi to number of matrices is quite interpretable. Phi measures recovery of the matrix weights. Because these are recovered at the *individual matrix* level, however, recovering many matrix weights should not be different from recovering only a few.

On the other hand, the insensitivity of the other indices to matrix sample size is surprising. Clearly, the nonmetric information from only a very few matrices is sufficient to recover virtually all of the metric information in the original distances, the distorted distances, and the original stimulus configuration.

There are other considerations that may influence choice of matrix sample size in addition to metric recovery. For example, Rodgers (1985) showed that the power in statistical tests of group differences in WMDS weight spaces depends heavily on sample size (as would be expected for any parametric test). Also, neither the present nor the MacCallum and Cornelius (1977) study directly investigated the ability of WMDS to "tie down" the dimensions (e.g., Shiffman et al., 1981, p. 67), the WMDS feature in which interpretable directions tend to lie on the coordinate axes. Because this is an attractive property of the WMDS model, future research might investigate its relationship to matrix sample size. Also, because the rotational invariance of WMDS is related to the variability of the matrix weight vectors, this property might be more affected by matrix sample size than the measures of metric recovery from the present study.

Graphs similar to those in Figures 1 through 4 could be developed for CMDS and RMDS settings (as well as for unfolding and more exotic models), for different levels and types of error, and for other measures of recovery (e.g., STRESS or STRAIN). This approach also provides a way of formally comparing the recovery of metric information in different MDS models. For example, RMDS and WMDS can be applied to exactly the same type of data. WMDS produces an extra space and achieves some level of rotational invariance, but what does RMDS do with the "extra" information? Is the stimulus space it produces more metrically perfect than that from WMDS?

For years, researchers have been using the Pearson and Hartley (1951) power charts. These came from analytic procedures derived from parametric assumptions. MDS researchers have

not been successful in developing comparable methods. However, monte carlo procedures provide a way to empirically derive equivalent relationships in nonlinear settings like those in MDS models.

References

- Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology*. Palo Alto CA: Annual Reviews, Inc.
- Davison, M. L. (1983). *Multidimensional scaling*. New York: Wiley.
- Girard, R. A., & Cliff, N. (1976). A monte carlo evaluation of interactive multidimensional scaling. *Psychometrika*, *41*, 43–64.
- Isaac, P. D., & Poor, D. D. (1974). On the determination of appropriate dimensionality in data with error. *Psychometrika*, *39*, 91–109.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills CA: Sage.
- MacCallum, R. C. (1979). Recovery of structure in incomplete data by ALSCAL. *Psychometrika*, *44*, 69–74.
- MacCallum, R. C., & Cornelius, E. T. (1977). A monte carlo investigation of recovery of structure by ALSCAL. *Psychometrika*, *42*, 401–428.
- Mardia, K. V. (1972). *Statistics of directional data*. New York: Academic Press.
- Pearson, E. S., & Hartley, H. O. (1951). Charts of the power function for analysis of variance tests, derived from the non-central F distribution. *Biometrika*, *38*, 112–130.
- Ramsey, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, *42*, 241–266.
- Rodgers, J. L. (1985). Statistical tests of group differences in ALSCAL-derived subject weights: Some monte carlo results. *Applied Psychological Measurement*, *9*, 241–248.
- SAS Institute, Inc. (1983). *SUGI supplemental library user's guide*. Cary NC: Author.
- Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*, 287–315.
- Sherman, C. R. (1972). Nonmetric multidimensional scaling: A monte carlo study of the basic parameters. *Psychometrika*, *37*, 323–355.
- Shiffman, S. S., Reynolds, M. A., & Young, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press.
- Spence, I., & Domoney, D. W. (1974). Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika*, *39*, 469–490.
- Young, F. W. (1970). Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, *35*, 455–473.

Acknowledgments

The author expresses his appreciation to Robert C. MacCallum, who provided the data from MacCallum and Cornelius (1977) for reanalysis, and who provided suggestions that improved this study. Part of the preparation of this article was completed while the author was on sabbatical leave in the Psychology Department at Ohio State University in 1987/88. An earlier version was presented at the annual meeting of the Psychometric Society in Toronto in 1986.

Author's Address

Send requests for reprints or further information to Joseph Lee Rodgers, Department of Psychology, University of Oklahoma, Norman OK 73019, U.S.A.