An APL/360 program for interactive monotone distance analysis*

JAMES E. DANNEMILLER
Survey Research Office

and

DAVID JOHN GOW
Department of Political Science
University of Hawaii, Honolulu, Hawaii 96822

Since Shepard's (1962) innovative development in programming an algorithm for monotone distance analysis, a series of algorithms and programs have been developed for this task. The model specifying the representation of variables in a multidimensional space such that the interpoint distances form a monotone function with the values of the input matrix has been variously denoted as proximities analysis (Shepard, 1962), nonmetric multidimensional scaling (Kruskal, 1964), smallest space analysis (Guttman, 1968), and a submodel of conjoint analysis (Young, 1972). We choose to follow Lingoes's (1972) suggestion that the model may more informatively be known as monotone distance analysis (MDA), thereby reflecting important aspects about the reproduction function between the input matrix of similarities and the derived representational space. Regardless of the variance in nomenclature, all of the previous programs have been coded in FORTRAN and generally designed for running under a batch processing system.

APL/360, on the other hand, is an interactive language for the IBM 360 series and has been implemented with varying degrees of success on other machines. At the present time, all of the currently available MDA programs utilize iterative techniques to minimize a function reflecting departure from the monotonicity constraint. Therefore, the APL environment within which APLIMDA runs allows the user to maintain a high degree of supervision over the execution of his program by monitoring and modifying the characteristics of the algorithm for the data set at hand. In addition, the user can rapidly evaluate the consequences of the choice of alternative coefficients of similarity as well as the effects of the addition or deletion of variables from the matrix of similarities to be input for analysis.

Algorithm. APLIMDA employs the algorithmic sequences that Monte Carlo studies and numerical experimentation by a number of researchers (Spence, 1970; Lingoes & Roskam, 1973) suggest are optimal in terms of speed of convergence and robustness with respect to local minima, as well as further options that a researcher may have some felt need to consider. Commencing with the input of a correlation matrix, R, the initial configuration can be based on the principal axes decomposition of Torgerson's (1958, p. 256) B* matrix of scalar products formed over the numerical values of R. Optionally, the user may select the principal axes of Torgerson's B* matrix formed on the rank order of the elements of R; the Guttman-Lingoes initial configuration (Lingoes & Roskam, 1973); or alternatively, the principal components of R.

From the initial configuration, Guttman-Lingoes's double-phase iterations (Guttman, 1968) based on rank-images guide the configuration of points to a position where the loss function is, in most circumstances, close to the overall or global minimum. After the termination of the double-phase procedure, single-phase iterations also based on rank images tend to bring about convergence at the global minimum. The pivotal equation (cast into matrix notation) of the APLIMDA iterative process is based on the Guttman-Lingoes SSA-I algorithm (Guttman, 1968):

$$X^{(t+1)} = \frac{1}{n} C^{(t)} X^{(t)} \qquad (1)$$

where n is the number of variables, X is the matrix of coordinates, C is the correction matrix, and the superscripts are iteration indices.

The elements of the correction matrix, C, in Eq. 1 are formed through

$$c_{ij} = \begin{cases} 1 + \sum_{k=1}^{n} d^*_{ik}/d_{ik} & (i = j) \\ 1 - d^*_{ij}/d_{ij} & (i \neq j) \end{cases} \qquad (2)$$

where d^*_{ij} is the rank image (Guttman, 1968) of the Euclidean distance, d_{ij}.

Commencing with the initial configuration of the user's choice as the first estimate of the matrix of coordinates, X, the iterative sequences indicated above are utilized. There is no guarantee that the algorithm always avoids local minima traps; however, as one step toward alleviating the problem, we provide the user with the choice of four nonarbitrary initial configurations (two are metric; two are nonmetric) from which to commence the iterative process.

By utilizing the Guttman-Lingoes initial configuration, APLIMDA is in most respects identical to the SSA-I algorithm. The only significant difference is that we have set the number of inner iterations in the

*Funds for this project were provided by the Survey Research Office, University of Hawaii. The research was undertaken while the second author was a grantee, The East-West Center, Honolulu, Hawaii.

double phase to t = 3, whereas in the original SSA-I program t = 10, and in subsequent programs Lingoes (1973) set t = 5. We have found (on the basis of brief informal experimentation) that t = 3 improves speed and efficiency at the cost of only a small number of additional double-phase iterations.

Restrictions. The most meaningful restriction on APLIMDA is the response time (not the CPU time) that the user is willing to accept. Unless the IBM 360 user is in a position to use a privileged terminal (or devise a method for inducing this state), the user will often have to wait 60 sec for each iteration with a 25-order matrix on the IBM 360/65. However, since the program is interactive, the user is in a position to quickly determine (with very few iterations) the most appropriate dimensionality for the data set presently under analysis.

Program Availability. Program listings of APLIMDA are available from James E. Dannemiller, Acting Director, Survey Research Office, 1914 University Avenue, The University of Hawaii, Honolulu, Hawaii 96822. With the listing (which is free of charge), we provide a trial matrix together with the correct output.

REFERENCES

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 1968, 33, 469-506.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Lingoes, J. C. A general survey of the Guttman-Lingoes nonmetric program series. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences.* Vol. I. New York: Seminar Press, 1972. Pp. 49-68.

Lingoes, J. C. *The Guttman-Lingoes nonmetric program series.* Ann Arbor: Mathesis Press, 1973.

Lingoes, J. C., & Roskam, E. A mathematical and empirical study of two multidimensional scaling algorithms. Psychometric Monographs, 1974, in press.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-139.

Spence, I. Multidimensional scaling: An empirical and theoretical investigation. Unpublished PhD thesis, University of Toronto, 1970.

Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

Young, F. W. A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences.* Vol. I. New York: Seminar Press, 1972. Pp. 69-104.

Probabilities for Spearman rank correlation coefficients*

JERROLD H. ZAR

Department of Biological Sciences
Northern Illinois University, DeKalb, Illinois 60115

Spearman's rank correlation procedure (Spearman, 1904; Zar, 1974, pp. 243-244) is one of the oldest and best known of nonparametric methods. This program may be used alone or employed with a rank correlation program. It provides the one-tailed and two-tailed probabilities of given Spearman rank correlation coefficients.

Input. Any number of Spearman rank correlation coefficients, r_s, and their associated sample sizes.

Output. The probability (both one-tailed and two-tailed) of a Spearman rank correlation coefficient at least as large as each coefficient submitted.

This program is especially useful as a subroutine in conjunction with programs (e.g., Zar, 1970) for Spearman rank correlation analysis.

Method. For the Spearman rank correlation coefficient,

$$r_s = 1 - 6\,\Sigma d^2/(n^3 - n),$$

the value of Σd^2 is computed. Then the cumulative probability for all possible Σd^2 up to and including the computed Σd^2 is determined by the Pearson Type II curve approximation described by Olds (1938, 1949). This approximation is very accurate for moderate and large sample sizes and is much more accurate than employing Student's t for assessing the significance of the rank correlation coefficient (Zar, 1972). When dealing with small sample sizes, however (say, less than 10), it might be wiser to refer to tables of the exact distribution of Σd^2 (Owen, 1962, pp. 400-406; de Jonge & von Montfort, 1972; Otten, 1973) or to tables of critical values of r_s (Zar, 1974, pp. 498-499).

Program Language and Computer. This program is written in FORTRAN IV and was run on the Northern Illinois University IBM 360/67 system, utilizing 82K bytes of core for compilation.

Availability. A source program listing and documentation may be obtained without cost from: Jerrold H. Zar, Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115.

REFERENCES

de Jonge, C., & Van Montfort, M. A. J. The null distribution of Spearman's S when n = 12. Statistica Neerlandica, 1972, 26, 15-17.

Olds, E. G. Distributions of sums of squares of rank differences for small numbers of individuals. Annals of Mathematical Statistics, 1938, 9, 133-148.

Olds, E. G. The 5% significance levels for sums of squares of rank differences and a correction. Annals of Mathematical Statistics, 1949, 20, 117-118.

Otten, A. The null distribution of Spearman's S when n = 13(1)16. Statistica Neerlandica, 1973, 27, 19-20.

Owen, D. B. *Handbook of statistical tables.* Reading, Massachusetts: Addison-Wesley, 1962.

Spearman, C. The proof and measurement of association between two things. American Journal of Psychology, 1904, 15, 72-101.

Zar, J. H. FORTRAN IV programs for rank correlation. Behavioral Science, 1970, 15, 463.

Zar, J. H. Significance testing of the Spearman rank correlation coefficient. Journal of the American Statistical Association, 1972, 67, 578-580.

Zar, J. H. Biostatistical analysis. Englewood Cliffs, New Jersey: Prentice-Hall, 1974.

*Contribution No. 492 from the Department of Biological Sciences, Northern Illinois University.