Title

Coscal: Program for Metric Multidimensional Scaling

Author

Cooper, Lee

Publication Date

1972-01-05

COSCAL: Program for Metric Multidimensional Scaling, *Lee G. Cooper, University of California, Los Angeles* [CAMR 17].

DESCRIPTION: COSCAL provides a metric multidimensional scaling solution. It determines, for a specified number of dimensions, the coordinates of the stimuli on the axes of a Euclidean space. There is an assumed linear relationship between the original comparative interpoint distance, h_{jk}, between a pair of stimuli, jk, and the final absolute interpoint distance, d_{jk}:

$$(1) \qquad h_{jk} = d_{jk} + c + \delta_{jk}, \qquad j, k = 1, 2, \cdots, n, j \neq k$$

where δ_{jk} is an error term, and:

$$(2) \qquad d_{jk} = \left[\sum_{m}^{t} (a_{jm} - a_{km})^2 \right]^{\frac{1}{2}}$$

where a_{jm} is the coordinate of the jth stimulus of the mth axis of a Euclidean space. This is the standard Euclidean distance function. The solution is calculated so as to minimize the sum of squares of error:

$$(3) \qquad G = \tfrac{1}{2} \sum_{k}^{n} \sum_{j \neq k}^{n} \delta_{jk}^2 = \text{minimum.}$$

This is an iterative procedure which requires some initial guess at the values of the a_{jm} to start the routine. These initial estimates may be provided by the user or may be determined internally. If the user decides to let these starting points be determined by the program, the following procedure is used:

1. Determine an initial estimate of the additive constant by first finding the smallest comparative interpoint distance in the original data. The initial estimate of the additive constant is then that number which converts this smallest distance into an absolute distance which is slightly larger than zero. If h_s is the smallest comparative interpoint distance, then the desired constant is:

$$(4) \qquad c = -h_s + (1/10\, n^2) \sum_{j<k}^{n} (h_{jk} - h_{..})^2.$$

2. Convert the absolute interpoint distances into scalar products between stimuli, b_{jk}^*:

$$(5) \qquad b_{jk}^* = \tfrac{1}{2} \left(\frac{1}{n} \sum_{j}^{n} d_{jk}^2 + \frac{1}{n} \sum_{k}^{n} d_{jk}^2 - \frac{1}{n^2} \sum_{j}^{n} \sum_{k}^{n} d_{jk}^2 - d_{jk}^2 \right).$$

3. Resolve the matrix of scalar products into its principal axes; and use the coordinates of the first t principal axes as the starting points for a t-dimensional solution.

An index of the goodness of fit was developed to aid in selecting the appropriate number of dimensions for the solution:

$$(6) \qquad \text{FIT} = 1 - \frac{\sum_{j<k}^{n} \delta_{jk}^2}{\sum_{j<k}^{n} (h_{jk} - h_{..})^2}.$$

This index ranges from zero to one, a perfect fit. It is best to look for dramatic increases in fit from a solution in n dimensions to a solution in $n + 1$ dimensions. The solution in $n + 1$ dimensions should be chosen. An absolute judgment on the goodness of fit using this index might be: .85 = fair, .90 = good, .95 = very good, and .99 and above = too good: you are probably including too many dimensions. It is advisable not to put too much emphasis on these absolute judgments. A lot depends on the quality of the original data. Using a comparative, dimension-by-dimension judgment probably provides a good basis for selecting an appropriate number of dimensions.

A theoretical description of the iteration procedure used for solution can be found in [3, 4] and the theory linking the iteration procedure to the problems of metric multidimensional scaling can be found in [1, 2].

INPUT: The program is set up to be able to run more than one data set at a time. The number of stimuli to be scaled can range up to 20, and the number of dimensions desired in the solution can range up to 10. The original comparative interpoint distances must be provided only in F, E, or D format.

OUTPUT: Printed output includes the coordinates of the stimuli on the axes of a Euclidean space; the value of the error function G; the index of the goodness of fit; the additive constant c; and the sum of squares of the original comparative interpoint distances, averaged to a mean of zero.

Because the program is written in double precision and is dimensioned for the maximum case of 20 stimuli in 10 dimensions, the program requires a GO region with slightly under 25OK bytes of core. For somewhat smaller problems the amount of core required can be greatly reduced by reducing the size of the vector called E in the program. For 20 stimuli in 10 dimensions it requires a dimension of 20,100. In general it requires $(NS^2 \times NT^2 + NS \times NT)/2$, so that any reduction in the number of stimuli or the number of dimensions can lead to a considerable saving in core. The routine is quite flexible. It can be easily modified to accommodate special cases. For assistance in modifying the program and for a source deck, contact the author.

COMPUTER: IBM OS 360, for the FORTRAN H compiler.

PROGRAM LANGUAGE: FORTRAN IV. The program is written in double precision for all real variables.

AVAILABILITY: Information and a listing of the program are available from:

Lee G. Cooper
Graduate School of Management
University of California, Los Angeles
Los Angeles, California 90024

REFERENCES

1. Cooper, Lee G. "Metric Multidimensional Scaling and the Concept of Preference," Working Paper No. 163, Western Management Science Institute, University of California, Los Angeles, 1970.
2. ———. "A New Solution to the Additive Constant Problem in Metric Multidimensional Scaling," *Psychometrika*, in press.
3. Fletcher, R. and M. J. D. Powell. "A Rapidly Converging Descent Method for Minimization," *Computer Journal*, 2 (July 1963), 163–8.
4. Gruvaeus, G. T. and K. G. Jöreskog. "A Computer Program for Minimizing a Function of Several Variables," *Educational Testing Service Research Bulletin RB-70-14*, 1970.

Lee G. Cooper

DISTSIM: A Computer Simulation of a Distribution System, *Frederick E. Webster, Jr., Dartmouth College; Glen R. Kendall, Department of the Interior; and Neil M. Henderson, Dartmouth College* [CAMR 18].

DESCRIPTION: DISTSIM is a computer simulation of a distribution system for a single product from one or more factories through warehouses and retailers to customers. The simulation enables the user to analyze the interaction of the elements of total distribution cost, to test alternative distribution strategies, and to find the most profitable means of satisfying a defined pattern of customer demand.

For teaching purposes, DISTSIM can be used by students or managers in programs of continuing education. As a research tool, DISTSIM permits testing of alternative strategies for managing the distribution system. The consequences of a set of decision rules can be determined by allowing the program to run for several periods and evaluating the outcomes.

The parameters of the system, including all costs and times as well as the number of facilities, can be readily changed to model a variety of actual distribution situations. The stimulation is complex and requires large computer capacity (approximately 16,000 words core storage capacity), limiting the number of facilities that can be included.

The user is required to make decisions about the following elements of the distribution system:

1. Amount of product to be ordered by each facility, with flexibility as to date of order and date of receipt.
2. Inventory levels at each facility. There are normal and excess inventory level charges for each facility.
3. Source of supply for each facility.
4. Transportation mode between any source and any destination. Each mode has a distinct cost and time for moving products between any two points in the system; the faster, most dependable modes are most costly.
5. Information to be purchased concerning demand, inventory levels, previous costs, and the status of orders, back orders and shipments at any point in time.
6. Frequency of order. Frequency of order requires a specific decision (in interaction with the decision on amount to order) since there is a fixed cost for ordering. Orders for early shipment cost more than orders for later shipment.
7. Cost of back orders. Out of these decisions emerges a level of customer service. Inability to meet customer demand results in loss of orders and a cost of back orders.

All parameters of the system can be modified by the user. At the present time there are 5 retailers, 2 warehouses, and 1 factory, 3 modes of transportation, and a schedule of costs—for inventory charges, orders, back orders, transportation (by mode), and information purchases. Demand is generated from a probability distribution which can also be modified to fit the desires of the user. The program is designed so that relatively minor programming (6–8 hours) is necessary to make major changes.

When used as part of an introductory marketing course, DISTSIM requires the student to recognize and treat the various aspects of distribution within the context of the whole system. For example, he must make decisions regarding the trade-offs among cost of transportation, time in transit, and reliability of transportation.

The education process is in three phases:

1. The student uses pencil and paper to decide order quantity, shipper, mode of transportation, and day of shipment.
2. The student makes the same decisions for each day, for each facility, by interacting with the program through use of teletype inputs.
3. The student develops a set of explicit decision rules which are programmed and tested over several weeks of simulated time.

The final output is an income statement for the period simulated showing total revenue, total distribution expenses, and gross profit contribution.

DISTSIM can be used either in teletype (TTY) mode or in decision rule mode. In TTY mode, the user interacts with the simulation, providing decision inputs as requested by the program; this mode gives the user some feeling for the complexities of managing a distribution system as it progresses through time, and it does not require programming ability. In the decision rule mode, the user develops a set of decision rules for making ordering, inventory, and transportation decisions, programs these decision rules into the simulation, and