

SIMATRIX: A program using the SAS system for multidimensional scaling analysis of sorting data

**EDWARD A. GREENBERG and
MARINA STOCK McISAAC**
Arizona State University, Tempe, Arizona

Multidimensional scaling (MDS) has become a popular analysis tool for assessing the perceived degree of relationship among objects or concepts. There are several ways of measuring subjects' ratings of object (dis)similarities, but the "stimulus sorting method" is frequently used because of its simplicity and ease of data collection. In the stimulus sorting method, subjects sort a set of stimuli into as many groups, or clusters, as they wish so that the stimuli within a group are judged to be more similar to each other than they are to stimuli in other groups. This task is very easy for subjects to perform and is less tedious than providing pairwise comparisons of stimuli when the number of stimuli is large. Using pairwise similarity judgments, for example, $n(n - 1)/2$ ratings are required. Data resulting from a sorting task are usually analyzed using nonmetric MDS with the help of computer programs such as ALSCAL (Young & Lewyckyj, 1979). The ALSCAL program is also available as a supplemental author-contributed procedure to the SAS® system¹ (Young, Lewyckyj, & Takane, 1983). SAS (Ray, 1982) is a comprehensive computer system for data management, statistical analysis, graphics, and report writing.

In order to apply nonmetric MDS, the sorting data must be converted into stimulus x stimulus similarity matrices for each subject. This paper describes one method for using SAS to convert sorting data into similarity data that can be input directly to ALSCAL for MDS analysis. This program was used in a recent study (McIsaac, Mosley, & Story, 1984) of visual dimensions in photographs, but, with minimal modification, it can be applied to any stimulus sorting task. Similar programs have been developed for use on microcomputers (Takane, 1981, 1982).

Description. The program makes use of the extensive data-manipulation features of the SAS system. Data are input directly with subject-specific information (e.g., demographic data). The sorted data are converted to matrices, one for each subject. Each matrix is square, with as many rows and columns as the total number of sorted stimuli. Cell values are either "1" (similar) or "0" (not similar), indicating that the two stimuli that intersect at a matrix cell were placed in the

same group or were placed in different groups. In addition to arranging the sorted data into matrices for each subject, a matrix of averaged similarities for all subjects and a separate file of subject demographic data are generated. SIMATRIX handles varying numbers of stimulus groups for different subjects in the same sample.

Input. A single program parameter is entered to indicate the number of stimuli that have been sorted by the subjects. For each subject, a subject-identification code is entered (e.g., "S001"), followed by demographic data in a fixed format. The last value coded on this record indicates the number of groups into which the subject classified the stimuli. Then stimulus numbers are coded free-field on as many additional data lines as there are groups for that subject. All of the stimulus numbers are coded from 1 to n , where n is the total number of stimuli presented to the subject. The first stimulus number coded for a group is flagged by the program as the "most representative" of that group. Other than this consideration, the order of stimulus numbers within a group or the order of groups for the subject is irrelevant. For example, assuming that Subject S001 sorts 10 stimuli into three groups, the data might be coded:

```
S001 ... demographic data ... 3  
10 6 4 5 3  
2 8 7  
1 9
```

The "3" on the first record indicates that, for this subject, three records follow, one for each group of stimuli. The three groups contain five, three, and two stimuli, respectively. The stimulus that best represents the group is number 10 in the first group, number 2 in the second group, and number 1 in the third group. Data for the next subject begin with a new "subject record" and continue with the sorting data for that subject. It can be seen that the data-entry task is quite simple. In addition to the data on the subject record, only n (number of stimuli) codes need be entered. A restriction of this program is that only consecutive integers 1 through n may be used as stimulus codes. If fewer than n stimuli are counted for a subject, a warning message is printed and the similarity matrix for that subject is not generated.

Output. Primary output is a file of square $n \times n$ similarity matrices, one per subject. This file can be passed directly to the ALSCAL procedure for replicated MDS analysis. A matrix of mean similarities (averaged across subjects) can be used for classical MDS and, when printed, provides information about the relative similarity of any two stimuli. Because the ALSCAL procedure uses only the lower left triangular portion of similarity matrices, the main diagonal is used to tally the

Edward A. Greenberg is with Academic Computing Services, and Marina Stock McIsaac is Assistant Professor, Department of Educational Technology, both at Arizona State University, Tempe, AZ 85287.

occurrences of a stimulus appearing in a group by itself. This information is ignored by ALSCAL, but may be of interest to the researcher. Several other data sets are produced, one containing the information from the subject records and another consisting of one record for every group of stimuli as classified by a subject.

Language and Computer. This program was implemented using Data and Proc(edure) steps available in the base SAS program product, Release 82.3, on an IBM system. The ALSCAL procedure is furnished with the base product, along with other author-contributed procedures in the supplemental library. The SAS system is available on IBM OS, Prime, Digital VAX/VMS, and Data General Eclipse systems.

Availability. A program listing may be obtained free of charge from Edward A. Greenberg, Academic Computing Services, Arizona State University, Tempe, Arizona 85287.

REFERENCES

- McISAAC, M. S., MOSLEY, M. L., & STORY, N. (1984). Identification of visual dimensions in photographs using multidimensional scaling techniques. *Educational Communication and Technology Journal*, 32(3), 53-68.
- RAY, A. A. (Ed.). (1982). *SAS user's guide: Basics*. Cary, NC: SAS Institute, Inc.
- TAKANE, Y. (1981). MDSORT: A special-purpose multidimensional scaling program for sorting data. *Behavior Research Methods & Instrumentation*, 13, 698.
- TAKANE, Y. (1982). IDSORT: An individual differences multidimensional scaling program for sorting data. *Behavior Research Methods & Instrumentation*, 14, 546.
- YOUNG, F. W., & LEWYCKYJ, R. (1979). *ALSCAL-4 user's guide* (2nd ed.). Chapel Hill, NC: Data Analysis and Theory Associates.
- YOUNG, F. W., LEWYCKYJ, R., & TAKANE, Y. (1983). The ALSCAL procedure. In S. P. Joyner (Ed.), *SUGI supplemental library user's guide* (pp. 1-16). Cary, NC: SAS Institute, Inc.

NOTE

1. SAS is a registered trademark of SAS Institute, Cary, NC.

(Manuscript accepted for publication July 11, 1984.)