Behavior Research Methods & Instrumentation
1979, Vol. 11 (4), 465-466

BEMSD: A general-purpose program for estimating missing data in a sociomatrix

PERCY B. BROWN
School of Business Administration
University of California, Berkeley, California 94720

A sociometric test typically seeks to determine the presence or intensity of a choice criterion relation that may exist between individuals who are members of a social collectivity. Data derived from such tests are usually presented in matrix form, where row and column vector labels identify collectivity members as givers and receivers of choices, respectively. Cells within the sociomatrix represent the presence or absence of a criterion relation between collectivity members in a binary-valued sociomatrix. When a multivalued sociomatrix exists, cell values represent the intensity of a criterion relation that exists between collectivity members.

Sociometric tests administered under field conditions are likely to generate sociomatrices that contain some null row vectors as a result of collectivity members' absenteeism during test administration or unwillingness to participate in the test. Researchers who wish to use sociometric data to assess linkage patterns among all collectivity members, or to partition the collectivity and assign all members to subsets of the collectivity, are presented with a problem by the appearance of null row vectors in the sociomatrix. The most direct solution to the missing data problem is to exclude from analysis those collectivity members who did not participate in the sociometric test. Exclusion is accomplished by deleting all null row vectors and corresponding column vectors from the sociomatrix. This approach to the missing data problem is undesirable because it destroys potentially useful information (e.g., typically nonnull column vectors associated with null row vectors) and fails to meet the objective of total inclusion of collectivity members in the analysis.

Brown (Note 1) has devised a technique with which Bayesian estimation of missing sociometric data (BEMSD) may be accomplished for each cell in null row vectors of an incomplete sociomatrix. The BEMSD program attempts to construct those choices a nonparticipant would have given to other collectivity members had he or she participated in the sociometric test.

Description. BEMSD searches a sociomatrix for null row vectors. Upon detection of null row vectors, Bayes' theorem is used to compute posterior probabilities for values that may be acquired by null cells such that:

$$p(x_{ij} <u> | x_{ji} <v>) = \frac{p(x_{ij} <u>) \cdot p(x_{ji} <v> | x_{ij} <u>)}{\sum_{u=o}^{q} p(x_{ij} <u>) \cdot p(x_{ji} <v> | x_{ij} <u>)},$$

given x_{ij} is a null cell in the sociomatrix where the choice intensity level u, at which individual i chooses individual j, must be estimated; x_{ji} is a reciprocal of the null cell x_{ij} in the sociomatrix, where individual i has been chosen by individual j at choice intensity level v; $p(x_{ij} <u> | x_{ji} <v>)$ is the posterior probability that cell x_{ij} will acquire a choice intensity value u, given that cell x_{ji} contains the choice intensity value v; $p(x_{ij} <u>)$ is the prior probability that cell x_{ij} contains the choice intensity value u; $p(x_{ji} <v> | x_{ij} <u>)$ is the conditional probability that any x_{ji} will acquire the value, v, given that any reciprocal x_{ij} has acquired the value u; and

$$\sum_{i=o}^{q} p(x_{ij} <u> \cdot p(x_{ji} < v > | x_{ij} <u>)$$

is the rule of elimination with $0 \leqslant u \leqslant q$. Typically, u and v are integers.

Application of the above equation for each value of u that may be acquired by x_{ij} results in the creation of a posterior probability distribution that contains q + 1 probability values. A specific estimated cell value for each null cell x_{ij} in the sociomatrix is then selected from posterior probability distributions uniquely associated with each of these null cells.

Optional features of BEMSD include the capability to analyze binary or multivalued sociometric data; alternate methods for construction of conditional probability distributions; iterative generation of subsequent null cell estimates; the ability to compare a reproduced sociomatrix, containing estimated null cells, with a reference sociomatrix containing a full data complement; alternate methods for selection of null cell values; selective restriction of printed output; and storage of summarized output in machine-retrievable form.

Input. Program input consists of control information and data. The user may provide specifications for program options, title information, and a read format for row vectors of the sociomatrix, or he may utilize default options and formats built into the program. The sociomatrix must be preceded by a header card, or card image, upon which is specified the number of row vectors in the sociomatrix and the maximum value that may be assumed by any null cell. Successive row vectors of the sociomatrix are then read as unit records. The first entry on each unit record is a label that uniquely identifies each row vector. The row vector label is followed by row vector cell values.

If the option that provides for comparison of reproduced and reference sociomatrices is selected, the user must also provide (1) the reference sociomatrix, (2) prior probability distributions for all cells in the reference sociomatrix, (3) conditional probability distributions for values contained in the reference socio-

0005-7878/79/040465-02$00.45/0

matrix, and (4) the distribution of choice intensities in the reference sociomatrix.

Output. Complete output includes echoed input control information, for verification of options selected by the user; identification of null row vectors; prior, conditional, and posterior probability distributions; and descriptive statistics derived from comparison of reproduced and reference sociomatrices.

If a sociomatrix that contains a full data complement is supplied as program input, BEMSD assumes that such a sociomatrix is a reference matrix and generates requisite prior, conditional, and choice intensity probability distributions. The reference sociomatrix and associated probability distributions are then stored in an auxiliary disk data file.

Limitations. The present version of BEMSD is capable of processing a 70 by 70 sociomatrix in which cells may assume integer values that range between zero and eight.

Language and Computer. The program was developed at the University of California, San Francisco, on an IBM 370/148 with virtual machine facility (VM)/370 Release 5. BEMSD source code is written in FORTRAN IV, and object code is obtained from the FORTRAN G compiler. Although the program is not interactive, the IBM conversation monitor system (CMS) has been used for program editing, compilation, and execution.

Core and Time Requirements. The main program and associated subprograms require 45K bytes of core. Necessary work space is approximated by $4[N^2(2q + 6\sqrt{N} + 7) + 3q^2 + 16q + 20]$, where N is the size of the collectivity over which the sociomatrix is formed and q is the maximum value that any cell in the sociomatrix may acquire. Therefore, when $N = 70$ and $q = 8$, approximately 454K bytes of core are needed for work space and maximum core required is about 500K bytes. Work space may be adjusted according to the user's needs.

Typically, a sociomatrix with 40% missing data, $N = 20$, $q = 8$, and the complete output option specified requires 80 sec of processing time under CMS.

Availability. The BEMSD user's guide, a current version of the program object code, and a test problem are available at no cost from Percy B. Brown, School of Business Administration, 350 Barrows Hall, University of California, Berkeley, California 94720.

REFERENCE NOTE

1. Brown, P. B. *A Bayesian approach to estimation of missing sociometric data.* Unpublished manuscript, University of California, Berkeley, California, 1979.

(Accepted for publication May 13, 1979.)