COMPUTER PROGRAM ABSTRACTS

An IBM 360/67 Program for Guttman-Lingoes Multidimensional Scalogram Analysis—III[1], *James C. Lingoes, The University of Michigan*[2]. **(CPA 307)**

Description. Multidimensional scalogram analysis is the most general and basic technique available for a *complete* representation of the three fundamental facets of data, that is, the Population of objects observed, the Items selectively chosen to delimit the observations, and the Categories, labels, or values into which the first two facets are mapped:

$$(1) \qquad PI \rightarrow C \quad \text{or} \quad C_i\,,$$

depending upon whether the categories are the *same* for all items (for example, agree, indifferent, disagree) or are *specific* to each item (for example, for item i: high, medium, low; and for item j: red, blue, green, and so on) $-C$ and C_i, respectively. The notation PI is to be read as the Cartesian product of two sets, where a particularized mapping is denoted by:

$$(2) \qquad \begin{aligned} &qj \rightarrow b \quad \text{or} \quad b_j \\ &(q \in P; \quad \text{and,} \quad b \in C \quad \text{or} \quad b_j \in C_j). \end{aligned}$$

Each data facet has a geometrical or coordinate representation in some m-dimensional space (members of the set P are represented as *points*, elements of I are *partitions* of the space, and members of C or C_i are *regions* of the space or regions within each partition, respectively). As such, multidimensional scalogram analysis (MSA) is a method for the analysis of the simple relation of "belongingness", free of any restrictive assumptions regarding item scaling properties or distribution form or what is the most appropriate index of relationship. Indeed, one can just as easily perform a conceptual analysis of some theory (where the items are relevant concepts and their categories are nominal characterizations of particular observations) as an analysis of purely quantitative data based on ratio scales. In either case, a perfect solution can be achieved, such that the original data can be reproduced or recovered exactly. The goal of MSA, of course, is to determine the minimum dimensionality within which such recovery is possible and, in such a manner, to reveal the implicit orderliness (or lack thereof), or pattern of relationships for the purposes of interpretation. The inferences that are possible are a function of the restrictions that are imposed on the solution. One important specification relates to the nature of regional boundaries, that is, what shape do they assume?

G-L(MSA-I), the first program in the multidimensional scalogram series (Lingoes, 1966; 1968), assumes that categories are specific to each item and the exact shape of the category boundaries is left unspecified provided that the regions are contiguous, that is, only points belonging to a specific category of an item appear within the region defining that category. MSA-I is thus the most general formulation of the pure analysis of belongingness. Commensurate with its generality, of course, is its weakness as a model for drawing conclusions about item and category interrelationships. The dimensions themselves (or any rotation of them) do not, in general, have psychological significance. Interpretations must flow directly from the configuration of points and any uniformities that emerge in respect to regional forms or to partitionings *despite* the lack of specification as to their nature. As a consequence, more than three dimensions presents (at the present time at least) an insuperable hurdle for interpretation, while, generally, only a one or two dimensional solution is practically manageable. The value of the method resides in its wide applicability to many kinds of problems which are in the stage of initial investigation.

G-L(MSA-II) (Lingoes, 1967; 1968), representing a stronger model for analysis, assumes that the categories are general to all items and, accordingly, imposes *circular* boundaries on the category regions. As part of the solution a *radius of inclusion* is defined, such that sitting on any person point and drawing a circle (more generally, a sphere), all categories into which that person falls will be included. Similarly, sitting on any category point, all persons falling in that category will be included. Thus, from the MSA-II solution one can make interpretations regarding both interpersons and intercategory relations. The dimensions (or some rotation thereof) have potential meaning. It then becomes feasible to make inferences for higher dimensional spaces. Interitem relationships, on the other hand, can be surmised only when very special configurations obtain, since the method lacks a specification to reveal them automatically.

---

[1] This research in nonmetric methods is supported in part by a grant from the National Science Foundation (GS-929).

[2] Prepared while on sabbatical to the *Faculté des Lettres et Sciences Humaines d'Aix*, France.

G-L(*MSA*-III), the present method, imposes a restriction that will permit inferences regarding interitem relationships as well as yielding information on category ordering within items. *MSA*-III assumes that the categories are specific to each item and, accordingly, constrains the boundaries to be both straight line and parallel to reflect the nature of the data and the kinds of inferences to be drawn, respectively. In common with *MSA*-II, metric information can be obtained from purely nonmetric considerations and the dimensions have potential meaning.

A succinct way of characterizing each of the *G-L* methods is afforded by what is called the *characteristic function* of observation and reproduction. Given these two mappings, a definition of the loss function to be minimized, and the set of side conditions at once reveals the nature of the method. To this matter we now turn for *MSA*-III.

Observation Function

$$(3) \qquad e_{pc_i} = \begin{bmatrix} 1 \text{ if } p \in c \text{ for } i \\ 0, \text{ otherwise} \end{bmatrix}$$

Reproduction Function

$$(4) \qquad \hat{e}_{pc_i} = \begin{bmatrix} 1 \text{ if } d_{pc_i} < d_{pb_i} \text{ for } i\ (b \neq c) \\ 0, \text{ otherwise} \end{bmatrix}$$

Subscript ranges are: [$p = 1, 2, \cdots, N$; $i = 1, 2, \cdots, n$; $b_i, c_i = 1, \cdots, k_i$ (number of categories in the i^{th} item)]. The d's are Euclidean distances calculated from:

$$(5) \qquad d_{pc_i} = \sqrt{\sum_{a=1}^{m} (x_{pa} - x_{c_ia})^2},$$

where the x's are the Euclidean coordinates, m is the dimensionality. *Side constraint:* within a fixed item i for $k_i \geqq 3$, any triple of categories: a_i, b_i, c_i must satisfy the following equality:

$$(6) \qquad d_{a_ib_i} + d_{b_ic_i} = d_{a_ic_i},$$

where the distance on the righthand side of (6) is the largest of the three distances.

Let

$$(7) \qquad \lambda = \frac{\sum_{p=1}^{N} \sum_{d_i=1}^{k_i} d_{pc_i} d^*_{pc_i}}{\sum_{p=1}^{N} \sum_{c_i=1}^{k_i} d^2_{pc_i}},$$

where $d^*_{pc_i} = d_{pc_i}$ when d_{pb_i} is the smallest distance *and* $e_{pb_i} = 1$, otherwise: $d^*_{pb_i} = d_{pc_i}$ and $d^*_{pc_i} = d_{pb_i}$ when $e_{pc_i} = 1$ and d_{pb_i} is the smallest distance between p and b within i, it being understood that the categories are mutually exclusive and exhaustive, that is,

$$(8) \qquad \sum_{c_i=1}^{k_i} e_{pc_i} = 1 \qquad (p = 1, 2, \cdots, N; \; i = 1, 2, \cdots, n).$$

The *loss function* is the $G - L$ coefficient of alienation:

$$(9) \qquad K = \sqrt{1 - \lambda^2}.$$

And, finally, the *solution conditions* are: a) K a minimum for b) m a minimum.

Output. 1) The set of coordinates, K; 2) The coefficient of reproducibility, R (see: Lingoes, 1967; 1968); 3) K, the measure of goodness-of-fit; 4) The reproduced score matrix; and, 5) The plots for each pair of coordinates ($m \geqq 2$), containing the person points and the category points. The latter serve the basis for constructing category boundaries for items by the drawing of perpendicular bisectors for adjacent categories within each item.

Capacity. Same as *MSA*-II.

Running time. Same as *MSA*-II.

Availability. See CPA 294, *Behav. Sci.*, 1968, 13.

REFERENCES

Lingoes, J. C. An IBM 7090 program for Guttman-Lingoes multidimensional scalogram analysis—I. *Behav. Sci.*, 1966, 11, 76–78;—II. *Behav. Sci.*, 1967, 12, 268–270.

Lingoes, J. C. The multivariate analysis of qualitative data. *Mult. Behav. Res.*, 1968, 3, (in press).

A CD3200–3600 Program for Nonlinear Factor Analysis[1], *Roderick P. McDonald, University of New England, New South Wales.* (CPA 308)

A new computer program for nonlinear factor analysis, PROTEAN (PRoduct/Polynomial ROtation TEst ANalysis) has been written. This effectively supersedes the group of programs previously described (McDonald, 1965*b*, 1967*c*), in that it incorporates virtually all the work of these programs and also includes a number of new possibilities. The relevant theory is developed and illustrated in McDonald (1962, 1965*a*, 1967*a*, 1967*b*, 1967*d*, and *in press*), and in McDonald and Burr (1967).

Preprocessing. Any standard program for orthogonal common factor analysis *may* be used for the initial processing of a matrix of raw data, in preparing data for submission to PROTEAN,

[1] The research was supported by a grant from the Australian Research Grants Committee.

though in most cases minor modifications and additions will be needed.

Let Z be an N ("subjects") by n (variables) sample score matrix, in which the variables have been standardized in the sample. Let F be a $n \times r$ ($r < n$) matrix of commonfactor coefficients (orthogonal model), and U^2 a $n \times n$ diagonal matrix of unique variances, estimated from Z, so that

$$(1) \qquad R \equiv \frac{1}{N} Z'Z \approx FF' + U^2,$$

to an acceptable approximation. It is strongly recommended that F be obtained by Canonical Factor Analysis (Rao, 1955). Given F, U, and Z, one then obtains "constructed" factor scores by Bartlett's method, that is the $N \times r$ matrix

$$(2) \qquad \begin{aligned} V &= [v_{ip}] = (F'U^{-2}F)^{-2}F'U^{-2}Z, \\ i &= 1, \cdots, N, p = 1, \cdots, r, \end{aligned}$$

(see McDonald and Burr, 1967). Whatever factor analysis program is employed, it should be modified if necessary so that F and V can be output by rows on punched cards, in the format ($A6$, $4x$, 10F7.4), where the first field should contain a row label. The N cards containing V should be randomly assigned to an Exploratory Sample and a Test Sample, of approximately equal size. (Limits: $r \leqq 10$, $n \leqq 80$. N not limited at this stage, but 500 is the limit for submission to PROTEAN, so that if N exceeds 1000, successive subsamples of $\leqq$500 must be used in exploratory and test runs.)

Description. The user sets up a hypothesis that one or more prescribed columns of V, suitably transformed by an orthonormal matrix, are in one to one correspondence with prescribed functions in the elements of one or more other (prescribed) columns, apart from "errors" due to unique variations.

One may hypothesize that

$$(3) \qquad v_{ij} = h_p(v_{ik}) + \text{"error"}, \qquad i = 1, \cdots, N,$$

where $h_p(*)$ is a polynomial of prescribed degree p, and j, k are prescribed.

One may hypothesize that

$$(4) \qquad v_{ij} = v_{ia}v_{im} \cdots v_{iu} + \text{"error"}, i = 1, \cdots, N,$$

where a, m, $\cdots$, u are prescribed. Under the hypothesis (3), the program seeks an orthonormal transformation such that the quantity

$$(5) \qquad \varphi_p = \frac{1}{N} \sum_{i=1}^{N} v_{ij} h_p(v_{ik})$$

is a maximum, where the parameters of the function $h_p(*)$ are determined by the moments of v_{ik}, according to the theory given by McDonald (1967b). Under the hypothesis (4), the program seeks an orthonormal transformation such that the quantity

$$(6) \qquad \varphi = \frac{1}{N} \sum_{i=1}^{N} v_{ij} v_{ia} v_{im} \cdots v_{iu}$$

is a maximum, according to the theory given by McDonald (1967a, and in press). One may combine such hypotheses, and the corresponding maximization procedures, in that the program can be set to maximize a weighted sum of two or more quantities of the type (5) and (6), where the weights are prescribed.

A third form of maximization procedure is available in the program. It may be shown that if either of the hypotheses (3) or (4) is true, a reasonable approximation to the desired orthonormal transformation will often be obtained if we maximize the sample fourth moment

$$(7) \qquad \varphi_f = \frac{1}{N} \sum_{i=1}^{N} v_{ij}^4 .$$

An initial, exploratory use of the program, seeking to maximize a weighted sum of two or more such fourth moments, should yield graphical output that suggests precise hypotheses of the type of (3) or (4). (In artificial, "error"-free data, this procedure works very well.)

Note that it is possible to combine hypotheses of the type (3) or (4) so that "function" columns and "argument" columns involved in separate hypotheses are not necessarily distinct. For example, we might postulate that the second column is a quadratic function of the first, that the fourth is a quadratic function of the third, and that the fifth is a product of the second and fourth (these in turn being functions of the first and third).

The program proceeds by rotating pairs of columns together in steps of a size prescribed in degrees by the user, through an arc of prescribed size. At each step, those functions of the type (5), (6), (7) that will vary as a result of rotation are evaluated, a weighted sum of these functions plus the unchanging functions is formed, and the step is selected at which its maximum is attained, over the prescribed arc. With this step as a new starting point, the next pair of columns are rotated together in the same way. When all pairs of columns have been rotated together, this whole procedure is iterated until either no change, at the prescribed step size, is indicated as necessary, or the numbers of iterations allowed has been reached. Up to four repetitions of the iterative procedure are allowed. The user may specify up to four hypotheses to examine in a single run on one problem, with a fixed arc size (necessarily 360 degrees) and a fixed step size (preferably coarse, say 20 degrees), the object being to select a preferred hypothesis for more accurate examination in a second run. Alternatively, he may fix on a single hypothesis and specify up to four arc/step sizes, each nested within the previous, the object then being to minimize the numerical work by finding a coarse approximation to the desired maximum, then running over the region surrounding the maximum so obtained, with a finer step size.

The output from one run may serve as input to the next. By combining the orthonormal transformation and other parameters output by a previous run using an exploratory sample, with the factor score matrix from a test sample, and suppressing the rotation procedure, hypotheses formed on the basis of the exploratory sample may be tested by a large-sample statistic. This statistic is also output at the end of the rotation procedure, but it may be spuriously large due to capitalization on chance in the rotation process itself.

Availability. A full writeup and Fortran listing are given in McDonald (1967*e*).

REFERENCES

McDonald, R. P. A general approach to nonlinear factor analysis. *Psychometrika*, 1962, 27, 397–415.

McDonald, R. P. Difficulty factors and nonlinear factor analysis. *Brit. j. Math. Stat. Psychol.*, 1965, 18, 11–23. (*a*)

McDonald, R. P. Some *IBM* 7090-94 programs for nonlinear factor analysis. *Research Memorandum 65-11*, Princeton, N.J.: Educational Testing Service, 1965. (*b*)

McDonald, R. P. Factor interaction in nonlinear factor analysis. *Research Bulletin 67-46*. Princeton, N.J.: Educational Testing Service, 1967 (*a*)

McDonald, R. P. Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, 1967, 32, 77–112. (*b*)

McDonald, R. P. Some IBM 7090-94 programs for nonlinear factor analysis. *Behav. Sci.*, 1967, 12, 72–74. (*c*)

McDonald, R. P. Nonlinear factor analysis. *Psychomet. Mono. No. 15*, 1967. (*d*)

McDonald, R. P. Protean—a comprehensive *CD* 3200/3600 program for nonlinear factor analysis. *Research Memorandum 67-26*, Princeton, N.J.: Educational Testing Service, 1967. (*e*)

McDonald, R. P. & Burr, E. J. A comparison of four methods of constructing factor scores. *Psychometrika*, 1967, 32, 381–401.

Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, 20, 93–112.

Insofar as we are geometricians, then, we reject the unforeseeable. We might accept it assuredly, insofar as we are artists, for art lives on creation and implies a latent belief in the spontaneity of nature. But disinterested art is a luxury like pure speculation. Long before being artists we are artisans; and all fabrication, however rudimentary, lives on likeness and repetition, like the natural geometry which serves as its fulcrum.

HENRI BERGSON, *Creative Evolution*