5. MDPREF (MultiDimensional PREFerance scaling)

1. OVERVIEW

Concisely: MDPREF (MultiDimensional PREFerence Scaling) provides internal analysis of two-way data of either a set of paired comparisons matrices or a rectangular, row-conditional matrix by means of a vector model, using a linear transformation of the data.

In the terminology developed by Carroll and Arabie (1979) MDPREF may be described as:

Data:	Model:
Two mode	Scalar-product
Two- or three-way	Two sets of "points"
Interval level	One space
Row-conditional	Internal
Complete or incomplete	

1.1 ORIGINS, VERSIONS AND ACRONYMS

MDPREF is based on a model developed at Bell Laboratories by J.D. Carroll and J.J. Chang (see Carroll , 1973). In this paper they develop two types of solution, one iterative and the other analytical, making use of the Eckart-Young decomposition theorem (1936). The MDPREF program implements this latter type, since the solutions obtained were virtually identical. A quasi-non-metric version (N-MDPREF) has been developed, but is not currently available. The MDS(X) version of MDPREF additionally includes directional statistics programmed by Charles Jones.

1.2 FURTHER SPECIFICATION

The MDPREF program may be thought of as providing internal analysis of preference data. This involves a set of subjects making preference or any similar sort of judgment about a set of stimuli (objects). From the data the program positions the stimuli as points in a Euclidean space, and represents each subject by a vector or line directed towards

the region where that subject's highest preference lies. In the case of perfect fit, the projections of the stimuli on this line correlate perfectly with the subject's preference scores.

1.3 RELATION OF MDPREF TO OTHER PROGRAMS IN THE LIBRARY

MDPREF analyses 'preference' data by means of a point vector or "ideal vector" model. Each subject or judge is represented in the space as a vector directed (which indicates the direction of increasing preference. The stimuli are represented as points in the same space, so that the projections of the stimuli onto a given subject's vector maximally reproduce his (her) preferences.

The same point vector model is implemented both in phase IV of PREFMAP and PROFIT, although in these cases the scaling is 'external' in the sense that the configuration of stimulus points is known beforehand and the subjects are fitted into this space as vectors. In MDPREF by contrast both subject vectors and stimulus points are positioned simultaneously from the information in the data, a so-called 'internal' analysis. (Note however that PREFMAP phase IV does allow a quasi internal analysis q.v.)

The same data as used in MDPREF may also be internally scaled by the non-metric distance model ('unfolding analysis') implemented in MDS(X) as MINIRSA. In this case, both subjects and stimuli are represented as points in the same space.

2. DESCRIPTION

2.1 INPUT DATA

MDPREF accepts input data in either of two main forms: as a set of pair-comparisons matrices (see David (1963), Ross (1934)) or as a set of rankings or ratings forming a rectangular, so-called "first-score" matrix. Options within the program differ with different data input and the type of input is chosen by the DATA TYPE parameter on the PARAMETERS card. In the following the "first-score" input is dealt with in sections 2.1.1 and 2.1.1.1 and the method of pair-comparisons and its associated options in sections 2.1.2 to 2.1.2.1.1. Further options are discussed in section 2.3.

2.1.1 The first-score matrix (DATA TYPE 1-4)

Suppose a set of N subjects is asked to rank in order of, say, preference, or give a rating to the set of p stimuli. The resultant data forms a rectangular 'row-conditional' matrix with N rows (subjects) and p columns (stimuli), called the "first score matrix" in the program. Each row of the matrix represents the preference rank or score assigned by that subject to the stimuli.

Such a matrix can also be obtained by taking the pair comparison matrix for a given subject and summing each row. The resultant column of scores gives that subject's rank order of preference for the stimuli and these may be collected to form the "first-score matrix".*

2.1.1.1 Ranks or Scores ?

Preference judgments may be represented for MDPREF (as in MINIRSA and other programs) in four distinct ways. The major distinction is that between a rank and a score. If a subject is asked to write down in his order of preference for five stimuli, he might respond with:

ACDEB

---

*The program in fact converts pair-comparison input into "first-score" form in this way before proceeding with the analysis.

If these letters (or stimulus names) are given numeric values this becomes:

13452

This is the rank-ordering method (analogous to Coombs's I-scales) and means that stimulus 1 is preferred to 3 which is preferred to 4 etc.

Data may be input to MDPREF in this form by specifying DATA TYPE(1). In various data-collection techniques it may be that the ordering obtained begins with the least-preferred stimulus so that the previous example would in this case be written as: BEDCA, signifying that B is least preferred, followed by E, and so forth. If this is the case then the data should be specified as: DATA TYPE(2).

A different way of representing such data is by the 'score' method. In this method each column represents a particular stimulus and the entry in that column gives the score or rating of that stimulus (for that subject) in his 'scale of preference'. Thus, in our original example the I-scale ACDEB (where A is preferrred to C, which is preferred to D etc.) would in this method be represented as follows:

		A	B	C	D	E
subject	i	1	5	2	3	4

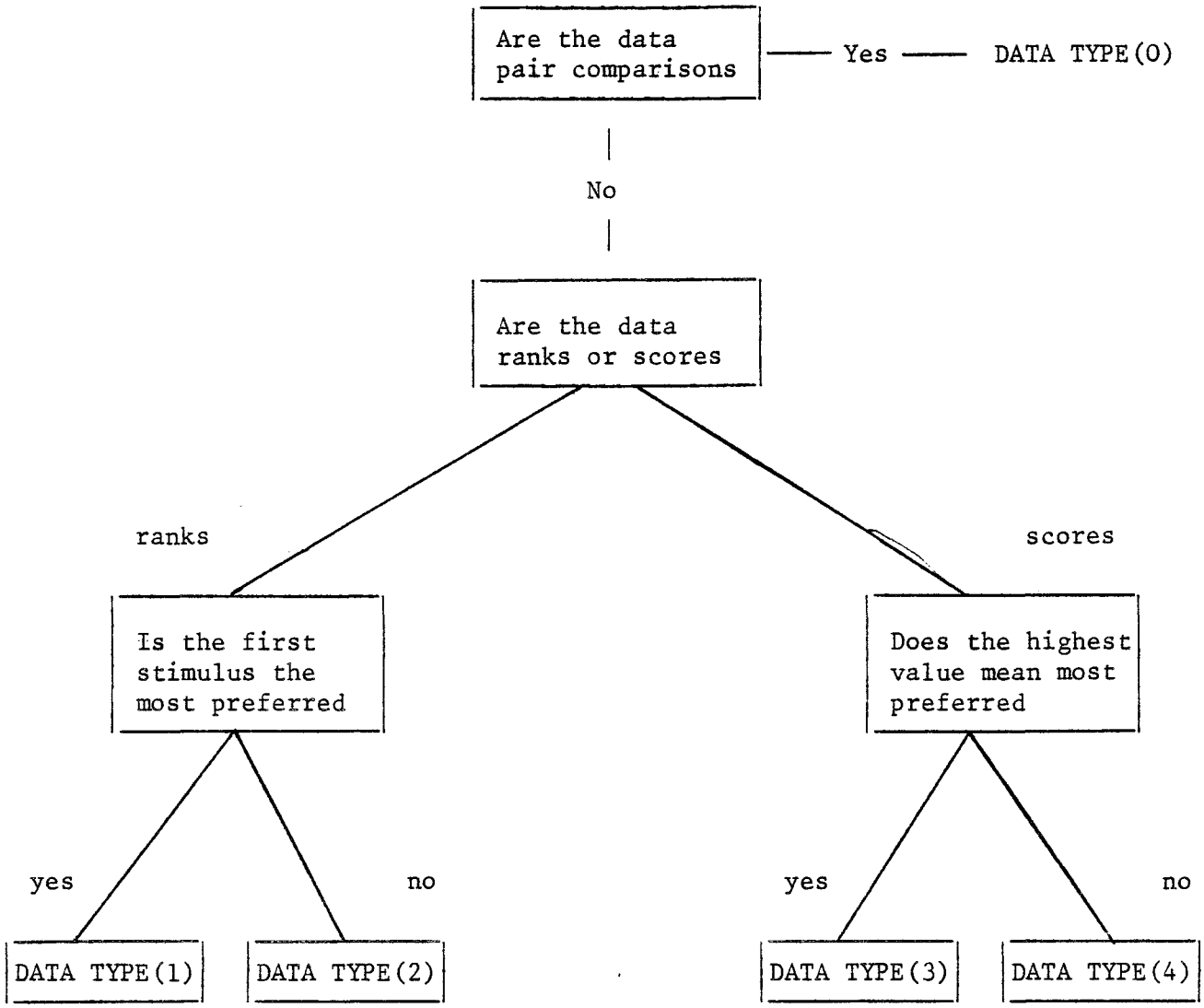
In this instance, the lowest number ('1') is used to denote the most preferred stimulus and the highest ('5') to represent the least preferred. This option is chosen by: DATA TYPE(3). Alternatively, the highest number might have been used to represent the most preferred stimulus and if this is so, DATA TYPE(4) should be specified.

(Although in illustrating the score method we have used the number 1 to 5, the data might equally well have been numerical ratings).

For an example see 2.1.2.1.1

Figure 1 provides a simple means of identifying the appropriate DATA TYPE value.

Figure 1.



2.1.2 The pair-comparisons matrices (DATA TYPE(Ø))

Suppose a subject is asked to consider all possible pairs of p stimuli and for each pair to indicate which stimulus (s)he prefers (or which stimulus possesses more of a given attribute). (S)he is asked to make p (p-1)/2 judgments of preference. (Since, with a large data set this number of pairs might become prohibitively large, strategies exist to reduce the number of judgments (see 2.3.1)). The data thus obtained may be collected into a square, asymmetric matrix whose rows and columns each represent the p stimulus points, whose entry a_{ij} takes the value 1 if the subject prefers stimulus i to stimulus j, and the entry a_{ji} will normally be Ø, meaning that the subject does not prefer stimulus j to stimulus i (but see 2.3.1). The subject may be allowed to express indifference between the stimuli, or leave blank a particular pair comparison. Allowance is made for these options in the program, and the relevant coding conventions are described in section 2.3.

If there are N subjects performing this test of preference, then there will be N such matrices. These are input to MDPREF by specifying on the PARAMETERS card the value DATA TYPE(0).*

2.1.2.1 Coding of paired comparisons matrices

In the example above the entry '1' was taken to stand for preference by the particular subject for the row-stimulus over the column stimulus, and the value 'Ø' for its converse. Further values are required to represent indifference between stimuli and missing data. Since coding conventions vary, the program allows the users to specify their own. This is done by means of the control card READ CODES (which has no operand field and has associated with it its own INPUT FORMAT card). READ CODES instructs the program to read in four values for the codes, the first of which will represent preference, the second its opposite ("anti-preference"), the third indifference and the fourth a missing data value.

---

*This value happens to be the default value, hence it need not in fact be specified if pair comparisons data are input.

2.1.2.1.1 Example

col.1	col.16
:	
INPUT FORMAT	(4A2)
READ CODES	
1 0 I B	
:	
:	

It will be noted that the codes must be specified as alphanumeric (A-type) variables. This allows, as in the example, for non-numeric, mnemonic codes to be used. Thus our example has the program to read

1 as the code for preference
Ø as the code for "antipreference"
I as the code for indifference
B as the code for a missing datum

Note also that even if, in a particular analysis, fewer than four codes are used, four values should nevertheless be specified and read under READ CODES, leaving the relevant column(s) blank (see Section 4).
The N pair comparisons matrices will be read by the READ MATRIX card, according to its INPUT FORMAT card. This card should specify the format of one row of the input matrices and the individual matrices should follow each other without separation. (For example, see 5.1).
Also note that if there are missing data then MISSING(1) should be specified on the PARAMETERS card.

2.1.3 Example of data types

When eliciting judgments by means of pair comparisons we need three things: a set of subjects who will evaluate a set of stimuli on a given criterion. Each subject vector will then represent the direction

in which that subject sees the criterion increasing over the configuration of stimulus points. Suppose we were interested in the 'user-friendliness' of the accompanying documentation of various computed packages. We might ask computing centre advisers to fill in the following:

... Taking each pair in turn please indicate by ticking in the box provided, which of the pair of packages is more "user friendly" ...

SPSS	☐ 3	GENSTAT	☐ 1
GENSTAT	☐ 1	CLUSTAN	☐ 4
MDS(X)	☐ 2	SPSS	☐ 3
BMDP	☐ 5	BMDP	☐ 5
....		....	

And we would go on to list (probably in random order) all twenty pairs of these five programs. For each adviser we would then construct a matrix similar to this:

Subject 32

	GENSTAT	MDS(X)	SPSS	CLUSTAN	BMDP
GENSTAT		1	1	1	1
MDS(X)	Ø		B	1	1
SPSS	Ø	B		1	N
CLUSTAN	Ø	0	0		1
BMDP	Ø	0	N	0	

This subject believes that GENSTAT is more 'user-friendly' than all the other packages, MDS(X) than CLUSTAN and BMDP and CLUSTAN than BMDP. Furthermore, (s)he left the pair SPSS/MDS(X) blank (hence B) and decided that there was No difference between BMDP and CLUSTAN.

2.1.3.1 Data for 'First-score'

In the example above, five stimuli were presented in pairs, twenty in all. If we were concerned with more than that number of stimuli we might feel that the number of pairs was too large for the subject to manage without boredom, error or bloody mindedness taking its toll. We might then decide to abandon the pair comparison method (which is, of course, particularly sensitive to intransitivities in a subject's data) and use instead a method of ranking. For instance, we might ask:

Please place the letters corresponding to the packages listed in the box provided so that the first letter represents the program which you feel to be most 'user-friendly' and the last the one you feel to be least 'user-friendly'.

A: GENSTAT
B: MDS(X)
C: SPSS
D: CLUSTAN
E: BMDP
F:
G:

(Most) User-friendly (Least)

This method is obviously less time-consuming but less sensitive than the method of pair comparison. In this case we simply take each subject's list of letters (I-Scale) and collect them onto cards with the subject numbers:

```
      .
      .
      .
S023    ABCDEFG
S024    GFEDCBA
S025    ACEGBDF
      .
      .
      .
```

Here we would specify DATA TYPE(1) to MDPREF to denote the fact that our data are ranked (I-Scales) with the highest 'preference' first.

2.2 THE MODEL

The MDPREF model represents the preferences of a subject for a group of stimuli as a vector through the configuration of stimulus points. This vector indicates the direction in which his (her) preference increases over the space. Substantively this makes strong assumption about the nature of preference, in that the model implies an "ideal" point - i.e. a point of maximum preference - at infinity (which is similar to the classic econometric assumption of insatiability. In MDPREF, where the point of maximum preference is at infinity, the contours are perpendicular to the vector). There is no reason to cavil, for instance at the idea of seriousness (Coxon 1980) or, as in our earlier example, "user friendliness" increasing uniformly over the space.

MDPREF is a linear (or metric) procedure and the measure of goodness-of-fit of the model to the data is a product-moment correlation. Consider one subject vector passing through a configuration of stimulus points with the perpendicular lines drawn from the points onto the vector. It is the values given to the points at which these perpendicular lines meet the vector which are maximally correlated with that subject's data. (This is guaranteed by the Eckart Young decomposition).

The subject vectors are normalised (for convenience only) to the same length, i.e. so that their ends lie at a common distance from the origin of the space, forming a circle, sphere or hypersphere as the case may be. Thus when a solution of more than 3 dimensions is represented (as it must be) as a set of 2-dimensional plots, some of the vectors will not, in fact, lie on the boundary circle since they will have been projected down from the higher dimensions. The length of the vector in the sub-space is related to the amount of variation in that subject's data explained by those two dimensions of the solution space.

2.2.1 Description of the Algorithm

1. If the input is in the form of pair comparisons matrices, these are converted into a "first-score" matrix.

2. The major and minor product-moment matrices are formed.

3. The inter-subject and interstimuli correlations are calculated.

4. The p-m matrices are factored by the Eckart-Young procedure to provide coordinates of the stimulus space and of the subject vector ends.

5. The first r columns of the relevant factor matrices are taken. These form the two configurations output as solution.

2.3 FURTHER OPTIONS

2.3.1 Dimensionality

The program prints out the latent roots of the matrices. The number of positive roots will be equal to the number of stimuli or the number of subjects, whichever is the smaller. The magnitude of the roots gives an indication of the amount of variation in the data accounted for by that dimension. The largest root will always be first and the others will follow in decreasing order. Some may be zero. An appropriate dimensionality may be chosen by means of the familiar scree-test.

2.3.2 Normalising and Centering

With the data in the form of a first score matrix the user may choose how the matrix is to be centred and normalised using the parameters CENTRE and NORMALISE. The default for these parameters is 0 and means no action.

Other options allow various courses. CENT(1) instructs the program simply to subtract the row means. This will, in a rating exercise, remove any effect due to differences in the actual values used by particular subjects. NORM(1) allows the program not only to subtract the row means but also to take out any effect due to differences in the range or spread of scores involved by normalising each row by dividing it by its standard deviation.

CENT(2) and NORM(2) perform the same operation on the column elements, i.e. subtracting column means and column normalising respectively. This latter option has the effect of taking out the unanimity effect in subjects judgements and leaving only the significan differences in judgements (see Forgas (1979)). CENT(3) instructs the program to double centre the matrix by subtracting both row and column means. NORM(3) does this and normalises the entire matrix.

2.3.3 Weighting of pair comparison matrices

Since pairwise judgements are often difficult to make, the user may sometimes wish to accord to each judgement a 'weight'. This might represent the degree of confidence which the subject attaches to his judgement, or perhaps the reliability which the researcher ascribes to each judgement.

If weights are input then there must be one weights matrix per subject.

The weights matrix immediately follows its associated pair comparisons matrix. This is read according to a WEIGHTS FORMAT control card which should be suitable for real (F-type) numbers. (For an example see Section 4.2.

2.3.3.1 The SAME PATTERN parameter

If, as often happens, there is more than one identical weights matrix, then the number of such matrices should be specified as the SAME PATTERN parameter. In this case, the weights matrix follows the first pair comparisons matrix and is read according to the WEIGHTS FORMAT card. Those pair comparisons matrices having the same pattern of weights then follow each other without separation.

2.3.4 Blocking of pair-comparisons data

If the number of pair-comparisons judgements has been thought too great then the researcher may resort to the use of incomplete data, i.e. certain element-pairs may not be presented to the subjects (see Burton & Nerlove, 1971). The resulting data-matrix will have 'blocks' missing. If one of these strategies is used and the data are arranged in blocks then BLOCK(1) must be specified on the PARAMETERS card so that allowance can be made in the calculation of row- and column-sums.

2.3.5 Interpretation of the solution

The MDPREF program positions the N subject vectors and the p stimulus points in a space of user-specified dimensionality. Interpretation of the stimulus configuration should proceed as for any MDS configuration, although it should be borne in mind that since this is an interval scaling model, the stimulus points have been positioned to secure maximum agreement with the subject's vectors. Consequently, interpretation of the position of stimulus points should be made with regard to the principal

direction(s) and spread of the subject vector ends.

The identification of 'outliers' amongst the subjects by visual inspection is straightforward. Certain one-, two- and multi-sample tests for mean direction also exist and have known properties, giving directional analogues to the analysis of variance. Appendix 3 gives a brief summary of statistics available in MDPREF and fuller description may be found in Pearson and Hartley (1972) and Mardia (1972). (See also Stephens (1962; 1969)).

3. INPUT PARAMETERS

MDPREF allows data to be input in two forms:

1. A "first-score" matrix in which case an N × p matrix is input.

2. A set of pair comparisons matrices in which case there will be N matrices, each p × p.

Options available with each type of option differ. The type of input is chosen by the parameter:

DATA TYPE	0	0: Data are in a pair-comparisons matrix.
		1: Data are in first-score form, ranked with first stimulus the most preferred.
		2: Data are in first-score form, ranked with first stimulus the least preferred.
		3: Data are in first-score form, as scores with highest score being the most preferred.
		4: Data are in first-score form, as scores with the highest score being the least-preferred.

3.1 OPTIONS WITH THE FIRST SCORE MATRIX

Keyword	Default	Function
MATFORM	0	0: The matrix is punched subjects (rows) by stimuli (columns). 1: The matrix is punched stimuli (rows) by subjects (columns).
GROUPS	0	The number of groups present in an analysis of variance should be specified. (See Appendix 3).
CENTRE	0	0: The data are not centred. 1: Row-means only are subtracted. 2: Column means only are subtracted. 3: Matrix is double centred.
NORMALISE	0	0: Matrix is not normalised. 1: Rows are centred and normalised. 2: Columns are centred and normalised. 3: Both rows and columns are centred and normalised.

3.2 OPTIONS WITH PAIRED COMPARISONS MATRICES

Keyword	Default	Function
SAME PATTERN	0	Sets the number of subjects whose pattern of missing data or weights matrices are the same.
WEIGHTS	0	0: No weights are input 1: Weights are input
BLOCK	0	0: The data are not arranged in blocks 1: The non-empty cells are arranged in blocks or are to be treated as such. (NOTE: Weights cannot be used with this option).
MISSING	0	0: There are no missing data 1: There are missing data in the matrix.

3.3 NOTES

1. The READ CONFIG card is not valid with MDPREF.

2. Note that even if only two or three codes are used in the paired comparisons matrices, the READ CODES card must specify four codes, which must be in the order specified.

3.4 PROGRAM LIMITS

Maximum number of stimuli	60
Maximum number of subjects	100
Maximum number of dimensions	5 5
Maximum number of groups	15

3.5 PRINT, PLOT AND PUNCH OPTIONS

The general format for printing, plotting and punching output is described in the Overview. In the case of MDPREF, the options are as follows.

3.5.1 PRINT options

Option	Form	Description
FINAL	p × r N × r	The stimulus matrix followed by subject matrix.
FIRST	N × p	The first-score matrix. (This is the input matrix after being modified i.e. centred/normalised). Means & standard deviations of subjects are printed.
CROSS-PRODUCTS		Four matrices are printed:
	N × N	1. the cross-product matrix (subjects)
	p × p	2. " " " " (stimuli)
	N × N	3. the correlation (PPM) matrix (subjects)
	p × p	4. " " " (stimuli)
SECOND	N × p	The second-score matrix.

Option	Form	Description
ROOTS		The latent roots.
RESIDUALS	N × p	The first-score matrix less the second-score.
CORRELATIONS	N	The correlation for each subject between the data and the stimulus projections is printed.

The default option allows for only the final configuration to be printed.

3.5.2 PLOT options

Option	Description
SUBJECTS	The n(n-1)/2 plots of the subject vectors in chosen dimensionalities.
STIMULI	The n(n-1)/2 plots of the stimulus points in the chosen dimensionalities.
JOINT	Both the above.
SHEPARD	In this case simply the first-score plotted against the second-score.
ROOTS	A scree diagram.
RESIDUALS	Histogram of residual values
GROUPS	A plot showing the groups (if chosen).

The default options allow for the first two dimensions of the joint space in each dimensionality only.

3.5.3 PUNCH options

Option	Description
SUBJECT SPACE	The final configuration of subjects is punched.
STIMULUS SPACE	The final configuration of stimuli is punched.

Nothing is punched by default.

4. EXAMPLES

4.1 EXAMPLE OF A SIMPLE RUN

```
col.1                col.16
-------------------------------------------------------------
RUN NAME             TEST RUN OF MDPREF
TASK NAME            FIRST SCORE OPTION
N OF SUBJECTS        20
N OF STIMULI         16
DIMENSIONS           2,3
INPUT FORMAT         (16X, 16F3.0)
PARAMETERS           DATA TYPE(1), NORMALIZE(1)
COMMENT              *****
                     THE PARAMETERS CARD SPECIFIES FIRST SCORE
                     MATRIX AS INPUT. THIS MATRIX IS TO BE
                     NORMALISED BY ROW
                     *****
READ MATRIX

   <the 20×16 first score matrix follows here>

PRINT                CROSS-PRODUCTS(2), SECOND(2,3)
COMPUTE
TASK NAME            PAIRED COMPARISONS OPTION
N OF SUBJECTS        20
N OF STIMULI         10
DIMENSIONS           2
COMMENT
                     N.B. THE FOLLOWING INPUT FORMAT CARD DESCRIBES
                     ONLY THE CODES FOLLOWING...
INPUT FORMAT         (4A2)
READ CODES
1 0 9 B
INPUT FORMAT         (9X,10A2)
COMMENT
                     ... WHEREAS THIS ONE REFERS TO THE INPUT MATRICES

                     NO PARAMETERS CARD IS INSERTED AS
                     ALL DEFAULT OPTIONS ARE ASSUMED
PLOT                 SHEPARD, RESIDUALS
READ MATRIX

   <20 square matrices, each of order 10 follow here>

COMPUTE
FINISH
-------------------------------------------------------------
```

4.2 EXAMPLE OF A RUN WITH WEIGHTS ADDED

```
col.1                col.16
RUN NAME             MORE MDPREF TEST DATA
TASK NAME            ... THIS TIME WITH WEIGHTS
N OF SUBJECTS        10
N OF STIMULI         5
DIMENSIONS           2,3
PARAMETERS           WEIGHTS (1)
COMMENT              *****
                     NEXT A FORMAT CARD FOR THE CODES
                     *****
INPUT FORMAT         (4A2)
READ CODES
1 0  B
COMMENT              *****
                     NOW ONE FOR THE INPUT MATRICES
                     *****
INPUT FORMAT         (5F1.0)
COMMENT              *****
                     AND FINALLY A WEIGHTS FORMAT
                     *****
WEIGHTS FORMAT       (5F2.0)
COMMENT              *****
                     WE NOW INPUT FOR EACH OF THE 10
                     SUBJECTS A P-C MATRIX AND A WEIGHTS
                     MATRIX WITHOUT SEPARATION
                     *****
READ MATRIX
 1111          \
0 111          |
00  1          } PAIRED COMPARISONS
000 1          |
0000           /
 0 2 1 9 4     \
 3 0 3 6 2     |
 8 5 0 3 1     } WEIGHTS
 4 8 2 0 9     |
 3 4 5 8 0     /
  <here, without break, follow 9 other such pairs of matrices>
PLOT                 SHEPARD (2)
COMPUTE
FINISH
```

BIBLIOGRAPHY

Bradley, R.A. (1954; 1955) The rank analysis of incomplete block designs I and II, Biometrika, 41, 502-537 and 42, 450-470.

Burton, M.L. and S.B. Nerlove (1976) Balanced designs for triads tests: two examples from English, Soc.Sci.Res., 5, 247-67.

Carroll, J.D. (1964) Non-parametric multidimensional analysis of paired comparisons data, Bell Telephone Labs.

Carroll, J.D. and P. Arabie (1979) Multidimensional scaling in M.R. Rosenzweig and L.W. Porter (eds.) (1980) Annual Review of Psychology, Palo Alto Ca. Annual Reviews.

Carroll, J.D. and J.J. Chang (1973) Models and algorithms for multidimensional scaling, conjoint measurement and related techniques, Bell Telephone Labs, mimeo (1968) How to Use MDPREF.

David, H.A. (1963) The method of paired-comparisons, London: Griffin, (Chapter 5).

Eckart, C. and G. Young (1936) Approximation of one matrix by another of lower rank, Psychometrika, 1, 211-218.

Forgas, J.P. (1980) Multidimensional scaling: a discovery method in social psychology, in G.P. Ginsburg, Emergent techniques in social psychological research, London: Wiley.

Mardia, K. (1972) Statistics of directional data, London: Academic Press.

Pearson, E.S. and H.D. Hartley (eds.) (1972) Biometrika tables for statisticians, vol.II, C.U.P.

Ross, R.T. (1934) Optimum orders for the presentation of pairs in the method of paired comparisons, J.Educ.Psychol., 25, 375-382.

Slater, P. (1960) The analysis of personal preferences, B.J.Stat.Psych., 13, 119-135.

Stephens, M. (1962) The statistics of directions, Ph.D. dissertation, University of Toronto.

Stephens, M. (1969) Multi-sample tests for the Fisher distribution for directions, Biometrika, 56, 169-81.

Tagg, S.K. (1980) The analysis of repertory grids using MDS(X), in Coxon and Davies (eds.) Working papers in multidimensional scaling, MDS(X) project, Cardiff.

Takane, Y., F.W. Young and J. de Leeuw (1977) Nonmetric individual differences multidimensional scaling. An alternating least squares method with optimal scaling features, Psychometrika,42, (1), pp 7-67.

Tucker, L.R. (1955) Description of paired comparisons preference judgments by a multidimensional vector model, Princeton N.J: ETS, RM 55-7.

Tucker, L.R. (1960) Dimensions of preference, Princeton N.J: ETS, RM 60-7.

APPENDIX 1: THE RELATION OF MDPREF TO PROGRAMS NOT IN MDS(X)

MDPREF is analogous to the INGRID program widely used in the analysis of repertory grids (Slater, 1960). The use of various MDS(X) programs in this type of analysis is described in detail by Tagg (1980); see also Forgas (1979). A similar model is used by Tucker; see Tucker (1955; 1960). A MDPREF-like model is not included in either ALSCAL or the G-L series but an approximation is implemented by the Takane-Young-de Leeuw program PRINCIPALS (see Takane et al, 1975).

APPENDIX 2: THE METHOD OF COMPUTATION*

Given N subjects and p stimuli, we have N matrices of order p.

$$\underset{\sim}{D}^{i} = \{d^{i}_{jk}\} \qquad \begin{aligned} i &= 1,2,\ldots N \\ j,k &= 1,2,\ldots p \end{aligned}$$

where

$$d^{i}_{jk} = \begin{cases} +1 & \text{if individual } i \text{ judges } j > k \\ -1 & \text{if } \quad " \quad\quad i \text{ judges } j \quad k \\ +\emptyset & \text{if } \quad " \quad\quad i \text{ judges } j = k \\ -\emptyset & \text{if no response} \end{cases}$$

Further, assume two solution matrices

$$\underset{\sim}{X} \equiv \{x_{ia}\} \qquad \begin{aligned} i &= 1,2,\ldots N \\ a &= 1,2,\ldots r \end{aligned}$$

a configuration of N subject vectors in an r dimensional space and

$$\underset{\sim}{Y} \equiv \{y_{ja}\} \qquad \begin{aligned} j &= 1,2,\ldots p \\ a &= 1,2,\ldots r \end{aligned}$$

a configuration of p stimulus points an r dimensional space. We then define the first score matrix, $\underset{\sim}{S}$ $(N \times p)$

$$\underset{\sim}{S} = \{s_{ij}\} = \underset{\sim}{X}.\underset{\sim}{Y}'$$

where

$$s_{ij} = \underset{\sim}{x}.\underset{\sim}{y}'$$

(This is the so-called second score matrix)

---

*This appendix is based on Carroll (1964), which is used with permission

We further define

$$\delta^i_{jk} = s_{ij} - s_{ik}$$

as the difference between the preference for j over k for subject i.

Given these we define a criterion of agreement between data and a given configuration

$$C = \sum_i W_i \left(\frac{(\sum_{j \neq k} d^i_{jk} \delta^i_{jk})^2}{\sum_{j \neq k} (\delta^i_{jk})^2} \right)$$

where W_i is the (optional) WEIGHT assigned by subject i.
This may be maximised by defining a matrix $\underset{\sim}{S}^*$ whose general entry

$$\underset{\sim}{S}^*_{ij} = \sqrt{W_i} \sum_{j \neq k} (d^i_{jk} - d^i_{kj})$$

and factoring $\underset{\sim}{S}^*$ such that

$$\underset{\sim}{S}^* = \underset{\sim}{U}\ \underset{\sim}{B}\ \underset{\sim}{V}'$$

From these factors we derive

$$\underset{\sim r}{U}\quad \underset{\sim r}{B}\quad \underset{\sim r}{V}$$

Where $\underset{\sim r}{U}$ and $\underset{\sim r}{V}$ are the matrices given by the first r columns of $\underset{\sim}{U}$ and $\underset{\sim}{V}$ respectively, $\underset{\sim r}{B}$ is a diagonal matrix of order r (i.e. the first r rows and columns of $\underset{\sim}{B}$).
The solution matrices $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are then given by

$$\underset{\sim}{X} = \underset{\sim r}{U}\ \underset{\sim r}{B}$$

$$\underset{\sim}{Y} = \underset{\sim r}{V}$$

This factorisation is carried out by means of the Eckart-Young (1936) procedure.

The matrix $\underset{\sim}{U}$ is a matrix with eigenvectors of $\underset{\sim}{S}\,\underset{\sim}{S}'$ as its columns and the matrix $\underset{\sim}{V}$ has as columns eigenvectors of $\underset{\sim}{S}'\underset{\sim}{S}$. $\underset{\sim}{B}$ is the diagonal matrix of corresponding eigenvalues. If the eigenvalues are ordered according to magnitude (and the columns of $\underset{\sim}{U}$ and $\underset{\sim}{V}$ permuted correspondingly), then $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ is the matrix of rank r yielding the best least-squares approximation to $\underset{\sim}{S}^*$. Carroll (1964) proves that the criterion C is indeed maximised by factorization of the $\underset{\sim}{S}^*$ matrix.

APPENDIX 3: STATISTICS FOR DIRECTIONAL DATA

A3.1 Definitions

We shall be concerned with differences and similarities between subjects' preferences, i.e. between the vectors. A sample of vectors may be thought of as drawn from a population whose overall direction is the polar vector. The average direction for the sample set of vectors is called the modal vector. The vector sum of a set of vectors is a resultant vector and its sum of squares its length (R).

A3.2 Measures of distribution

It is clear that the greater the length of the resultant vector, the more agreement exists in the sample.

The probability density of distribution of vectors around the polar vector is given by kappa, high values of which imply a concentrated symmetrical distribution of vectors around the polar, while a zero value gives a uniform distribution around the circle or sphere.

Kappa may be estimated from sample data by

$$K = \frac{N-1}{N-R}$$

where N is the total number of vectors (and also, obviously, the sum of the lengths of N unit vectors) and R the length of the resultant. Note, however, that this approximation is only accurate when $R/N > 0.7$ (i.e. kappa > 3.3).

A3.3 Tests of significance

A directional analogy to one-way analysis of variance is an approximate test for comparison of polar vectors from two or more samples. The parameter $2K(N-R)$ is distributed approximately as chi-square with $2(N-1)$ degrees of freedom.

It is possible, arguing from the analogy with analysis of variance, to partition the chi-square for the concentration of vectors from two independent samples about a common estimated mean vector. The overall χ^2 is the sum of the components from (a) the concentration of vectors in each sample about their mean vectors, and (b) the concentration of the two estimated mean vectors.

An approximation to the F-test compares 'between-group' and 'within group' components. With S samples an F-distribution is approximated by

$$U = \frac{(N-S)\quad (\sum_i R_i - R)}{(S-1)\quad (N - \sum_i R_i)}$$

In the three-dimensional (spherical) case this statistic has (2S-2) and (2N-2S) degrees of freedom in the numerator and denominator respectively. In the circular (two-dimensional) case these values are respectively (S-1) and (N-S).

The statistical theory which would allow us to proceed to a two-way analysis of variance has not yet been developed.

A3.4 Input parameters for statistics

Statistics are only available with the 'first-score' option. If the user wishes to use the program to perform the one way analysis (s)he should specify the number of groups on the GROUPS parameter on the PARAMETERS cards. Each row of the matrix (i.e. each subject) should then be assigned to a group. This is done by appending to each row the number of the group to which that subject is assigned. The INPUT FORMAT card should be amended to read this number as an integer (I-type) value.