13. PREFMAP (PREFerence MAPping)

1. OVERVIEW

Concisely: PREFMAP (PREFerence MAPping) provides external analysis of two-way, row-conditional data (usually a preference measure) by four different models, using either a monotonic or a linear transformation of the data.

Alternatively, following the categorisation developed in Carroll and Arabie (1979) PREFMAP may be described as:

Data:	Two mode	Model:	External
	Two way		Constrained
	Interval or Ordinal level		Scalar products and distance models
	Row conditional		Two sets of "points"
	Complete or incomplete		

1.1 ORIGIN, VERSIONS AND ACRONYMS

The hierarchy of preferential analysis programs was developed at Bell Laboratories by J.D. Carroll and J.J. Chang. The present program is PREFMAP2 which superseded the original PREFMAP in 1972.

1.2 PREFMAP IN BRIEF

PREFMAP provides 'external' analysis of preference data: i.e. it seeks to relate subjects' preferences for a set of stimuli to an existing configuration of the stimulus points by means of four different models. These models are the four 'phases' of the program and they form a hierarchy inasmuch as each phase is a special case of the one preceding it. The subjects are mapped into the stimulus configuration as points of "maximum preference" (or, in phase IV, vectors) so that the preference scores are maximally reproduced in the distances from each subject's point to the stimulus point. (The program also allows for subject points to represent least preference on some or all

dimensions).

PREFMAP may be used to provide a quasi-internal analysis by generating a stimulus configuration from the preference data.

1.3 THE RELATION OF PREFMAP TO OTHER PROGRAMS IN THE LIBRARY

The PROFIT program with linear options is exactly analogous to the linear version of phase IV of PREFMAP, while phase III implements essentially the same model as MINIRSA, except insofar as that latter program provides a wholly internal analysis of the data and only permits points of maximum preference on all dimensions. Similarly, MDPREF provides an internal analysis with a model analogous to the linear version of PREFMAP phase IV. (PREFMAP may, however, be used to give a quasi-internal analysis (see 2.3.1 below).

2. DESCRIPTION

2.1 DATA

PREFMAP expects two matrices to be input in the usual case:

i) a matrix defining the configuration of stimulus points

ii) a matrix of subjects' preferences.

2.1.1 The configuration

The configuration of stimulus points into which the program maps subjects' preferences may be an *a priori* arrangement of some sort or the result of a previous MDS analysis.

Alternatively, the program may generate a configuration from the preference data themselves, thus providing a quasi-internal analysis. The analysis is not truly internal, since once the configuration has been generated, it is not changed during the course of the analysis to provide a better fit to the data (but see 2.3.3 below). If this option is required the user should specify INITIAL(1) on the PARAMETERS card.

2.1.2 The preference matrix

As is usual in MDS "preference" is used as a shorthand for any proximity or (dis)similarity type of judgement or data. Any data which can be thought of as resulting from a question of the type "which has the more of attribute x" are amenable to analysis in PREFMAP and other "preference" analysis programs in MDS(X).

Specifically, the data will consist of:

1. a set of *subjects* judging
2. a set of *stimuli*
3. on the basis of some given *criterion*.

Such data might be collected by means of the method of pair comparisons (although PREFMAP does not accept input in this form). Alternatively, methods of ranking or rating may be used.

2.1.2.1 Ranks or Scores

Preference judgements may be represented for PREFMAP (as in MDPREF 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

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

13452

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

Data may be input to PREFMAP 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 preferred to C, which is preferred to D etc.) would in this method be represented as follows:

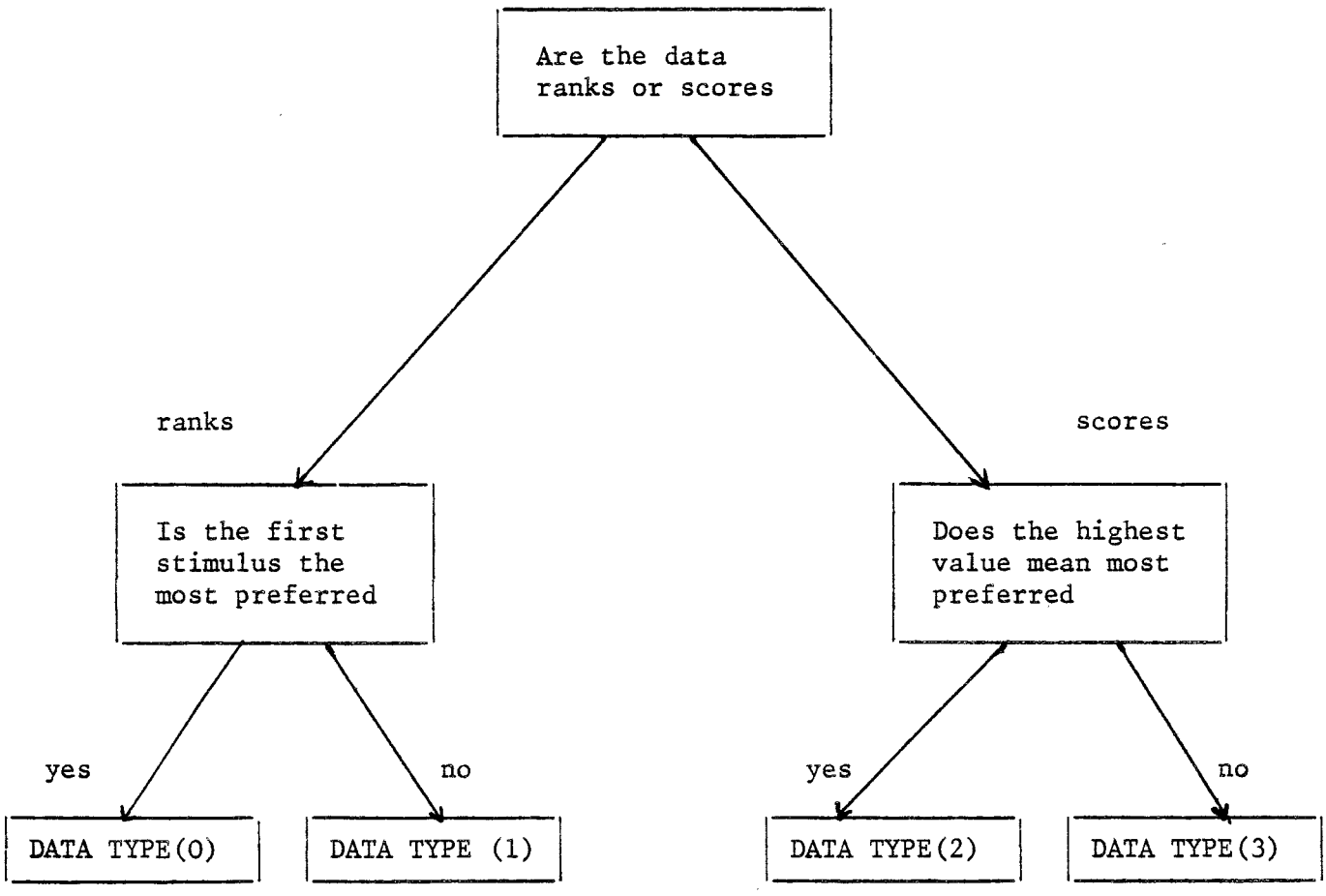
		A	B	C	D	E
subject	i	1	5	2	3	4

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).

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

Figure 1



2.2 THE MODEL

The four phases in PREFMAP are in fact four distinct models, which are "nested" in the sense that each phase is a special case of the one preceding it. In all four phases the program seeks to represent the preference information within the given configuration. As the program moves from phase 1 to phase 4 the restrictions become more strict.

Phase I: The general unfolding model

In this model each subject is allowed

1. orthogonally to rotate the axes of the similarity space to his own reference dimensions;

2. to assign to each of his dimensions a different evaluative weight or salience.

Within the configuration space each subject is represented as a point located by his most preferred position on each of the constituent dimensions of this space, i.e. placed at his ideal point.

Phase II: The weighted unfolding model

In this model subjects

1. are assumed to share the same reference axes (see 2.3.3) but

2. are allowed to weight each of the common dimensions (axes) differentially.

The subjects' ideal points are again mapped into the space.

Phase III: The simple unfolding model

In this model the subject may neither

1. rotate axes

nor

2. differentially weight them

but each ideal point is mapped into the stimulus space.

Phase IV: The vector model

The vector model represents each subject's preference as a vector directed towards his region of maximum preference. The projections of the stimulus points onto the vector reproduce the subject's preference values. Moreover, the angle which the vector makes with each dimension can be thought of as representing the salience of that dimension in the preference judgement.

The program thus forms a hierarchy of models in the sense that each higher numbered phase is a special case of the lower numbered.

The user may choose which of the models (s)he wishes the data to be analysed by. This is done by means of the parameters S-PHASE and E-PHASE (Start-phase and End-phase which may have the same value, in which case only that one model is implemented). By default, the program will begin at phase I and compute solutions by the other models in order of increasing restriction to end at phase IV.

The program computes the F statistic as a measure of improvement as it moves from phase to phase (see 2.3.6).

2.2.1 The transformation

In addition to containing four separate models for representing preference data, PREFMAP allows the user to choose the form of the scaling function linking data to the solution, which is tantamount to defining the level of measurement of the preference data.

2.2.1.1 The linear option

By specifying FIT(0) on the PARAMETERS card the user signifies that, in phases I-III the distances from the subject's ideal point to the stimulus points will be (a least-squares approximation to) a linear transformation of the rating values ascribed to the stimuli by the subject. In phase IV it is the distances between the projection of the points onto the subject's vector which are so related.*

This option is applicable if the data are believed to have interval (or, indeed ratio) level properties.

2.2.1.2 The non-linear (monotone) option

Alternatively, the user may believe that the data will bear only ordinal interpretation and that the solution distances should be allowed to be (as close as possible to) some ordinal transformation of the original data values. This option is chosen by specifying FIT(1) on the PARAMETERS card. FIT(1) should only be chosen, however, if there are no tied data values. If there are equal data values then the user has the choice of two monotonic approaches.

2.2.1.2.1 The primary approach (FIT(2))

By specifying FIT(2) the user indicates that the information contained in the ties is important and that ties should be matched by ties in the fitting values. This is to say that any equal data values which have associated with them unequal distances in the solution will decrease the goodness-of-fit of solution to data.

---

*Phase IV of PREFMAP is exactly analogous to the PROFIT analysis and users are referred to that documentation for a detailed description of the model.

2.2.1.2.2 The secondary approach (FIT(3))

If, however, the user believes that existence of tied data values is not significant information, (s)he may then specify FIT(3). The program will then not be constrained to matching equal data with equal fitting values, and will 'break' ties if in so doing the goodness-of-fit is improved.

Strictly speaking the non-linear version of PREFMAP is only "quasi-non-metric" inasmuch as the program first computes the linear solution and then uses this as the starting point for a series of iterations by Kruskal's (1964) monotone regression procedure rather than proceeding directly to a monotonic estimation.

2.2.2 The PREFMAP algorithm

The linear procedure involves at each phase the solution of a multiple regression equation. The coefficients of the equation are output as beta-weights (see 2.3.5).

The non-linear procedure solves first the linear regression then proceeds to use the linear solution as a starting-point for the monotone regression, which is an iterative procedure.

1. A set of fitting-values is defined, which is a montone rescaling of the data.

2. The measure of departure of the distances in the configuration between the subject and stimulus points and the rescaled data values, $STRESS_2$, is calculated.

3. A number of tests are performed: e.g. Has STRESS reached a minimum ? Has improvement in STRESS been great enough in the last few iterations to warrant continuing ? etc. If the answer to any of these is Yes, then the configuration is output as solution. If not,

4. The direction on each axis in which each subject point has to move in order to minimize STRESS, together with the probable size of that movement is calculated.

5. The ideal points are moved in accordance with 4 and the program returns to step 2.

2.3 FURTHER FEATURES

2.3.1 Quasi-Internal analysis

Although PREFMAP is primarily a program for external analysis, the facility exists to perform a quasi-internal analysis. If no configuration is input by the user, the program will generate one from the preference data themselves by an Eckart-Young factorisation of the minor product of the preference data matrix (see Carroll and Chang, 1971), which has often been transformed in some way (see below). A number of options is allowed to the user in choosing how this configuration is formed by means of the INITIAL parameter.

2.3.1.1 When INITIAL(0)

When INITIAL is given the value 0 the preference matrix is doubled-centred. This means that both row- and column-means are extracted and the overall mean of the matrix added back in. This means in essence that only interaction effects remain. The configuration is then formed in the manner indicated above. When this option is chosen with FIT(0) and S-PHASE(3), the resulting analysis is equivalent to internal metric unfolding (Carroll and Chang, 1971).

2.3.1.2 When INITIAL(1)

Alternatively, before the configuration is formed only the row-means of the preference matrix are extracted. This has the effect of removing from the configuration any influence due to the actual values used by different subjects, though not any effect due to the spread in the scores.

2.3.1.3 When INITIAL(2)

If, in addition, the user wishes to remove influence of the actual spread of preference scores from the initial configuration, then choice of INITIAL(2) instructs the program to standardise the preference scores before extracting the configuration.

2.3.1.4 When INITIAL(3)

On occasion, the user may wish simply to remove the row effects and the column effects without fully double-centring the matrix. In this case the user should choose INITIAL(3).

These options are not available when READ CONFIG is used to input a separate stimulus configuration.

Once the configuration has been generated by the program it does not change during the analysis (but see 2.3.3). Carroll (1971) has shown that the use of this method of generating an initial configuration gives a general least-squares solution to the metric unfolding problem.

2.3.2 Normalisation of scale values

Independent of whether the analysis is external or internal and what information is taken into account if an when generating the configuration, the user may choose whether, in mapping the preference data into the configuration, (s)he wishes to ipsatize the subjects' data. If NORMALISE(1) is chosen each subject's preference scores are normalised by removing the subject's mean score and dividing by their standard deviation. The default option NORMALISE(0) leaves the scores unchanged.

2.3.3 Canonical rotation: the KEEP parameter

In phase II subjects are thought of as applying evaluative weights to the dimensions of the stimulus space. The orientation of these axes is then, in a substantive sense, non-arbitrary. As part of the solution to phase I (where arbitrary rotation of the axes is allowed) the program produces a canonically-rotated space. This is the stimulus configuration with its axes rotated to those of an "average subject" whose data are formed by averaging the subject's preference scores. These axes have certain optimal properties (see Carroll, 1972) and

will not normally correspond to (e.g.) the principal axes of the configuration.

In the normal way the program, in passing from phase I to phase II, will use this average subject space as the basis of its calculations in this and lower phases. If, however, the user feels that the axes of the input configuration have substantive significance and wishes to retain them in the lower phases, then (s)he should set the parameter KEEP to 1 on the PARAMETERS card.

If a solution is not being computed in phase I, then the program will simply use the original input configuration.

2.3.4 Initial weight estimates

If the user begins the analysis at phase III (i.e. S-PHASE(3)) then (s)he has the option of having the program read in estimates of the dimension weights. This is done by means of the READ WEIGHTS control card and its associated INPUT FORMAT card. The weights (one per dimension) should follow the READ WEIGHTS card.

2.3.4.1 Example

```
col 1                    col 16
DIMENSIONS               4
INPUT FORMAT             (4F5.3)
READ WEIGHTS
0.54_0.21_0.71_0.5
```

2.3.5 The BETA weights

In the linear procedure the program seeks to represent the preference data as a linear function of the (squared) distance between the subject's ideal point/vector and the stimulus points. This is done in a multiple

regression. For this to be done the distance equation must be multiplied out into its constituent terms. Since the form of the distance equation differs in each phase with the differing weighting options, the number of terms involved in this expansion will differ at each phase. The program estimates a standardised regression coefficient for each of the terms in this expansion. These are the so-called BETA weights. Appendix 3 gives the form of the regression equation for each phase and the order in which the weights are output if requested. For a fuller explanation see van Schuur (1976).

2.3.6 The F-statistic

The F-statistic is computed for two purposes by the program.

2.3.6.1 Within phases

Within each phase the statistic is computed to ascertain whether the variance explained by the model differs from the residual (unexplained) variance.

2.3.6.2 Between phases

Between phases the statistic measures the extent to which the higher-level model explains significantly more variation than the lower level one.

In both cases the F-statistic is only approximate since the stages are not independent.

2.3.7 Phases and strategies

It should be noted that choice of S-PHASE affects the form of the subsequent analysis. In particular note the use of the canonical reference space noted above (2.3.3). Further considerations include the following.

2.3.7.1 When S-PHASE(1)

At the end of this phase the original input configuration is replaced by the phase 1 "private space" of the average subject (see 2.3.3). The rationale is that the average subject's configuration is likely to provide a better fit than the original configuration to this data at subsequent stages.

2.3.7.2 When S-PHASE(2)

At the end of this phase, the original input configuration is replaced by the phase 2 "private space" of the average subject (i.e. differentially weighted to conform but to the averaged data). This removes the canonical reference space for subsequent phases.

2.3.7.3 When S-PHASE(3)

If the user begins at phase 1 or 2, it is quite possible that the individual ideal point at phase 3 will be a "saddle point", i.e. a point which is a mixture of optimal preference on some one or more dimensions (signalled by a positive value) and pessimal preference (negative value) on others (see Carroll, 1972, pp 121-123). However, if one begins with S-PHASE(3), the original configuration is unchanged and all ideal points will be positive. In many applications the user may wish to start at a higher phase and finish at phase 3 or 4, but then perform a separate run with S-PHASE(3) if the simple unfolding model turns out to be adequate. In this case a solution is obtained where all ideal points are constrained to be positive.

2.3.7.4 When S-PHASE(4)

The program will simply implement the vector model.

3. PARAMETERS

3.1 LIST OF PARAMETERS

Keyword	Default Value	Function
S-PHASE	1	Sets starting-phase for analysis.
E-PHASE	4	Sets end-phase for analysis.
NORMALISE	1	0: Preference matrix not normalised. 1: Matrix is row-normalised.
AVERAGE	1	0: Average subject's scores are computed in starting phase only. 1: Average subject's scores are computed at all phases.
DATA TYPE	0	0: Data are ranks: first stimulus most preferred. 1: Data are ranks: first stimulus least preferred. 2: Data are scores: highest value = most preferred. 3: Data are scores: lowest value = most preferred.
FIT	0	0: Linear fit. 1: Monotone fit: no ties. 2: Monotone fit: secondary approach to ties. 3: Monotone fit: primary approach to ties.
MATFORM	0	0: Input configuration punched stimuli (rows) by dimensions (columns). 1: Input configuration punched dimensions (rows) by stimuli (columns).
ORIGINAL	0	0: At each new phase begin with fitting values from previous phase. 1: At each new phase return to original preference matrix. (Only applicable if FIT(0))

Keyword	Default Value	Function
KEEP	0	0: Stimulus configuration becomes the canonically reotated space from phase I. 1: Return to original configuration.
CRITERION	0.005	Sets the criterion for terminating iterations. (Not applicable if FIT(0))
GENERATE	0	0: Configuration is generated from the double-centred preference matrix. 1: Configuration is generated after row-means of preference matrix are extracted. 2: Configuration is generated from preference matrix with standardised rows. 3: Configuration is generated from preference matrix with both row- and column-means extracted.

3.2 NOTES

1. In the parameters S-PHASE and E-PHASE, the hyphens are significant characters.

2. The preference data are submitted to the program as a matrix of F-type numbers.

3. The data matrix must have subjects as rows and stimuli as columns.

4. Only one dimensionality per run (task) may be analysed by PREFMAP.

5. Program limits:

Maximum number of subjects	= 49
Maximum number of stimuli	= 100
Maximum number of dimensions	= 5

3.3 PRINT, PLOT AND PUNCH OPTIONS

Since the output from PREFMAP is extensive, the form of the PRINT, PLOT and PUNCH commands has been modified to limit the amount of information generated.

The format of the commands remain the same, viz:

col 1	col 16	
PRINT PLOT PUNCH	ALL ALLBUT EXCEPT	keyword(arg.), keyword(arg.)...

In other MDS(X) programs the argument takes the form of a number (or numbers) which indicate the dimensionality (dimensionalities) for which the particular output specified by the keyword is required.

In PREFMAP the analysis is performed in only one dimensionality per task and thus the argument refers to the phase (or phases) for which the particular output is required. Obviously then, the argument will consist of numbers between 1 and 4 only.

Specification of ALL on the PRINT command will generate a detailed set of output for all phases from S-PHASE to E-PHASE. This is detailed and voluminous and its use is not recommended. Default options, as usual provide the essential information.

3.3.1 PRINT options

Both the original preference matrix and the input stimulus configuration are echoed by means of the PRINT DATA command.

Option	Form	Description
GENERATED	p × r	If the program generates the stimulus configuration from the preference matrix then this is printed.
STIMULI	p × r	The 'solution' matrix of stimulus coordinates is printed. This will differ at each phase unless KEEP CONFIG is specified.
SUBJECTS	N × r	The matrix of subject locations is printed. For phases I-III the matrix is of point coordinates. For phase IV direction cosines are printed.
ROTATIONS	N matrices each r × r	Phase I only. For each subject the cosines of the angles through which (s)he rotates each dimension is printed.
WEIGHTS	N × r	Phases I-III only. The weights which each subject assigns to each dimension is printed.
COMPOSITE	N matrices each r × r	Phase I only. The composite transformation matrix for each subject is printed. This is the matrix which transformed the original space into the weighted and rotated 'private' space.
IDEAL	N × r	Phase I and II only. The coordinates of the subject ideal points in the 'private' spaces (i.e. the space as rotated and/or weighted) is printed.
PRIVATE	N matrices each p × r	Phases I and II only. For each subject the coordinates of the stimulus with respect to the axes of the rotated and/or weighted space are printed.

Option	Form	Description
BETA	-	The regression weights for each subject are printed.
NORMALISED	N × p	This keyword takes no argument and the matrix of preference values as normalised by the program will be printed once only.
FITTING	N × p	For the linear procedure the estimated values are printed. For the non-linear procedure the fitting values are printed.
DISTANCES	N × p	Phases I-III: The matrix of squared distances between each subject and the stimuli is printed. Phase IV: The projections of the stimulus points onto each vector are printed.
RESIDUALS	N × p	The residual values (DISTANCES - FITTING) are printed.
JOINT	N × r p × r	Prints both the subset and stimuli final co-ordinates.

By default, the following are printed

STIMULI
SUBJECTS
ROTATIONS (phase I)
IDEAL (phases I-II)
Within- and between-phase F-tests
Individual correlation of data to solution

3.3.2 PLOT options

Option	Description
INITIAL	This keyword takes no argument. The configuration of stimulus points as input is plotted once only.
STIMULI	The configuration of stimulus points at each selected phase is plotted. There will be r(r-1)/2 two-way plots.
SUBJECTS	The configuration of subject ideal points (vectors) is plotted in the selected phases in the form of r(r-1)/2 plots.
JOINT	The configuration of stimulus points and subject points (vectors) is plotted.
SHEPARD	A Shepard diagram of the data values plotted against the (squared) distances is produced.
CORRELATIONS	A histogram of individual (subject) correlations is produced.
RESIDUALS	A histogram of residual values is produced.

By default only the first two dimensions of the joint space will be plotted.

3.3.3 PUNCH options

Option	Description
SPSS	Each line of the matrix punched contains the following for each phase selected I : the subject index J : the stimulus index INPUT : the corresponding data value NORMALISED: the data value as normalised FITTING : the corresponding fitting value DISTANCE : the squared distance between I and J RESIDUAL : the corresponding residual value in a fixed format.
GENERATED	The matrix of stimulus values if this is generated by the program is punched.
SOLUTION	The two solution matrices are punched, i.e. the matrix of stimulus coordinates and the matrix of subject coordinates (or, in phase IV, cosines).

No punched output is produced by default.

4. EXAMPLE

col 1 col 16

```
RUN NAME                PREFMAP TEST DATA
TASK NAME               DATA FROM OCCUPATIONAL COGNITION PROJECT
PRINT DATA              YES
N OF SUBJECTS           5
N OF STIMULI            16
ITERATIONS              50
DIMENSIONS              2
PARAMETERS              DATA(3), FIT(2), AVERAGE(0)
                        CRITERION (0.001)
PRINT                   SUBJECT, STIM, DIST, FITT
INPUT FORMAT            (11X, 2F12.5)
READ CONFIG
     <configuration of stimulus points here>
INPUT FORMAT            (16F3.0)
READ MATRIX
     <subjects' preferences here>
COMPUTE
FINISH
```

BIBLIOGRAPHY

Carroll, J.D. (1968) Generalisation of canonical correlation analysis to three or more sets of variables. Proceedings of the 7th Annual Convention of the APA, 227-228.

Carroll, J.D. (1972a) 'Individual differences and multidimensional scaling', in R.N. Shepard, A.K. Romney and S.B. Nerlove (eds), Multidimensional scaling: theory and applications in the behavioral sciences, vol.1.

Carroll, J.D. (1972b) 'Review of Delbeke (1968)', Psychometrika, 35, pp 178-281.

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

Carroll, J.D. and J-J Chang (1967) Relating preference data to multidimensional scaling solutions via a generalization of Coombs' unfolding model, Murray Hill, N.J., mimeo.

Carroll, J.D. and J-J Chang (1971) An alternative solution to the "metric unfolding problem". Paper presented at the Psychometric Society Meeting, April 1971, St. Louis, Missouri.

Kruskal, J.B. (1964a) Multidimensional scaling by optimizing goodness-of-fit to the nonmetric hypothesis, Psychometrika, 29, 1-27.

Kruskal, J.B. (1964b) Nonmetric multidimensional scaling: a numerical method, Psychometrika, 27, 115-129.

van Schuur, W.H. (1977) 'Preference mapping: an introduction to Carroll and Chang's PREFMAP', Edinburgh: Program Library Unit IURCS Report No.39.

APPENDIX 1: RELATION OF PREFMAP TO OTHER PROGRAMS

No program implements the exact equivalent of the PREFMAP phases. A number of programs and packages such as ALSCAL 4 give the user the option of fixing certain points in the configuration, and this is equivalent to using an external stimulus configuration in PREFMAP. By such options it is possible to set up ALSCAL 4 and KYST to perform the equivalent of phase III (External Unfolding by the simple distance model) and ALSCAL 4 to perform a close analogy to phase II (External Weighted Unfolding). But no equivalents of phases I and IV exist.

APPENDIX 2: TECHNICAL DESCRIPTION OF THE PREFMAP ALGORITHM

This appendix is based on Carroll (1972) which is used with permission. We begin by detailing the linear model in all four phases.

A2.1 The general model (Phase I)

Define

$$\underset{\sim}{X} \equiv \{x_{ja}\} \qquad \begin{matrix} j = 1,\ldots,p \\ a = 1,\ldots,r \end{matrix} \tag{1}$$

to be a matrix of coordinates for p stimulus points in r dimensions, and

$$\underset{\sim}{S} \equiv \{s_{ij}\} \qquad \begin{matrix} i = 1,\ldots,N \\ j = 1,\ldots,p \end{matrix} \tag{2}$$

to be the matrix of preference scale values for N individuals over the p stimuli. We assume that the smaller scale values represent the more preferred stimuli.

We further define

$$\underset{\sim}{Y} \equiv \{y_{ia}\} \qquad \begin{matrix} i = 1,\ldots,N \\ a = 1,\ldots,r \end{matrix} \tag{3}$$

to be the matrix of individuals' ideal points in r dimensions.

In the linear case we assume that the scale values are linearly related to the (square of the) Euclidean distance, i.e.

$$s_{ij} = a_i d^2_{ij} + b_i + e_{ij} \tag{4}$$

where a_i is one arbitrary constant ($a_i \geq \emptyset$)

b_i is the slope (regression) coefficient for subject i

and e_{ij} is an error term.

In this model we assume that both $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are transformed by an orthogonal matrix $\underset{\sim i}{T}$. and the squared distances are computed from these transformed values.

We define

$$\underset{\sim}{X}^{*} = \underset{\sim}{X}\ \underset{\sim i}{T}. \tag{5}$$

$$\underset{\sim}{Y}^{*} = \underset{\sim}{Y}\ \underset{\sim i}{T}. \tag{6}$$

and

$$d_{ij}^{2} = \sum_{a}^{r} w_{ia}\ (x_{ja}^{*} - y_{ia}^{*})^{2} \tag{7}$$

Equation (7) may be given in matrix form as

$$\begin{aligned} d_{ij}^{2} &= (\underset{\sim}{X}^{*} - \underset{\sim}{Y}^{*})\ \underset{\sim i}{W}.\ (\underset{\sim}{X}^{*} - \underset{\sim}{Y}^{*})' \\ &= \underset{\sim}{X}^{*}\ \underset{\sim i}{W}.\ (\underset{\sim}{X}^{*})' - 2\underset{\sim}{Y}^{*}\ \underset{\sim i}{W}.\ (\underset{\sim}{X}^{*})' + \underset{\sim}{Y}^{*}\ \underset{\sim i}{W}.\ (\underset{\sim}{Y}^{*})' \end{aligned} \tag{8}$$

where $\underset{\sim i}{W}.$ is a diagonal matrix whose non-zero entries are the weights w_{ia}.

Substituting (5) and (6) in (8) we get

$$d_{ij}^{2} = \underset{\sim}{X}\ \underset{\sim}{T}\ \underset{\sim i}{W}.\ \underset{\sim}{T}'\underset{\sim}{X}' - 2\underset{\sim}{Y}\ \underset{\sim}{T}\ \underset{\sim i}{W}.\ \underset{\sim}{T}'\underset{\sim}{X}' + \underset{\sim}{Y}\ \underset{\sim}{T}\ \underset{\sim i}{W}.\ \underset{\sim}{T}'\underset{\sim}{Y}' \tag{9}$$

Define

$$\underset{\sim}{R}^{*} = \underset{\sim}{T}\ \underset{\sim i}{W}.\underset{\sim}{T}' \tag{10}$$

and letting c_{i}^{*} represent the last term in(9) (since this is a constant w.r.t. $\underset{\sim}{X}$.). we get

$$d_{ij}^{2} = \underset{\sim}{X}\ \underset{\sim}{R}^{*}\underset{\sim}{X}' - 2\ \underset{\sim}{Y}\ \underset{\sim}{R}^{*}\underset{\sim}{X}' + c_{i}^{*} \tag{11}$$

Substituting (11) in (4) and letting

$$\underset{\sim}{R} = a_{i}\ \underset{\sim}{R}^{*} \tag{12}$$

and

$$\underset{\sim}{B} = -2\,\underset{\sim}{Y}\,\underset{\sim}{R} \tag{13}$$

we have

$$s_{ij} \approx \underset{\sim}{X}\,\underset{\sim}{R}\,\underset{\sim}{X}' + \underset{\sim}{B}\,\underset{\sim}{X}' + c_i \tag{14}$$

which is the equation of a general quadratic function of $\underset{\sim}{X}$ and which may be solved by treating the problem as a multiple linear regression problem.

The regression coefficient for x_a will be an estimate of b_{ia}, the regression coefficient for the dummy x_a^2 (i.e. $x_a \cdot x_{a'}$, $(a \leq a')$) will be an estimate of r^i_{aa}, and one half the regression coefficient of $x_a \cdot x_{a'}$, $(a < a')$ will be an estimate of $r^i_{aa'}$. Having thus solved for $\underset{\sim}{R}$ and $\underset{\sim}{B}$

$$\underset{\sim}{Y} = -\tfrac{1}{2}\,\underset{\sim}{B}\,\underset{\sim}{R}^{-1} \tag{15}$$

which is derived from (13).

Furthermore $\underset{\sim}{T}$ is the transpose of the matrix of characteristic vectors of $\underset{\sim}{R}$ and $\underset{\sim i}{W}$ is the diagonal matrix of latent roots.

A2.2 Phase II

Phase II is a special case of phase I in that the transformation matrix is restricted to identity transformation. Thus, in examples (10) and (12) $\underset{\sim}{R}$ is simply $a_i\,\underset{\sim}{W}$ i.e. a diagonal matrix whose non-zero entries are proportional to the weights. Thus, corresponding to (14) for phase II we have (absorbing a_i into w)

$$s_{ij} \approx \underset{\sim}{X}\,\underset{\sim i}{W}\,\underset{\sim}{X}' + \underset{\sim}{B}\,\underset{\sim}{X}' + c_i \tag{16}$$

while (13) becomes

$$\underset{\sim}{B} = -2\,\underset{\sim}{Y}\,\underset{\sim i}{W} \tag{17}$$

so that we may solve for $\underset{\sim}{Y}$ by the equation

$$\underset{\sim}{Y} = -\tfrac{1}{2} \underset{\sim}{B}\, \underset{\sim i}{W}^{-1} \tag{18}$$

which is particularly simple.

We may estimate B and W in phase II in much the same way as in phase I except that no cross-product terms are required. Also it is not necessary to find the characteristic roots of any equation since

$$y_{ia} = \tfrac{1}{2} b_{ia} / w_{ia} \tag{19}$$

A2.3 Phase III

Phase III is the simple unfolding model (cf. MINIRSA) except that it allows negative weights.

Phase III is equivalent to phase II when all $w_{ia} = \pm 1$ Since these are independent of i we define

$$\underset{\sim}{U} \equiv \{u_{aa'}\} \qquad a,a' = 1,\ldots,r \tag{20}$$

$$u_{aa'} = \begin{cases} \pm 1 & \text{iff } a = a' \\ 0 & \text{otherwise} \end{cases}$$

The equation corresponding to (14) or (16) in phase II becomes

$$s_{ij} \approx a_i \underset{\sim}{X}\, \underset{\sim}{U}\, \underset{\sim}{X}' + \underset{\sim}{B}\, \underset{\sim}{X}' + c_i \tag{21}$$

or, in summation form

$$s_{ij} \quad a_i \left[\sum_a^r u_a x_{ja}^2 \right] + \sum_a^r b_{ia} x_{ja} + c_i \tag{21a}$$

The term in brackets is independent of i and may be regarded as a single pseudo-independent variable in the regression equation.

Solving the equation in r + 1 independent variables gives us, immediately, a_i and b_{ia}.

The coordinates of y then become

$$y_{ia} = \tfrac{1}{2} (b_{ia} / a_i u_a) \tag{22}$$

A2.4 Phase IV

In phase IV the equation corresponding to (14), (16), (21) is

$$s_{ij} \approx a_i \underset{\sim}{Y} \underset{\sim}{X}' + c_i \tag{23}$$

where $\underset{\sim}{Y}$ is the vector of projections.

If we define

$$\underset{\sim}{B} \equiv a_i \underset{\sim}{Y} \tag{24}$$

then (23) becomes

$$s_{ij} \approx \underset{\sim}{B} \underset{\sim}{X}' + c_i \tag{25}$$

so that the regression equation contains only linear terms. Having estimated $\underset{\sim}{B}$ then

$$y_{ia} = b_{ia} / \sqrt{\sum_{a'}^{r} b_{ia'}^2} \tag{26}$$

A2.5 The quasi-non-linear procedure

If we now relax the assumption of linearity we may replace (4) by

$$\hat{s}_{ij} = d_{ij} + e_{ij} \tag{27}$$

where

$$\hat{s}_{ij} + M_i (s_{ij}) \tag{27a}$$

where M_i is a monotone (non) decreasing function.

PREFMAP is a "quasi-non-metric" procedure insofar as it first estimates the linear parameters and then uses these as initial estimates for a Kruskal's (1964) monotone regression procedure.

APPENDIX 3: THE BETA WEIGHTS

At phase IV the program estimates

$$s_{ij} \cong \sum_{a}^{r} b_{ia} x_{ja} + c_i$$

where

s_{ij} is the preference of subject i for stimulus j

x_{ia} is the coordinate of stimulus j on the a'th dimension of the r-dimensional space.

The r+1 beta weights are printed in the order

$$c_i, b_{i1}, b_{i2}, \ldots, b_{ir}$$

At phase III the program estimates

$$s_{ij} \cong a_i \sum_{a}^{r} x_{ja}^2 + \sum_{a}^{r} b_{ia} x_{ja} + c_i$$

The r+2 beta weights are printed in the order

$$c_i, b_{i1}, b_{i2}, \ldots, b_{ir}, a_i$$

At phase II the program estimates

$$s_{ij} \cong \sum_{a}^{r} w_{ia} x_{ja}^2 + \sum_{a}^{r} b_{ia} x_{ja} + c_i$$

The 2r+1 weights are printed in the order

$$c_i, b_{i1}, b_{i2}, \ldots, b_{ir}, w_{i1}, w_{i2}, \ldots, w_{ir}$$

At phase I the program estimates

$$s_{ij} \cong \sum_{a}^{r} \sum_{a'}^{r} r_{aa'}^{i} (x_{ja} x_{ja'}) + \sum_{a}^{r} b_{ia} x_{ja} + c_i$$

The $(r^2+3r+2)/2$ beta weights are printed in the order

$$c_i, b_{i1}, b_{i2}, \ldots, b_{ir}, r_{11}, r_{12}, \ldots, r_{1r}, r_{2r}, \ldots, r_{rr}$$