12. PINDIS (Procrustean INdividual DIfferences Scaling)

1. OVERVIEW

Concisely: PINDIS (Procrustean INdividual DIfferences Scaling) provides an internal analysis of a set of configurations by a procrustean fitting model which uses a similarity transformation of the data.

Alternatively, following the categorisation suggested by Carroll and Arabie (1979) the program may be described as follows:

Data: A set of configurations:	Model:
Three-way	P0: Similarity
Three-mode	P1: Dimensional weighting
Non-symmetric	P2: Dimensional weighting and rotation
Dyadic	
Ratio level of measurement	P3: Perspective (vector)
Matrix conditional	P4: Perspective and translation
Incomplete (missing dimensional co-ordinates)	P5: Double weighted
	Two spaces
One replication	Internal/External

1.1 ORIGIN, VERSIONS AND ACRONYMS

PINDIS was developed by Lingoes and Borg at the University of Michigan. A number of early versions of the program exist. The present program was adapted from the 1975 version which is documented in Borg (1977).

1.2 PINDIS IN BRIEF

PINDIS provides means of dealing with the question of individual differences. It takes as input a set of configurations obtained from previous scaling analyses. From these it derives a 'centroid configuration' which is an optimal fit to the input configurations.

The way in which these are related to each other differs in the six models which make up the program.

1.3 THE RELATION OF PINDIS TO OTHER PROGRAMS IN MDS(X)

PINDIS differs from all other programs in the MDS(X) library in accepting configurations as data. However, most of the models have affinities with other programs:

P0 Procrustean rotation is not related to any other MDS(X) program.

P1 and P2 are distance models.

P1 (Dimension weighting) is very similar to INDSCAL in permitting individual weighting of fixed dimensions. The parallels are discussed in Borg and Lingoes (1978).

P2 (Rotated and weighted distance) is very similar to the Carroll and Chang's IDIOSCAL model in permitting individual rotation of the dimensions followed by differential weighting of the dimensions.

P3 and P4 are weighted vector models.

P5 is a double weighting (dimensional and vector weighting) model.

None of them has a parallel in any other program.

2. DESCRIPTION

2.1 DATA

The PINDIS program takes as its input data a number of configurations. These will normally be the result of some previous scaling analysis, although any technique giving dimensional output is suitable. The number of points in each of the configurations should be the same although the dimensionalities of the spaces may differ.

The intuitively most apparent form of the data might be a three-way analysis where each configuration results from the scaling of a given individual's judgements of a set of stimuli.

The maximum number of dimensions in any one configuration is given on the DIMENSIONS card, the number of configurations on the N OF SUBJECTS card. The number of points in the configuration is given on the N OF STIMULI card and the data are read by the READ CONFIGS control card. These may be input either stimuli (rows) by dimensions (columns) or vice versa (in which case MATFORM(1) should be specified on the PARAMETERS card). The INPUT FORMAT card should read the longest row of the configurations.

2.2 THE MODEL

PINDIS stands for Procrustean INdividual DIfferences Scaling, and consists of a set of six models for dealing with the question of how different configurations are to be related to each other. In psychological terms, the general assumption is that each subject is systematically distorting a common, shared structure. The configuration obtained from a given individual is thought of as being a systematic distortion of a "master" configuration, the 'group space', and the program seeks both to derive this 'group space' and to relate the given configuration to it. The program contains six models which define different modes of distortion.

It will be seen that it is quite possible that different subjects will be best fit by different models. The first main output of PINDIS is an estimate of this shared aggregate group space or centroid configuration as it is known in the program. This is normally generated by the program from the input configurations in the manner described below but it is possible to input a fixed reference configuration and then use PINDIS for an external analysis (see 2.3.1).

2.2.1 The basic model (PO): Similarity transformation (Unit weighting)

The basic "model" of the PINDIS is simple Procrustean fitting and depends on the fact that MDS solutions are unique up to translation, rotation and reflection and uniform stretching or shrinking rescaling of axes. This is simply to say that in a configuration from, say, MINISSA, the significant information is contained in the relative distances between the stimulus and, in particular:

1. that the position of the origin is arbitrary and may be moved (translated) without destroying any of the significant information in the solution. (This is not the case for factor analytic solutions (see 2.3)).

2. that the axes of the configuration are in an arbitrary, though possibly convenient, position and may be (rigidly) rotated without destroying the salient information in the solution.

3. that a configuration may be reflected without loss of information. Intuitively this means that a configuration may come out of an analysis "back-to-front". Geometrically reflection is merely a special case of rotation.

4. that the actual numbers assigned to the distances are not significant information but may be made uniformly bigger or smaller at will. Intuitively, this means that the actual configuration may be enlarged or reduced so long as this process is uniform.

These operations, translation, rotation (with which we include reflection) and rescaling (uniform stretching etc.) comprise a <u>similarity transformation</u> and are known in the model as the "permissible transformations" in that changing a configuration by any (or all) of them gives a configuration which contains neither more nor less information than the original.

The program's first step is to take each pair of configurations in turn and, by applying the permissible similarity transformations, move them into maximum conformity with each other. Having done this, the program has effectively eliminated any differences in the configurations due to the conventions of the program producing them and has left the substantive differences - the differences due to random error and differential cognition. The centroid configuration is formed simply by taking the average position of each point over all the configurations. The model at this stage implies that in reporting their perceptions, subjects make no <u>systematic</u> distortions to the group space (the centroid).

The communality of each configuration to the centroid is then calculated. This may be regarded as the proportion of variance (r^2) in that particular configuration explained by the centroid.

The higher order models allow that subjects may systematically distort this centroid configuration. It is the mode of distortion which differs in these models.

2.2.2 The dimensional salience or dimension weighting (P 1)

In this case the mode of distortion is analogous to that of the INDSCAL model in that subjects, in arriving at their perceptual spaces, are thought of as applying differential weights to the dimensions of the group space (the centroid). Substantively this amounts to saying that subjects will attach greater salience to certain (fixed) aspects of the difference between stimuli than to others or that they will be prone to make finer distinctions on some criteria over others.

The user may choose whether these differential weights are to be applied to the centroid obtained at PO or whether this configuration is to be rotated to some optimal position before the weights are applied. The default option allows for this latter course and may be expected to result in substantively more interpretable solutions. If, however, the user wishes to fix the centroid after PO or has input a hypothesis configuration with 'meaningful' axes, then ROTATE(0) should be specified as the PARAMETERS card.

The communality of the centroid to each of the input matrices is then calculated. This and the similar values obtained from higher models should be compared to the value from PO which is treated as the baseline from which the more complex models are assessed. Final choice of the preferred explanatory model is made on the basis of the increase in the fitting value (r^2) which takes into account the fact that at each stage the number of free parameters increases dramatically.

2.2.3 Dimensional salience with idiosyncratic orientation (P2)

In this model each subject is thought of as distorting the centroid by first rotating the axes of the configuration to his/her own preferred orientation and then applying differential weights to these new axes. (It should be noted that if ROTATE(0) has been specified then this solution will be identical to P1).

The substantive interpretation of this model is that subjects are not only affording differential salience to the same criteria but also using different criteria.

In models P1 and P2 the mode of distortion which took the centroid into the subject configurations was essentially a dimensional weighting. In models P3 and P4 the distortions are applied directly to the actual stimulus points, which are considered as vectors from the origin of the space.

2.2.4 Perspective model with fixed origin (vector weighting) (P3)

Let us remind ourselves that the aim of the PINDIS procedure is to get the points of the centroid configuration (the group space) as close as possible to each of the individual input configurations in turn. This model seeks to do this by differentially stretching or shrinking each stimulus vector drawn from the origin of the space. What does this mean ? Essentially the process may be conceived of in this way. Take a subject configuration and plot it on top of the centroid so that the origin and axes coincide. Now draw a line to connect the origin with a particular stimulus point in the centroid configuration and produce it beyond both the point and the origin. The point on this line which is nearest to the corresponding point in the subject configuration is the point we are looking for.

The substantive justification for this model relies on the axes and origin of the space being interpretable/meaningful and asserts that the significant information in the configuration is the balance (actually the ratio) between the coordinates on the constituent axes. It is sometimes called the "unscrambling" model since a weight applied to a stimulus vector moves the position of that stimulus in the space.

2.2.5 The perspective model with idiosyncratic origin (P4)

A moment's thought will convince the user that although the actual orientation of the axes of the configuration do not affect the direction of the stimulus vector, the position of the origin is crucial. The idiosyncratic vector model additionally allows the subjects to move the origin of the centroid space to an idiosyncratic position before the vector weighting operations are performed.

If the centroid configuration has a rational origin and it does not make sense to shift it about in this manner, then the user should specify TRANSLATE(0) on the PARAMETERS card (see also 2.3).

2.2.6 The double weighted (dimension and vector weighting) model (P5)

This model allows both dimensional and vector weighting simultaneously. Although the number of free parameters in this model is large, it has been found that the goodness-of-fit of this particular model is often surprisingly low. This may indicate that the geometrical processes which define it have little psychological rationale (it is largely within the psychological field that it has been tried) though other substantive applications may find one.

The double weighting solution may be suppressed by specifying SUPPRESS(1) on the PARAMETER card.

2.2.7 Some general points

The user will have noticed that the centroid configuration is different in each model.

For each of the models the program calculates the communality between the centroid and each of the subject configurations. Choice of a particular model should be made by comparing this value for each subject for each model against the communality at PO. Some improvement should manifest itself as the number of free parameters increases. If a higher level model has virtually the same communality (for a given subject) as a lower one then obviously parsimony suggests that the lower one be preferred.

The number of parameters estimated in each model in finding a given subject configuration is a function of the dimensionality of the configuration (r) and the number of stimulus points (p).

P0 = 0	(simply permissible transformations)
P1 = r	(dimension weights)
P2 = r + (r(r-1)/2)	(dimension weights and pairwise rotation certificate)
P3 = p	(stimulus vector weights)
P4 = p + r	(stimulus vector weights and r-dimensional origin)
P5 = p + (p + r)	(dimension weights, stimulus vector weights and origin).

The models form a semi-lattice thus:

(distance) (vector)

P2 P5 P4

P1 P3

PO
(similarity)

2.3 FURTHER FEATURES

2.3.1 External analysis

The user may wish to use the PINDIS program to effect an external analysis by inputting, as well as the subject configurations, a fixed centroid configuration which may be an a priori arrangement of points or the result of a previous MDS or other dimensional analysis. This configuration is input to the program by means of the READ HYPOTHESIS control card which is peculiar to PINDIS, under its associated INPUT FORMAT card. This configuration will form the centroid at PO and will be rotated, weighted, etc., in the other models and users are urged to pay particular attention to the values given to the ROTATE (see 2.2.2 and 2.2.3) TRANSLATE (see 2.2.5) and ORIGIN (see below) parameters to ensure that they do not violate the logic of the configuration.

2.3.2 The use of the ORIGIN parameter

We note at 2.2.4 the importance of the position of the origin of the space in the weighted vector models. One way of making substantive sense of vector weighting is by moving the origin to a substantively meaningful position rather than at an arbitrary centroid and considering each of the other points as directions of distinction from that point. Consider this hypothetical example. Suppose we were interested in the perceptions of political parties. We might take the configurations belonging to members of a particular party and place the origin of the space at the point representing that party. The distance to the other party points (the length of the stimulus vectors) is then proportional to the perceived difference between the party of affiliation and the others but the direction will also have significance in representing the mode of difference (say right vs. left, populist vs. elitist). It may very well be the case that there is virtual consensus over the modes of difference, i.e. the ways in which the parties differ but disagreement over how different they are. Some right wing Conservatives may, for instance, be very anxious to dissociate themselves from the

National Front and while acknowledging the fact that the N.F. is more right-wing, will insist on the difference between the Front and the Tories being made as large as between, say, the Tories and the Labour party. Other members of the Conservative party, of a more moderate bent, might be less neurotic about admitting the similarity between the two. In this case, the weighted vector model provides a feasible model of the differences between the two groups.

The user may use this option by specifying the number of the point to be regarded as the origin as the argument to the ORIGIN parameter.

3. PARAMETERS

3.1 LIST OF PARAMETERS

Keyword	Default	Description
SUPPRESS	1	0: Double-weighted solution (P5) is performed. 1: Double-weighted solution (P5) is suppressed.
ROTATE	1	0: Idiosyncratic rotations of the centroid are not allowed, i.e. P2 is not performed. 1: Idiosyncratic rotations are allowed.
TRANSLATE	0	0: No translation of the origin allowed, i.e. P4 is not performed. 1: Translation of origin to idiosyncratic position is allowed.
ORIGIN	0	0: The origin is situated at the centroid of the space <any positive integer> gives the number of the point to be regarded as the origin.
MATFORM	0	0: The input configurations are input stimuli (rows) by dimensions (columns) 1: The input configurations are input dimensions (rows) by stimuli (columns).

3.2 NOTES

1. The READ CONFIGS control card is obligatory in PINDIS.

2. The READ MATRIX card is not valid with PINDIS.

3. Maximum number of dimensions = 6
   Maximum number of stimuli = 50
   Maximum number of configurations = 50

3.3 PRINT, PLOT AND PUNCH OPTIONS

The general format for printing, plotting and punching output is described in the Overview. The particular options for PINDIS are as follows:

3.3.1 PRINT options

Option	Form	Description
CENTROID	p × r	The centroid configuration is printed at each phase.
SUBJECTS	N(p × r)	The subject matrices are printed at each phase.

Both of these are produced by default.

3.3.2 PLOT options

Option	Description
CENTROID	The centroid configuration is plotted at each phase.
SUBJECTS	The subject configurations at each phase are plotted.

Both configurations are plotted by default.

3.3.3 PUNCH options

Option	Description
CENTROID	The coordinates of the centroid configuration are output - a fixed format.

No output is punched by default.

4. EXAMPLES

4.1 TEST RUN

```
col 1                 col 16

RUN NAME              RUN OF TEST DATA FOR PINDIS
PRINT DATA            YES
NO OF SUBJECTS        5
NO OF STIMULI         16
DIMENSIONS            3
COMMENT
                      FIVE CONFIGURATIONS ARE TO BE INPUT.  EACH HAS SIXTEEN
                      POINTS IN THREE DIMENSIONS
PLOT                  ALL
INPUT FORMAT          (14X,3F12.5)
COMMENT
                      ALL PARAMETERS WILL ASSUME DEFAULT VALUES
INPUT MEDIUM          DISC
READ CONFIGS
COMMENT
                      THE DATA IS IN A SEPARATE DISC FILE DEFINED AS
                      CHANNEL 4
COMPUTE
REWIND                DISC
TASK NAME             RUN WITH HYPOTHESIS MATRIX
COMMENT
                      NOTE THAT THOUGH THE DATA HAVE BEEN 'REWOUND' ALL OTHER
                      PARAMETERS MUST BE (RE)SET
N OF SUBJECTS         5
NO OF STIMULI         16
DIMENSIONS            3
PARAMETERS            TRANSLATE(1)
INPUT MEDIUM          CARD
INPUT FORMAT          (1X,3F8.0)
READ HYPOTHESIS

   <the hypothesis (target) matrix follows here>

INPUT MEDIUM          DISC
READ CONFIGS
COMMENT
                      THE DATA WILL NOW BE RE-READ FROM THE DISC FILE
COMPUTE
FINISH
```

BIBLIOGRAPHY

Borg, I. (1977) Representation of individual differences, in J.C. Lingoes (ed.) Geometric representations of relational data, Mathesis Press, Ann Arbor, Michigan.

Borg, I. and J.C. Lingoes (1977) A direct transformational approach to multidimensional analysis of three-way data matrices, Zeit. F. S-Psych., 8, 98-114.

Gower, J.C. (1975) Generalized procrustes analysis, Psychometrika, 40, 33-51.

Lingoes, J.C. and I. Borg (1976) Procrustean individual difference scaling, J. Market.Research, 13, 406-407.

Lingoes, J.C. and I. Borg (1977) Optimal solutions for dimension and vector weights in PINDIS, Zeit. F.S-Psych., 8.

Lingoes, J.C. and I. Borg (1978) A direct approach to individual differences scaling using increasingly complex transformations, Psychometrika, 43.

Lingoes, J.C. and P.H. Schonemann (1974) Alternative measures of fit for the Schonemann-Carroll matrix fitting algorithm, Psychometrika, 39, 423-427.