4. INDSCAL–S (INDividual differences SCALing)

1. OVERVIEW

Concisely: INDSCAL-S (INDividual Differences SCALing: Short version) provides internal analysis of a three-way data matrix consisting of a set of (dis)similarity matrices, by a weighted distance model using a linear transformation of the data.

Following the categorisation developed by Carroll & Arabie (1979) the program may be described as:

Data:	Three-way	Model:	Weighted Euclidean
	Two mode		Two sets of points
	Dyadic		Internal/External
	Linear		
	Unconditional		
	Complete		
	One replication		

1.1 ORIGIN, VERSIONS AND ACRONYMS

INDSCAL was developed by J.D. Carroll and J.J. Chang of Bell Telephone Laboratories. The original INDSCAL program performed two types of analysis: INDIFF, which is the most commonly used part of the program and often referred to simply as INDSCAL, and CANDECOMP. It is this former analysis (the INDIFF option) which comprises the present program (INDSCAL-S). The CANDECOMP option appears as a separate program within MDS(X). The present program is specially adapted from the 1972 version of INDSCAL.

A quasi non-metric INDSCAL known as N-INDSCAL exists but is though to be unstable.

In what follows we shall follow the convention of referring to the model as INDSCAL and this program as INDSCAL-S.

1.2 INDSCAL IN BRIEF

INDSCAL was originally developed to explain the relationship between subjects' differential cognition of a set of stimuli. Suppose that there are N subjects and p stimuli. The program takes as input a set of N matrices each of which is a square symmetric matrix (of order p) of (dis)similarity judgments/measures between the p stimuli. The model explains differences between subjects' cognitions by a variant of the distance model. The stimuli are thought of as points positioned in a 'group' or 'master' space. This space is perceived differentially by the subjects in that each of them affords a different salience or weight to each of the dimensions of the space. The transformation which is assumed to take the data into the solution is a linear one.

1.3 RELATION TO OTHER MDS(X) PROGRAMS

The INDSCAL model is a special case of the PINDIS hierarchy of models.

INDSCAL is also a special case of CANDECOMP where the second and third 'way' of the data matrix are identical. In the Carroll-Wish terminology INDSCAL is three way, two mode; CANDECOMP three way, three mode (actually N-way, N-mode where $3 \leq N \leq 7$).

2. DESCRIPTION

2.1 DATA

Imagine that a group of subjects is asked to assess the dissimilarity between a set of objects. It is inevitable that these judgments will differ. The problem then arises of the relationship between the sets of judgments. The INDSCAL model assumes that subjects are systematically distorting a shared space in arriving at their judgments and it seeks to reconstruct both the individual private (distorted) spaces and the aggregate "group" space.

There is no reason why the judgments of (dis)similarity should come from "real" individuals. They may be different occasions, methods, places, groups etc., in which case they are often referred to as 'pseudo-subjects'.

The mode of distortion which the INDSCAL model proposes is this. The basic, shared configuration (known as the Group Space in INDSCAL) has a given number of fixed dimensions. In making their dissimilarity estimates different subjects are thought of as attaching different salience to different dimensions. Thus, for instance, in judging the differences between two houses an architect might primarily distinguish between them in terms of style, whereas a prospective buyer might attach relatively little weight to that aspect but a great deal to the difference in price.

2.1.1 Example

Suppose we were interested in how people perceive the distances between 6 different areas of a city, and asked them to give their estimates of the distance between each of the pairs of areas (fifteen in all). These estimates we collect into three lower-triangle matrices as follows:

3.6					Subject 1
6.7	9.2				
7.0	3.1	3.1			
6.0	4.1	3.0	3.1		
4.1	5.0	3.6	6.7	4	
5.7					Subject 2
7.3	9.4				
7.1	3.3	4.3			
6.0	4.2	4.2	3.3		
5.7	6.4	4.6	7.3	4	
7.3					Subject 3
9.0	12.0				
9.9	4.3	3.3			
8.4	5.7	3.0	4.3		
4.2	5.8	4.1	9.0	5.6	

The fifteen judgments of each subject are collected into the lower triangle of a square symmetric matrix which would be submitted to INDSCAL-S as shown in section 4.1

2.2 MODEL AND ALGORITHM

The INDSCAL model interprets 'individual differences' in terms of subjects applying individual sets of weights to the dimension of a common 'group' or 'master' space. Hence the main output of an INDSCAL analysis is a 'Group Space' in which the stimuli (in our example, the area locations) are located as points. The configuration of stimuli in this Group Space is in effect a compromise between different individuals' configurations, and it may conceivably describe the configuration of no single individual.

Complementing the Group Space is a 'Subject Space'. This space has the same dimensions as the Group Space but in it each individual is represented as a point, located by the set of co-ordinates which are the values of the numerical '<u>weights</u>' which he assigns to each dimension. These individual weights or saliences are solved for by the program and are its next most important output.

Thus the subject whose individual cognition corresponds exactly with the "group space configuration" - if that subject exists - would be situated in a two-space on a line at 45° between the axes, whereas someone who paid no attention to one of the axes would be situated at zero on that axis.

Having obtained the 'Group Space' and an individual's set of weights, it is often useful to take the Group Space Configuration of stimuli points and transform it into that individual's 'Private Space'. A Private Space is simply the Group Space with its dimensions stretched or contracted by the weights which that subject has assigned to them.

2.2.1.1 Some properties of the INDSCAL model

It should be noted that INDSCAL produces a <u>unique</u> orientation of the axes of the Group Space, in the sense that any rotation will destroy the optimality of the solution and will change the values of the subject weights. Moreover, the distances in the Group Space are <u>weighted</u> Euclidean, whereas those in the private spaces are simple Euclidean. Because of this, it is not legitimate to rotate the axes of a Group Space to a more 'meaningful' orientation, as is commonly done both in factor analysis and in the basic multidimensional scaling model. It has generally been found that the recovered dimensions yield readily to interpretation.

Secondly, each point in the Subject Space should be interpreted as a vector drawn from the origin. The length of this vector is roughly interpretable as the proportion of the variance in that subject's data accounted for by the INDSCAL solution. All subjects whose weights are in the same ratio will have vectors oriented in the same direction. Consequently, the appropriate measure for comparing subjects' weights is the angle of separation between their vectors.

2.2.2 The Algorithm

1. The program begins by converting each subject's dissimilarities into estimates of euclidean distances by estimating the additive constant (see Torgerson 1958; Kruskal 1972).

2. These distance estimates are then double-centred to form a scalar-product matrix.

3. These scalar-products may be considered as the product of three numbers. The first of these will come to be considered as the subject weight. The other two give at this stage two distinct estimates of the value of the stimulus co-ordinates.

4. An initial configuration is input by the user or generated by the program (see 2.3.3).

5. The scalar-products between the points in this configuration are calculated and serve as an initial estimate of the solution parameters.

6. For each scalar-product at each iteration a pair of these three numbers is held constant in turn and the value of the other is estimated.

7. When maximum conformity to the data is reached by this iterative process, the two estimates of the stimulus coordinates are set equal and one more iteration is performed.

8. The matrices are normalised and output as solution.

2.3 FURTHER OPTIONS

2.3.1 Data

Consider again the example given above (section 2.1.1). In it we had three subjects judging six stimuli. Thus each subject generates a lower triangle matrix of five rows if the diagonals are omitted. These are input to the program after the READ MATRIX card sequentially, i.e. the matrix of subject I is followed by that of subject II which is followed by that of subject III, without break, fifteen lines in all.

The program will also analyse other types of data including correlation or covariance matrices. In this case the 'stimuli' will be the variables which are correlated and the 'subjects' perhaps replicative studies.

At the beginning of an INDSCAL analysis each input matrix of similarities, dissimilarities, or distances is converted into a matrix of scalar products. To equalize each subject's influence on the analysis these data are normalized by scaling each scalar products matrix so that its sum of squares equals one. Data input as covariances or correlations are not converted to scalar products and are not normalized in this way, thus it is essential to signal this type of input by means of the DATA TYPE parameter (see Section 3).

2.3.2 Number of dimensions

Some experimentation is generally needed to determine how many dimensions are appropriate for a given set of data. This involves analyzing the data in spaces of different dimensionality. For each space of r dimensions the program uses as a starting configuration the solution in (r + 1) dimensions less the dimension accounting for the least variance. Usually between two and four dimensional solutions will be adequate for any reasonable data set.

2.3.3 Starting configuration

The analysis begins with an initial configuration of stimulus points. This may be supplied by the user and read under a READ CONFIG card. This configuration should contain stimuli coordinates in the maximum dimensionality required.

Alternatively the program can generate a configuration either by a method similar to that used in IDIOSCAL or by picking pseudo-random numbers from a rectangular distribution. If the value of the parameter RANDOM is 0 then the IDIOSCAL procedure is used, otherwise the value is used as a seed to generate the random numbers. Since sub-optimal solutions are not uncommon with this method users are strongly recommended to make several runs with different starting configurations. A series of similar (or identical) solutions may be taken to indicate that a true 'global' solution has been found.

Alternatively, the user may wish to overcome this particular difficulty by submitting, as an initial configuration one obtained from, say, a MINISSA run in which the averaged judgements have been analysed. This method will also reduce the amount of machine time taken to reach a solution.

2.3.4 External analysis

On occasion a user may wish to determine only subject weights for some previously determined stimulus configuration, such as a previous INDSCAL solution, or, some known configuration (as in our example the actual geographical location of the city areas). This option requires that an input configuration be supplied under the READ CONFIG card. The full set of data should be read in under the READ MATRIX card but FIX POINTS should be set to 1 on the PARAMETERS card and the program will then solve only for the subject weights.

2.3.4.1 Large data sets

This option is particularly useful when the user has more data than the program is capable of handling (see 3.2). The user can use the configuration obtained either from a MINISSA analysis of averaged judgments or from an INDSCAL analysis of some judiciously selected subset of subjects and fit to it any number of subjects' weights.

2.3.5 The SOLUTIONS parameters

The axes of the solution correspond to the major direction of variation in the subjects' data. They will not usually correspond to the principal axes of the configuration, in which, the coordinates on the axes are uncorrelated. In the INDSCAL solutions, by contrast, the coordinates will usually be correlated and these correlations are output as the scalar-products matrix for the stimulus configuration. A similar scalar-products matrix is output for the subject space. In this however, it is a dispersion matrix whose diagonal entries are variances, representing the degree to which subject variation is concentrated in that dimension, and whose off-diagonal entries represent the co-variation between dimensions in the subject weights.

If the user wishes to constrain the solution as closely as possible to orthogonality (i.e. in the sense that the correlation between the coordinates is zero) then the parameter SOLUTIONS should be set to 1 on the PARAMETERS card. Users are warned that this will necessarily produce a suboptimal solution.

2.3.6 Negative weights in INDSCAL solutions

There is no interpretation of a negative subject weight in an INDSCAL solution. Nevertheless, from time to time negative values do occur in the subject matrix. If these are close to zero, then the occurrence is likely to be due to rounding error and should be regarded as zero for interpreting the solutions. Large negative values on the other hand suggest a more substantial error or that the model is not appropriate to the data.

2.3.7 Individual correlations as a measure of goodness-of-fit

Being a 'metric' procedure the index of goodness-of-fit of model to data is the correlation between the scalar products formed from the subject's data and those implied by the model. The program outputs a correlation coefficient for each subject and also the average correlation for all subjects and a root-mean-square coefficient which indicates the proportion of variance explained.

2.3.8 The stopping criterion

At step 7 of the algorithm the improvement in correlation is computed. If this is less than the value specified on the CRITERION parameter on the PARAMETERS card, then the iterations are ended. Users should make this value larger if they wish to essay a number of exploratory analyses or to test a number of starting configurations.

3. INPUT PARAMETERS

All parameter keywords may be shortened to the first four letters. All subsequent mis-spellings are ignored.

3.1 LIST OF PARAMETERS

Keyword	Default Value	Function
SOLUTIONS	0	0: Compute all dimensions simultaneously. 1: Compute separate one dimensional solutions.
FIX POINTS	0	0: Iterate and solve for all matrices 1: Solve for subject weights only.
RANDOM	0	Random number seed for generating the initial configuration. (Used when the user does not provide the initial configuration by use of READ CONFIG card).
DATA TYPE	1	0: IDIOSCAL starting configuration. 1: Lowerhalf similarity matrix (without diagonals). 2: Lowerhalf dissimilarity matrix (without diagonals). 3: Lowerhalf euclidean distances (without diagonals). 4: Lowerhalf correlation matrix (without diagonals). 5: Lowerhalf covariance matrix (without diagonals). 6: Full symmetric similarity matrix (diagonals ignored). 7: Full symmetric dissimilarity matrix (diagonals ignored).
CRITERION	0.005	Sets criterion value for termination of iterations.
MATFORM	0	0: Input configuration punched stimuli (rows) by dimensions (columns). 1: Input configuration punched dimensions (rows) by stimuli (columns). Only valid with READ CONFIG.

3.2 NOTES

1. Program limits

Maximum number of dimensions	=	5
Maximum number of stimuli	=	30
Maximum number of subjects	=	30
N OF SUBJECTS × (N OF STIMULI)^2	=	18000
max (N OF SUBJECTS, N OF STIMULI) × maximum no. of dimensions × 3	=	2500

2. The program expects input in the form of real (F-type numbers), and the INPUT FORMAT card should allow for this.

3. The INPUT FORMAT card should read the longest line of the input matrices.

3.3 PRINT, PLOT AND PUNCH OPTIONS

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

3.3.1 PRINT options (output to line printer)

Option	Form	Description
INITIAL	N × r p × r p × r	Three matrices are printed: 1. the initial estimates of the subject weights. 2. & 3. separate estimates of the stimulus configuration.
FINAL	N × r p × r N r × r	Two matrices are printed being the matrix of subject weights and the coordinates of the group space. These are followed by the correlation between each subject's data and solution and the matrix of cross-products between the dimensions.
HISTORY		An iteration by iteration history of the overall correlation. (The final (3) matrices at convergence are also printed).
SUMMARY		Summary of results produced at end of each analyses.

By default only the solution matrices and the final overall correlation are printed.

3.3.2 PLOT options (output to line printer)

Option	Description
INITIAL	The initial configuration may be plotted only if one is input by the user.
CORRELATIONS	The correlations at each iteration are plotted.
GROUP	Up to r(r-1)/2 plots of the p stimulus points.
SUBJECTS	Up to r(r-1)/2 plots of the Subject Space.

By default the Subject and Group Spaces will be plotted.

3.3.3 PUNCH options

Option	Description
FINAL	Outputs the final configuration and the subject correlations in the following order: - each subject is followed by the coordinates of its weight on each dimension; - each stimulus point is followed by its coordinates on each dimension.
CORRELATIONS	The overall correlation at each iteration is output in a fixed format.
SCALAR PRODUCTS	The scalar product matrix is punched.

No punched output is generated by default.

4. EXAMPLES

4.1 TEST RUN

```
col 1                     col 16

RUN NAME                  INDSCAL TEST DATA
TASK NAME                 ...FROM EXAMPLE IN 2.1.1
N OF SUBJECTS             3
N OF STIMULI              6
DIMENSIONS                2
PARAMETERS                CORRELATIONS(1),RANDOM(34551)
COMMENT                   THIS IS THE SET-UP FOR THE EXAMPLE GIVEN.
                          NOTICE THE USE OF THE SHORTENED PARAMETER
                          DESIGNATION AS IN 'DATA(2)'
INPUT FORMAT              (5F3.0)
COMMENT                   IN THE DATA, THE UNDERLINE '_' DESIGNATES
                          A SPACE.
READ MATRIX
_36
_67_92
_70_31_31
_60_41_30_31
_41_50_36_67_40
_57
_73_94
_71_33_43
_60_42_42_33
_57_64_46_73_40
_73
_90120
_99_43_33
_84_57_30_43
_42_58_41_90_56
PRINT                     FINAL, HISTORY
PLOT                      ALL
COMPUTE
FINISH
```

BIBLIOGRAPHY

Bloxom, B. (1965) Individual differences in multidimensional scaling, Princeton University Educational Testing Service Research Bulletin, 68-45.

Carmone, F.J., P.E. Greeen and P.J. Robinson (1968) TRICON: an IBM 360/65 program for the triangularisation of conjoint data, Journal of Marketing Research, 5, 219-20.

Carroll, J.D. (1974) Some methodological advances in INDSCAL, mimeo, Psychometric Society, Stanford.

Carroll, J.D. and P. Arabie (1979) Multidimensional scaling, in M.R. Rozenzweig and L.W. Porter (eds.) 1980 Annual Review of Psychology, pp 607-649, Palo Alto Ca., Annual Reviews.

Carroll, J.D. and J.J. Chang (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition, Psychometrika, 35, 283-319.

Carroll, J.D. and M. Wish (1974) Multidimensional perceptual models and measurement methods, in E.C. Carterette and M.P. Friedman Handbook of Perception, Vol.2, New York: Academic Press (Ch.5 Individual differences in perception).

——— (1975) Models and methods for three way multidimensional scaling, in R.C. Atkinson, D.H. Krantz, R.D. Luce and P. Suppes (eds.), Contemporary Methods in Mathematical Psychology, San Francisco: Freeman.

Coxon, A.P.M. and C.L. Jones (1974) Applications of multidimensional scaling techniques in the analysis of survey data, in C.J. Payne and C.O'Muircheartaigh, Survey Analysis, London: Wiley.

Horan, C.B. (1969) Multidimensional scaling: combining observations when individuals have different perceptual structure, Psychometrika, 34, 2, pt.1, 139-165.

Jackson, D.N. and S.J. Messick (1963) Individual differences in social perception, British Journal of Social Clinical Psychology, 2, 1-10.

Kruskal, J.B. (1972) A brief description of the 'classical' method of multidimensional scaling, Bell Telephone Laboratories mimeo.

Torgerson, W.S. (1958) Theory and methods of scaling, New York: Wiley.

Tucker, L.R. (1960) Intra-individual and inter-individual multidimensionality, in H. Gulliksen and S. Messick (eds.), Psychological scaling: Theory and applications, New York: Wiley.

Wish, M. and J.D. Carroll (1974) Applications of individual differences scaling to studies of human perception and judgment, in Carterette and Friedman (1974): see Carroll and Wish 1974 above.

Wold, H. (1966) Estimation of principal components and related models by iterative least squares, in P. Krishnaiah (ed.), International Symposium on multivariate analysis, New York: Academic Press.

APPENDIX 1: RELATION OF INDSCAL-S TO PROGRAMS NOT IN MDS(X)

The INDSCAL model is one of the family of models included in the IDIOSCAL program. It is believed that the INDSCAL-like part of IDIOSCAL is, however, prone to suboptimal solution for unknown reasons.

A non-metric INDSCAL model may be approximated using ALSCAL.

APPENDIX 2:

The INDSCAL model and algorithm

Let subjects be indexed $i = 1,\ldots,N$
and stimuli be indexed $j,k = 1,\ldots,p$

The INDSCAL model allows each individual i to rescale each dimension a $(a = 1,\ldots,r)$ of the group space $\underset{\sim}{X}$ by applying a rescaling weight w to each dimension.

Thus we define a private space $\underset{\sim}{Y}$ rescaled according to a subjective metric, i.e.

$$\underset{\sim}{Y} \equiv \{y^i_{ja}\}$$

$$y^i_{ja} = \sqrt{w_{ia}}\, x_{ja} \qquad (1)$$

Within this space the conventional Euclidean distance model holds:

$$d^i_{jk} = \sqrt{\sum_a^r (y^i_{ja} - y^i_{ka})^2} \qquad (2)$$

By substituting (1) in (2) and manipulating, we derive a more general weighted distance model

$$d^i_{jk} = \sqrt{\sum_a^r w_{ia} (x_{ja} - x_{ka})^2} \qquad (3)$$

The Algorithm

This section is based on Carroll and Chang (1970) which is used with permission.

Data are first converted into distance estimates following Torgerson (1958 pp 254-259).

The distance estimates for each subject are converted to a matrix of scalar-products $\underset{\sim}{B}$. This is done by double-centring the matrix whose general entry is $-\frac{1}{2} d_{jk}^2$

This will be regarded as a matrix of scalar products such that

$$b_{jk}^{i} = \sum^{r} y_{ja}^{i} y_{ka} \qquad (4a)$$

or, by substituting (3)

$$b_{jk}^{i} = \sum^{r} w_{ia} x_{ja} x_{ka} \qquad (4b)$$

Let us rewrite (4a) as

$$z_{ijk} \cong \sum^{r} w_{ia} x_{ja}^{L} x_{ka}^{R} \qquad (5)$$

Where

$\cong$ means a least-squares approximation

L and R simply distinguish the x's

Let $\underset{\sim}{W}$, $\underset{\sim}{X}_L$ and $\underset{\sim}{X}_R$ represent the corresponding matrices (N × r), (p × r) and (p × r) respectively, and suppose we are given initial estimates for $\underset{\sim}{X}_L$ and $\underset{\sim}{X}_R$ and we want to derive a least-squares estimate for $\underset{\sim}{W}$

Letting $s = N(j-1) + k$, so that s varies from 1 to N^2, we define

$$g_{sa} \equiv x^L_{ja} \, x^R_{ka}$$

and

$$Z^*_{is} \equiv z_{ijk}$$

We may now rewrite (5) as

$$Z^*_{is} \cong \sum_a^r w_{ia} \, g_{.a} \qquad (6)$$

It is clear that a least-squares solution is available for $\underset{\sim}{W}$. Alternatively we may put (6) in matrix form

$$\underset{\sim}{Z}^* \cong \underset{\sim}{W} \, \underset{\sim}{G}^r \qquad (7)$$

(The columns of $\underset{\sim}{G}^r$ may be thought of as the Kronecker products of the corresponding column vectors of $\underset{\sim}{X}_L$ and $\underset{\sim}{X}_R$).
The least squares solution for $\underset{\sim}{W}$ ($\hat{\underset{\sim}{W}}$) is

$$\hat{\underset{\sim}{W}} = \underset{\sim}{Z}^* \, \underset{\sim}{G} \, (\underset{\sim}{G}^r \underset{\sim}{G})^{-1} \qquad (8)$$

Having solved for $\underset{\sim}{W}$ we may get a better estimate of $\underset{\sim}{X}_L$ say by similar means.

Let $u = p(i-1)+k$ (u varies from 1 to pN) and define

$$h_{ua} \equiv w_{ia} \,.\, x^L_{ka} \quad \text{and} \quad z^{**}_{iu} \equiv z_{ijk}$$

Now (5) may be rewritten as

$$z^{**}_{iu} \cong \sum_a^r x^L_{ja} \, h_{ua} \qquad (9)$$

or, in matrix form

$$\underset{\sim}{Z}^{**} \cong \underset{\sim}{X}_L \underset{\sim}{H}^r \tag{10}$$

from which the least-squares estimate for $\underset{\sim}{X}_L$ is

$$\underset{\sim}{X}_L = \underset{\sim}{Z}^{**} \underset{\sim}{H}(\underset{\sim}{H}^r \underset{\sim}{H})^{-1}$$

The process is repeated on $\underset{\sim}{X}_r$ and re-iterated until the process converges.

Note that while there is no constraint making $\underset{\sim}{X}_L = \underset{\sim}{X}_R$ the basic symmetry of the b^i_{jk} guarantees that at conversion $\underset{\sim}{X}_L$ will be related to $\underset{\sim}{X}_R$ as

$$\underset{\sim}{X}_L = \underset{\sim}{C} \underset{\sim}{X}_R$$

$$\underset{\sim}{X}_R = \underset{\sim}{C}^{-1} \underset{\sim}{X}_L$$

where $\underset{\sim}{C}$ is an $r \times r$ diagonal matrix.

In practice $\underset{\sim}{X}_L$ is set equal to $\underset{\sim}{X}_R$ and $\underset{\sim}{W}$ is recomputed ($\underset{\sim}{X}_r$ being the last computed).