



Notes and Comments on MINI-RSA

APMC 22 May 2001

MINI-RSA (Michigan-Israel-Netherlands-Integrated Rectangular Smallest Space Analysis) aka Non-metric Multidimensional Unfolding provides internal analysis of:

1. **two-way, two-mode *data* of a rectangular, row-conditional matrix**
 2. **by means of a distance *model***
 3. **using an ordinal (non-metric) *transformation* of the data.**
-

1. **MINI-RSA** is a modification by Roskam of an original Guttman-Lingoes Program(SSAR-II). Coombs' (1964) developed the original (unidimensional) unfolding analysis, and this was generalised to the multidimensional case by Bennett and Hays. The Roskam MINI-RSA program implements non-metric multidimensional unfolding and hence provides a distance model analysis of preference rankings data. It consists of a joint mapping of the columns (stimuli, represented as points) and the rows (subjects, each of which is represented by their "ideal-point" (SPPF¹) or point of maximum preference). RSA thus gives a "point-point" representation of the rows and columns of a rectangular matrix.

Although designed originally for the analysis of I-scales (preference rank-orderings), it can also be used for the analysis of scores or ratings data.

2. **MDSX DOCUMENTATION:**

MDS(X) Users Manual, Edinburgh 1981, ch. 7(MINI-RSA_TUM.pdf)

The User' Guide to MDS, Heinemann 1982 ,5.3.3 and 6.2.3

(MINIRSA_TUG533.pdf, MINI-RSA_TUG622.pdf)

¹ Single-peaked preference function: one that has a single maximum, and declines systematically in all directions.

3. MDSX DATA:

TEST INPUT:

(1) (*TESTMINIRSA_INP1.txt*)

Coombs' 4-point configuration and 18 implied I-scales in 2 dimensions (1964, Fig 7.4, p164; also used in Coxon TUG, Fig 5.8). This data-set can be used to test that a correct choice has been made for describing Data-type (see below)

(2) (*TESTMINIRSA_INP2.txt*)

80 Male & Female subjects making preference judgments of family size and composition (derived from their pair-comparisons data) reported in Delbeke 1968) (Self-documenting file. See Coxon's re-analysis of Delbeke/Bollen data in Documentation:

KTMD5_COXON_FamilyComp_Delbeke.jpg)

TESTOUTPUT:

(1) *TESTMINIRSA1_OUT.txt*

(2) *TESTMINIRSA1_OUT.txt*

4. COMMENTS:

MINI-RSA is strongly akin to Simple Correspondence Analysis and to Principal Components Analysis, which all use a SVD of the scalar-products matrix, and differ only in pre-processing of the data, and in normalisation of the solution (Weller and Romney 1990). In MINI-RSA, the row and/or column data can optionally be centred or standardised, and in the solution the row points are projected onto a unit circle. This facilitates the comparison of row-points by reference to column-points (i.e. as projections of stimuli points on to the subject vector), and thus avoids direct comparison of row-column locations (Carroll, Green, Schaffer 1986,7).

5. HINTS:

i) MINI-RSA is a widely used program, especially in marketing (for analysing characteristics of brands), and for joint representation of preference rankings/ratings in general. It can also be thought of as a form of profile analysis. The transformation of the row-data can have major consequences for the scaling representation. Initial analysis often reveals a strong dominating "consensus" factor. When removed (by centring or normalisation) , the subject variability increases dramatically. Note that double-centring is equivalent to making a MINI-RSA analysis a distance analysis.

ii) RSA is of course a non-metric program, and unless the number of stimuli and subjects is large enough and the data cover a wide enough area of the possible I-scales, sub-optimal results may occur. In such cases, you are advised to use a metric unfolding program.

iii) It is possible to carry out an [internal] METRIC unfolding by using PREFMAP (qv), using the GENERATE or INITIAL parameter (see PREFMAP documentation, 13.11 - 13.12

iv) NOTE: CONFUSION HAS ARISEN OVER THE DATA TYPE SPECIFICATION IN MINI-RSA. SEE APPENDIX TO THIS NOTE FOR CLARIFICATION.

iii) REFERENCES

BASIC REFERENCES:

Coombs, C.H. (1964) A Theory of Data, London: Wiley

Bennett J.F. & Hays, W.L. (1960) Multidimensional unfolding: determining the dimensionality of ranked preference data, Psychometrika, 25, 27-43

Roskam, E.E. (1979) A survey of the Michigan-Israel-Netherlands-Integrated Series, in J.C. Lingoes, E.E.Roskam & I. Bor, eds (1979) Geometric Representations of Relational Data, Ann Arbor: Mathesis Press, pp 289-312

Other:

Delbeke, L (1968) Construction of Preference Spaces: an investigation into the applicability of multidimensional scaling models: Louvain-Leuven: Publications of the University of Louvain

7. STATUS

The algorithm appears to be stable and reliable, but results can be sub-optimal if there are few subjects and/or not enough variation in the data (see cautionary case of Delbeke's 0 family size). Usage: Moderate

APPENDIX ON DATA TYPE SPECIFICATION IN MINI-RSA

There has been some confusion about the definition of the four input data-types used in MINIRSA and PREFMAP. (See 2.1.1 and Fig 1, 7.5 [MINIRSA] and 13.16 [PREFMAP]. Basically, preference judgments ordering can either be in the form of an I-scale (where the stimuli are placed in decreasing order of preference) or as a rank-score (as in a data-matrix where each stimulus has a fixed column allocated to it, and the entry is the rank-score (or rating-scale value) allocated to it. But in both cases, the ordering can either be from highest to lowest, or the reverse. This therefore generates four types of data. Moreover, programs do not normally allow alphabetical characters to be used to define an I-scale, so the corresponding number has to be used to replace letters (e.g. A=1, B=2, etc)

For example, if a subject were asked to rank 5 objects {A,B,C,D,E}, and gave the (I-scale) answer: *BDACE* (B most preferred, then D, then A ... finally E), and the numbers {1,2,3,4,5} were used in place of the letters, then *BDACE* could be represented in any of four ways:

i)	I-scale; first position is the MOST preferred:	2 4 1 3 5 = BDACE
ii)	I-scale; first position is the LEAST preferred:	5 3 1 4 2 = ECADB
		[A B C D E]
iii)	Fixed-position score; highest value is the MOST preferred:	3 5 2 4 1
iv)	Fixed-position score; highest value is the LEAST preferred:	3 1 4 2 5.

In practice, alternative (i) is normally used for I-scales and (iii) for Rank-scores respectively.

The confusion which has arisen in the MDS(X) Documentation is of two sorts:

- Different type numbers and ranges are given: The MINIRSA documentation labels the four types from 0,1,2 to 3 (Fig 1 on p.7.5 and the Parameters DATA TYPE specification), whilst the text immediately before this (pp7.3-7.4) refers to types 1,2,3,4. Moreover the example at 7.3-4 is identified in the text as DATA TYPE 1, whereas the PARAMETERS definition and Figure 1 identify it as DATA TYPE 0.
- Reversal of Data-types 3 and 4 between the Figure and Documentation. The difference between these two types on p7.4, p7.5 and 7.10 are not compatible.

In order to determine what the programs actually do with respect to DATA TYPE, a well known and simple data set was used, which specifies the 18 I-scales compatible with a

given 4-stimulus configuration (see Coombs 1964, p146 (F7.4); Coxon T.U.G. p 142² (F5.8) The I-scales were numbered from the (open) isotonic region immediate to the right of A (#1=ADCB) and clockwise round to #12 (ADBC) then immediately down to the closed isotonic regions #13 (DABC) clockwise to #18 (DBAC), thus forming a recognisable spiral or 2 concentric-circular pattern.

In order to determine unequivocally what the options of DATA TYPE are, and how they operate, these data were coded in terms of the four types as above [(i) to (iv)] and each type was then run in MINI-RSA with separate runs for DATA TYPE (0), 1,2,3,4. (For details, see *RSA_Data-Type_Experiment* in MINI-RSA Documentation).

The results are clear:

- **There is no difference between DATA TYPE(0) and DATA TYPE (1)**
- **The four types are as follows:**

DATA TYPE (1)	I-scale input (i). First stimulus is most preferred
DATA TYPE (2)	Reverse I-scale input (ii). First stimulus is least preferred
DATA TYPE (3)	Fixed-score input (iii). High score means high preference
DATA TYPE (4)	Reverse Fixed-score input (iv). High score means low preference
- (Remember also that the INPUT FORMAT *must* allow for a frequency field)

In the Documentation (MDS(X) User Manual), pages 7.4, 7.5 and 7.10 should be replaced, as follows:.

²In so doing it was discovered that there is an error in Figure 5.8 of T.U.G., which had never been noticed before: Open isotonic region 7 is identified in the book as ADCB, whereas it is the reverse of this: BCDA ! The test data set has been corrected.

(Replacement pages 7.4 and 7.5 MINI-RSA)

Data may be input to MINI-RSA in this form (as an I-scale) by specifying DATA TYPE (1). In some data-collection techniques, it can happen 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, with A being the most preferred. If this is the case, then the data should be specified as a reversed I-scale, i.e. DATA TYPE (2).

A different way of representing the data is by the (fixed-position)' score' method. In this method, each column represents a particular stimulus and the entry in that column gives the score (rank-score, rating-score) allocated by the subject to that stimulus. 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
<i>subject i</i>	5	1	4	3	2

In this instance, the highest number (5) is used to denote the most preferred stimulus (A), and the lowest number (1) is used to denote the least preferred. This option (Fixed-position score, with high value meaning high preference), is chosen as: DATA TYPE (3).

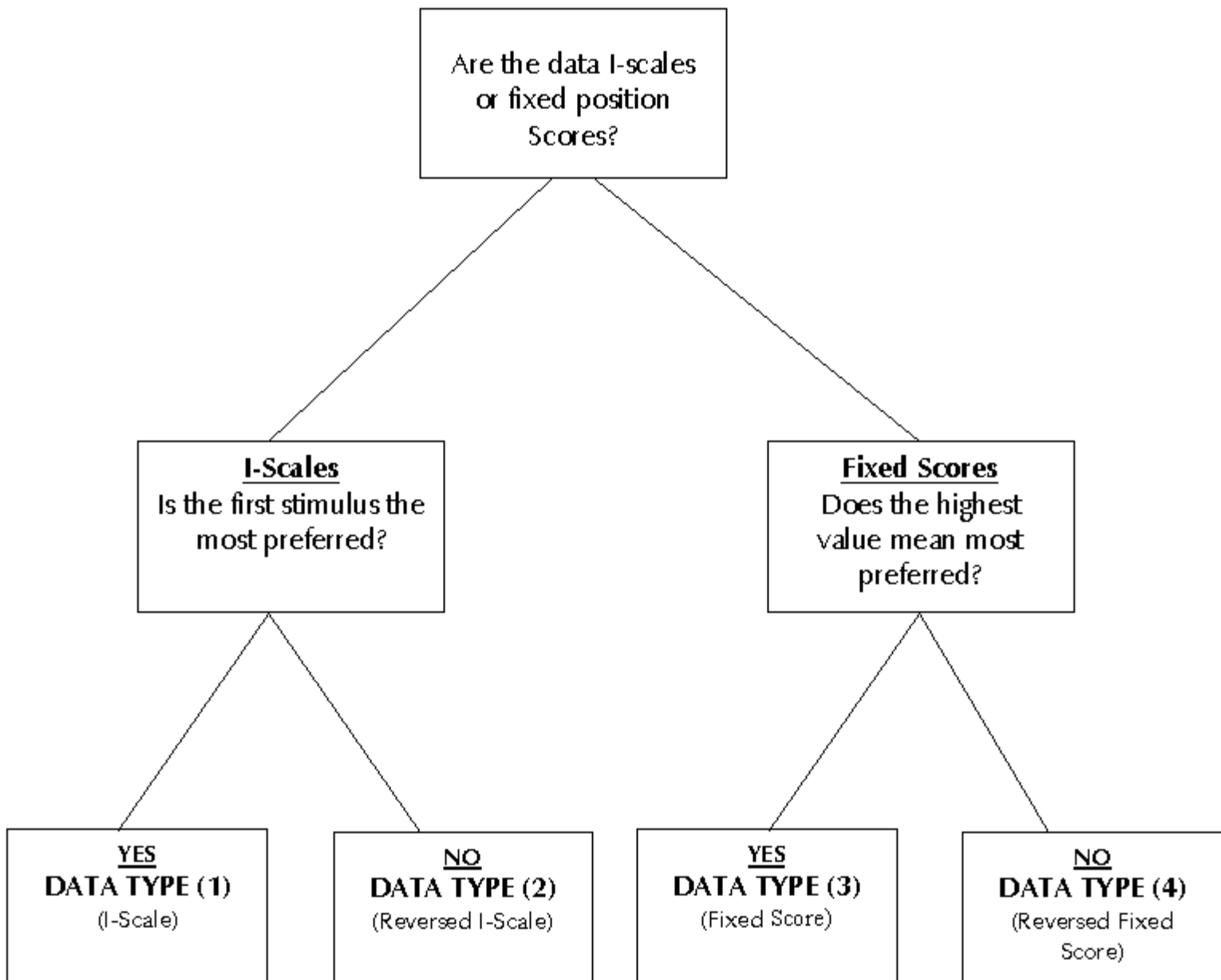
Alternatively, the lowest number (1) could have been used to represent the most preferred stimulus (A), and if this is so, DATA TYPE (4) should be specified:

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

Although in illustrating the Fixed Score method we have used the rank numbers 1 to 5, the data might equally well have been numerical ratings. Figure 1 provides a simple way of deciding upon the appropriate DATA TYPE value.

Figure 1 (Corrected)

Specification of DATA TYPE



(Replacement page 7.10, MINI-RSA)

3. INPUT PARAMETERS

3.1 LIST OF PARAMETERS

<u>Keyword</u>	<u>Default Value</u>	<u>Specification</u>
DATA TYPE	1	1: I-scale input. First stimulus is <u>most preferred</u> 2: Reverse I-scale input. First stimulus is <u>least preferred</u> 3: Fixed-position score input. High score means <u>high preference</u> 4: Reverse Fixed-score input. High score means <u>low preference</u>.
MINIMUM ITERATIONS	6	Sets minimum number of iterations to be performed before convergence test.
MISSING DATA	0	Sets the data value which is to be regarded as missing data.
MATFORM	0	<u>Note:</u> only relevant when READ CONFIG is used. 0: The input configuration is punched subjects and stimuli (rows) by dimensions (cols.). Subject co-ordinates preceded stimuli co-ordinates. 1: The input configuration is punched dimensions (rows) by subjects and stimuli (cols.)

3.2 NOTES

1. The {# |N|No} OF SUBJECTS instruction may be replaced by {# |N|No} OF ROWS
2. The {# |N|No} OF STIMULI instruction may be replaced by {# |N|No} OF COLUMNS
3. See 2.3.2 for how to input single rows with a frequency count.
N.b. The INPUT FORMAT must allow for the frequency filed, even if left blank (= 1)