18. WOMBATS: Work Out Measures Before Attempting to Scale

18.1 Overview

Concisely: WOMBATS (Work Out Measures Before Attempting To Scale), does just what its acronym says and computes from a rectangular data matrix one or more (dis)similarity measures suitable for input to other NewMDSX procedures.

18.1.1 WOMBATS in brief

The WOMBATS program is in effect a utility which takes as input a rectangular matrix either of raw data, and computes a measure of (dis)similarity between each pair of variables in the matrix. These measures are output in a format suitable for input either to other NewMDSX procedures or to other programs. This output format is chosen by the user.

18.2. DESCRIPTION OF THE PROGRAM

The following section describes briefly those aspects of the program pertinent to its use. The measures calculated in WOMBATS are those detailed in chapter 2 of `The User's Guide' (Coxon 1982). For a fuller discussion, see that reference.

Section 2.1 describes the type of data suitable for input, and its presentation to the program and section 2.2 the range of measures available. Section 2.3 describes further options including those for outputting the results.

18.2.1 Data

The basic form of input data for the WOMBATS program is a rectangular matrix in which the rows represent cases (or subjects) and the columns, variables (or stimuli). This may be a matrix of 'raw' data as collected by the user or exported from EXCEL, SPSS or a similar program.

The number of rows in the matrix is specified by the user in the N OF CASES command or, (alternatively, in N OF SUBJECTS). The number of columns fields is given by either N OF VARIABLES or N OF STIMULI. (In these commands 'N' may of course be replaced by either 'NO' or '#'.) The data are read by the program when it encounters a READ MATRIX command, and the INPUT FORMAT specification, if used, should describe one row of the data matrix. Otherwise, data values are be entered in free format, separated by spaces.

If the data to be input are for some reason in a matrix where the rows represent variables and the columns cases, then the user should specify MATFORM(O) in the PARAMETERS command.

The chosen measures are calculated between the entities designated as variables (so-called R-analysis). This will be the case whatever value is taken by the parameter MATFORM. If the user wishes measures to be calculated between cases rather than between variables (Q-analysis), see section 2.3.1 below.

N.B. The program expects data to be input as real numbers. The INPUT FORMAT statement, if used, must therefore be specified to read F - type numbers, even if the numbers do not contain a decimal point.

18.2.1.1 Levels of Measurement

The user must specify, for each of the variables in the analysis, the level of measurement at which it is assumed to be. Five levels are recognised by the program. The recognised levels are ratio, interval, ordinal, nominal and dichotomous. If a particular variable is not explicitly assigned to a

particular level by the user, then the program assigns it by default to the ordinal level of measurement.

Each of the measures in the program assumes that the variables on which it is operating have the properties of a particular level of measurement. If an attempt is made to compute a measure which assumes a level of measurement higher than that at which the variables have been declared to lie, the program will fail with an error message. No restriction is placed, obviously, on the attempt to calculate measures which assume levels lower than those declared.

The user signals the measurement level of the variables to the program by means of the LEVELS command, peculiar to the WOMBATS program. This consists of the command LEVELS, and one or more of the keywords RATIO, INTERVAL, NOMINAL, DICHOTOMOUS or ORDINAL. (Obviously, since the program defaults to ordinal, there is no need actually to specify variables associated with this last keyword). In parentheses following each keyword used are listed the variables which are to be assumed to be at that level of measurement. In these parentheses, ALL and TO are recognized. The following are valid examples of a LEVELS declaration.

```
LEVELS          INTERVAL (1, 2, 5, 7,), NOMINAL (3, 4, 6, 8)
LEVELS          RATIO (ALL)
LEVELS          NOMINAL (1 TO 4), INTERVAL (7 TO 11)
```

In the last example, variables 5 and 6 are presumed by default to be at the ordinal level.

18.2.1.2 Missing Data

Variables that include missing data are a problem. The user may specify, for each variable in which there are missing data, one code which the program will read as specifying a missing datum. Users will note however that an attempt to calculate certain measures between variables will fail if missing data are present. The measures for which this is the case are indicated in the discussion of the available measures in section 18.2.2.1.

The user signals the occurrence of missing data by means of the MISSING statement. This consists of the command MISSING followed by the value(s) to be regarded as signifying missing data. In parentheses following each missing data value is a list of the variables for which that value represents a missing datum. In these parentheses the forms ALL and TO are recognised. The following are valid examples of a MISSING declaration.

```
MISSING         -9.(1, 2, 7, 9),  99.(3, 4, 6, 8)
MISSING         0. (ALL)
MISSING         .1(1 TO 7), -.1(8 TO 16)
```

18.2.2 ANALYSIS

The aim of the WOMBATS program is to calculate for each pair of variables in the analysis a measure of the (dis)similarity between them. Having described the data to the program, the user must then choose the measure to be calculated. WOMBATS currently offers 26 different general measures. In comparing cases, however, where variables are of mixed types, it is also possible to apply Gower's General Coefficient of Similarity (1971).

The required measures are chosen by means of the MEASURE command. This contains the keyword MEASURE followed by one only of the keywords referring to the available measures described below. Only one measure is computed in each TASK of the run. If more than one measure is required on the same set of data, then a separate TASK NAME is necessary.

18.2.2.1 Available measures

It is convenient to consider the available measures in WOMBATS under their respective assumed levels of measurement.

18.2.2.1.1 Dichotomous measures

Sixteen measures of agreement between dichotomous variables are included in WOMBATS. These correspond to those described in `The User's Guide to MDS' pp.24-27. Missing data are allowed in all these measures.

In this section, the following notation will be crucial. Consider two dichotomous variables which we will assume to measure whether the objects under consideration do or do not possess a particular attribute. The co-occurrence(or frequency) matrix of these two variables looks as follows.

		Variable 1	
		1/Yes	0/No
Variable 2	1/Yes	a	B
	0/No	c	D

The cell `a' is the number of times that the attributes 1 and 2 co-occur, `b', the number of times attribute 2 is present when attribute 1 is not, `c' is the number of times attribute 1 is present and 2 is not and `d' is the number of objects possessing neither attribute 1 nor attribute 2. All the measures of agreement to be considered in this section result from the combination of these quantities in some way.

The measures available for the comparison of dichotomous variables are denoted by the `keywords' D1, D2, ..., D16 and it is these `keywords' that appear in the MEASURE command

For example, the command

MEASURE D15

will select Yule's Q as the measure to be calculated

Before choosing a dichotomous measure, users should consider:

- whether they wish "co-absences" (cell d) to feature in the assessment of similarity, and
- whether they wish the measure to have Euclidean properties. Gower and Legendre(1986) prove that if a similarity measure has non-negative values and the self-similarity s_{ii} is 1, then the dissimilarity matrix

  with entries $\delta_{ij} = \sqrt{(1-s_{ij})}$ is Euclidean.

Note that any similarity measure can be converted into a dissimilarity measure by a related transformation:

$\delta_{ij} = (1-s_{ij})$ if the similarity measure takes values between 0 and 1,

or $\delta_{ij} = (max-s_{ij})$ where max is the value of the greatest similarity.

D1 and D2 are undoubtedly the simplest and most commonly-used of these measures.

Each dichotomous measure is now considered:

Command	MEASURE D1
Type	Similarity measure
Range	low = 0, high = 1

Name	Jaccard's coefficient
Formula	$\frac{a}{(a+b+c)}$
Description	Excludes `d'. Represents the probability of a pair of objects exhibiting both of a pair of attributes when only those objects exhibiting one or other are considered. It is possible that a division by zero may occur in the calculation of this measure.

Command	MEASURE D2
Type	Similarity measure
Range	low = 0, high = 1
Formula	$\frac{a}{(a+b+c+d)}$
Name	Russell and Rao's measure
Description:	Represents the probability of a pair of objects in a pre-selected set exhibiting both of a pair of attributes.

Command	MEASURE D3
Type	Similarity measure
Range	low = 0, high = 1
Name	Sokal's measure
Formula	$\frac{(a + d)}{(a + b + c = d)}$
Description	Includes `d' in numerator and denominator. Represents the probability of a matching of two attributes.

Command MEASURE D4
Type Similarity measure
Range low = 0, high = 1
Formula

$$\frac{2a}{(2a+b+c)}$$

Name Dice's measure
Description Gives the positive matches `a' twice as much importance as anything else. Excludes entirely the mismatches. It is thus possible that a division by zero may occur in the calculation of this measure.

Command MEASURE D5
Type Similarity measure
Range low = 0, high = 1
Formula

$$\frac{2(a+d)}{(2(a+d)+b+c)}$$

Name no name
Description Includes `d' in both numerator and denominator. The matches (a and d) are given twice as much weight as the mismatches.

Command MEASURE D6
Type Similarity measure
Range low = 0, high = 1
Formula

$$\frac{a}{(a+2(b+c)}$$

Name no name
Description Excludes `d' entirely. The matches (b and c) are accorded twice as much weight as the matches. It is possible that a division by zero may occur in the calculation of this measure.

Command MEASURE D7
Type Similarity measure
Range low = 0, high = 1
Name Rogers and Tanimoto's measure

$$\frac{(a+d)}{(a+d+2(b+c)}$$

Formula

Description Includes `d' in numerator and denominator. The mismatches (b and c) are accorded twice as much weight as the matches.

Command	MEASURE D8
Type	Similarity measure
Range	low = 0, high = a + b + c + c + d - 1
Name	Kulczynski's measure

$$\frac{a}{b+c}$$

Formula

Description Excludes `d' entirely. This measure is the simple ratio of the positive matches (a) to the mismatches (cf. D9). it is possible that a division by zero could occur in the calculation of this measure and an undefined statistic occur. The maximum value otherwise is as stated.

Command MEASURE D9
Type Similarity measure (Sokal & Sneath)
Range low = 0, high = a + b + c + d − 1
Formula

$$\frac{(a+d)}{(b+c)}$$

Name no name
Description This measure is the simple ratio of all matches (positive and negative) to the mismatches (cf D8). The statistic may be undefined, due to a zero divisor. The maximum finite value is as stated.

Command MEASURE D10
Type Similarity measure
Range low = 0, high = 1
Name Kulczynski's measure

$$\frac{1}{2}\left(\frac{a}{a+c}+\frac{a}{a+b}\right)$$

Formula
Description Excludes `d' entirely. This measure is a weighted average of the matches to one or other of the mismatches. This statistic may be undefined.

Command MEASURE D11
Type Similarity measure
Range low = 0, high = 1
Formula

$$\frac{1}{4}\left(\frac{a}{a+c}+\frac{a}{a+b}+\frac{d}{b+d}+\frac{d}{c+d}\right)$$

Name no name
Description Includes `d' in numerator and denominator. This is the analogue of D10 with mismatches included.

Command	MEASURE D12
Type	Similarity measure
Range	low = 0, high = 1
Formula	

$$\frac{a}{\sqrt{\langle (a+c)(a+b) \rangle}}$$

Name	Ochiai's measure
Description	Excludes `d' from numerator. It uses the geometric mean of the marginals as a denominator. This statistic may have a zero divisor.

Command MEASURE D13
Type Similarity measure
Range low = 0, high = 1
Formula

$$\frac{ad}{\sqrt{\langle (a+c)(a+b)(b+d)(c+d) \rangle}}$$

Name no name
Description Includes `d' in numerator and denominator. It uses the geometric mean of the marginals as a denominator and will return a value of 0 iff either a or d is empty.

Command MEASURE D14
Type Similarity measure
Range low = -1, high = +1
Formula

$$\frac{(a+d)-(b+c)}{(a+b+c+d)}$$

Name Hamann's coefficient
Description Simply the difference between the matches and the mismatches as a proportion of the total number of entries. A value of 0 indicates an equal number of matches to mismatches. Some thought should be given to the interpretation of any negative coefficients before scaling the results.

Command MEASURE D15
Type Similarity measure
Range low = -1, high = +1
Formula

$$\frac{(ad)-(bc)}{(ad+bc)}$$

Name Yule's Q
Description This is the original measure of dichotomous agreement, designed to be analogous to the product-moment correlation. A value of 0 indicates statistical independence. Some thought should be given to the interpretation of any negative coefficients before scaling the results. This statistic may be undefined.

Command MEASURE D16
Type Similarity measure
Range low = -1, high = +1
Formula

$$\frac{(ad-bc)}{\sqrt{\langle (a+c)(a+b)(b+d)(c+d) \rangle}}$$

Name Pearson's Phi

Description

A value of 0 indicates statistical independence. Some thought should be given to the interpretation of any negative coefficients before scaling the results. The statistic may be undefined if any one cell is empty.

18.2.2.1.2 Nominal measures

Five measures are available in WOMBATS for the measurement of nominal agreement between variables. Four of these are based on the familiar chi-square statistic. The other is the Index of Dissimilarity.

18.2.2.1.2.1 Chi-square based measures

The following procedure is used to evaluate the chi-square statistic that forms the basis of four of the available measures.

Consider two variables x and y. We form the table whose row elements are the values taken by (or the categories of) the variable x and whose column elements are the values (categories) taken by variable y. (Obviously, since this is a nominal measure, these values have no numerical significance). The entries of this table are the number of cases which take on particular combinations of values of x and y i.e. the number of cases that fall into the particular combinations of categories.

The value of the chi-square statistic is calculated by comparing the actual distribution of these values in the cells of the table to that distribution which would be expected by chance (statistical independence occurs when $p(i,j) = p(i) \times p(j)$) . Thus, the higher the value of the statistic, the more the actual distribution diverges from the chance or expected one (0).

In the case of there being missing data in the original matrix, then the whole row or column corresponding to that value is deleted. Caution should be exercised if there are many missing data and particularly if these are unequally distributed around the variables since the value of the statistic is dependent on the number of values it considers and strictly speaking chi-square measures based on largely different numbers of cases are not comparable.

The other measures in this section seek to overcome the dependence of chi-square on the number of cases by norming it. The norming factor differs for each statistic.

The following notation will be used in discussing nominal measures:

- N will indicate the number of cases
- r will stand for the number of rows in the matrix i.e. the number of categories (values) taken by variable x and
- c will stand for the number of columns i.e. the number of categories in variable y.

Name	Chi - square
Command	MEASURE CHISQUARE
Type	Similarity measure
Range	low = 0, high = N x min(r,c)
Comment	A value of 0 indicates statistical independence. The maximum value is dependent on the value of N.

Name	Phi
Command	MEASURE PHI
Type	similarity measure
Range	low = 0, high = ≤(min(r,c)-1)
Comment	The phi coefficient is chi-square normed to be independent of N. Reaches a maximum for 2 x 2 tables in which case it reduces to the product-moment correlation.

It may, however, exceed 1 when the minimum of r and c is greater than 2.

Name	Cramer's V
Command	MEASURE CRAMER
Type	similarity measure
Range	low = 0, high = 1
Comment	Cramer's coefficient is chi-square normed to be independent of N and of the number of r and c. Reaches a maximum for non-square tables.

Name	Pearson's Contingency coefficient C
Command	MEASURE PEARSON
Type	similarity measure
Range	low = 0, high = 1
Comment	Pearson's coefficient is chi-square normed to be independent of N, originally developed as a measure for contingency tables. Cannot reach its maximum of 1 for non-square tables.

18.2.2.1.2.2 The index of dissimilarity

The remaining statistic in this section is the index of dissimilarity. In the case of the chi-square measures, the implicit comparison is between the actual (bi-variate) distribution and the expected (chance) one. In the case of the index it is two (univariate) distributions that are compared.

Consider again the table that is formed by cross-tabulating the values of variable x and those of variable y. If the two variables had identical distributions then all the off-diagonal cells would be empty. The index of dissimilarity is simply the proportion of cases that appear in these off-diagonal cells and may be thought of as the proportion of changes needed to change the one distribution into the other. The index does not require equal numbers of values in the variables.

Name	Index of dissimilarity
Command	MEASURE ID
Type	dissimilarity
Range	low = 0, high = 100

18.2.2.1.2 Ordinal level measures

At present, there are three measures of ordinal agreement in WOMBATS, all related to the basic tau (τ) measure of Kendall (19..). τ_b, τ_c and Goodman and Kruskal's gamma (γ). There are two important distinctions in considering these measures. First, we need to know if they measure weak or strong monotonic agreement between the variables and secondly how they treat tied values in them. This second distinction can be crucial since much ordinal level data, being composed of a relatively small number of categories, will contain a large proportion of tied data values.

Consider a two-way table between ordinal variables x and y. For any pair of individuals ***i,j*** , one of the following five conditions will hold:

a) Concordant (C): where X and Y order the individuals in the same way (if ***i*** is higher(lower) on X, the ***j*** is higher(lower) on Y)

b) Discordant (D): where X and Y order the individuals in opposite ways
c) Tied on X (T_x)
d) Tied on Y (T_y)
e) Tied on both X and Y (T_{xy})

The numerator of all the ordinal measures here considered is the difference between numbers of concordant and discordant pairs. They differ in the form the denominator takes.

Name Goodman and Kruskal's gamma (γ)
Command MEASURE GAMMA
Type similarity measure
Range low = -1, high = +1
Formula

$$\gamma = (C-D)/(C+D)$$

Comment Measures the weak monotonic agreement between the variables, taking the ratio of the difference between concordant and discordant pairs to their sum. It thus ignores the ties completely. For this reason it is possible that the value be undefined (i.e. there may be no cases). If there are no ties then the index reduces to Yule's Q (D15). Some thought should be given to the interpretation of the negative values before the results are scaled.

Name Kendall's tau-b (τ_b)
Command MEASURE TAUB
Type similarity measure
Formula

$$\tau_b = (C-D) / \{\sqrt{(C+D+T_y)}.\sqrt{(C+D+T_x)}\}$$

Range low = -1, high = +1

Comment Measures strong monotonic agreement in the variables, relating the difference between concordant and discordant pairs of the geometric mean of the quantities arrived at by adding in the ties to the denominator. This should be used only for square tables.

Name Kendall's tau-c (τ_c)
Command MEASURE TAUC
Type similarity measure
Formula (corrects for non-square tables)

Range low = -1, high = +1

Comment In the formula, m stands for the lesser of the number of rows and columns in the original matrix. The statistic may be used for non-square tables and reduces, in the case of square ones to tau-b.

18.2.2.1.4 Interval level measures

The interval level measures currently available in WOMBATS are product-moment measures (covariance and the product-moment correlation) and Euclidean distance.

Consider the conventional scatter-plot of, a number of cases measured on two variables. These cases may be represented as points in a space, the two dimensions of which are the variables concerned. (The statement holds for more than two variables, of course.) The Euclidean distance between the cases is the straight line distance between the points which represent them. The correlation between each pair of points is simply the cosine of the angle between the two vectors drawn from the origin to the points concerned and the covariance is that same cosine multiplied by the length of the vectors.

Command	MEASURE DISTANCE
Type	dissimilarity
Range	low = 0, high = maximum variance in the variables
Comments	If the ranges of the variables involved are markedly different, then some attempt at rescaling (i.e. normalisation) should be made so that differences in a highly valued variable do not swamp out differences in one of humbler dimensions. Does not take into account the extent to which the variables are correlated. (A measure which does so is Mahalanobis 1936, qv.)

Command	MEASURE COVARIANCE
Type	similarity
Range	low = 0, high = highest variance
Comments	The interpretation given to the negative values should be carefully thought out before scaling.

Command	MEASURE CORRELATION
Type	similarity
Range	low = 0, high = 1
Comments	The negative values may need to be given some thought before the results of this calculation are scaled.

18.2.3 FURTHER OPTIONS

18.2.3.1 Measures between cases

It may be that the user wishes to have the measures calculated between the cases (subjects, individuals) in the analysis rather than the variables. When variables are all of the same or comparable levels, this is accomplished simply by specifying in the PARAMETERS command, the keyword ANALYSIS, followed in brackets by the figure 1.

This command has the effect of calculating the measures between the entities designated as cases and is independent of the MATFORM parameter.

When variables are of mixed levels, and for up to a maximum of 200 cases, it is possible to calculate Gower's general similarity coefficient:

Command	MEASURE GOWER
Type	similarity
Range	low = 0, high = 1

Comments This is applicable as a general measure of similarity between cases, where the variables are of mixed levels. This measure also requires the statement **GOWER WEIGHTS**, followed on the next and subsequent lines of input by a series of (positive) weight values to be applied to the variables in calculating similarity values. The value 1.0 for all variables treats all variables as of equal weight. Gower's General Similarity Coefficient s_{ij} compares two cases i and j, and is defined as follows:

$$s_{ij} = \frac{\Sigma_k \; w_{ijk} s_{ijk}}{\Sigma_k \; w_{ijk}}$$

where: s_{ijk} denotes the contribution provided by the *k*th variable, and w_{ijk} is usually 1 or 0 depending upon whether or not the comparison is valid for the *k*th variable; if differential variable weights are specified it is the weight of the *k*th variable or 0 if the comparison is not valid.

The effect of the denominator $\Sigma_k \; w_{ijk}$ is to divide the sum of the similarity scores by the number of variables; or if variable weights have been specified, by the sum of their weights.

Gower defines the value of s_{ijk} for ordinal, interval and ratio variables as:

$$s_{ijk} = 1 - | \; x_{ik} - x_{jk} \; | \; / r_k$$

where: r_k is the range of values for the *k*th variable.

For a dichotomous variable, $s_{ijk} = 1$ if cases i and j both have attribute k "present" or 0 otherwise, and the weight w_{ijk} causes negative matches to be ignored. If all variables are binary, then Gower's General Similarity Coefficient is equivalent to Jaccard's Similarity Coefficient A/(A+B+C) (Measure D1 above) since the negative matches scored in cell D are ignored. If negative matches are not to be ignored, the variable should be specified as a nominal variable.

The value of s_{ijk} for nominal variables is 1 if $x_{ik} = x_{jk}$, or 0 if $x_{ik} \neq x_{jk}$. Thus $s_{ijk} = 1$ if cases i and j have the same "state" for attribute k, or 0 if they have different "states", and $w_{ijk} = 1$ if both cases have observed states for attribute k.

The weight w_{ijk} for the comparison on the *k*th variable is usually 1 or 0. However, if you assign differential weights to your variables, then w_{ijk} is either the weight of the *k* th variable or 0, depending upon whether the comparison is valid or not. This allows larger weights to be given to important variables, or for an external scaling of the variables to be specified.

18.2.3.2 Multiple analyses

Only one measure may be calculated at each TASK NAME. In order to calculate more than one measure on the same data at one time, more than one TASK NAME should be contained in one RUN. The TASK NAME command also resets PARAMETERS values to their original (default) values and it is necessary to reset these on subsequent runs, as required.

18.3. OUTPUT OPTIONS

The measures are output by default as a lower triangular matrix suitable for input to other procedures in the NewMDSX library. There is no need to signal this output with a command. Other options are available which match different conventions in other programs (see below) and in this case it is necessary to specify the output format for the measures.

18.3.1 Secondary output

If an OUTPUT FORMAT statement is included, specifying a valid FORTRAN format in brackets, this will be used to save the matrix in an optional secondary output file. By default, there is no secondary output.

18.3.2 Alternative output forms

By request, measures may be output as an upper triangular or as full (symmetric) matrix. This is accomplished by use of the keyword OUTPUT in the PARAMETERS command:

- The default specification OUTPUT(1) gives a lower triangle without diagonal, and
- OUTPUT(2) a lower triangle with digonal, and
- OUTPUT(3) a full matrix.

This parameter does not affect the operation of the OUTPUT FORMAT command, if used.

18.3 Examples

```
RUN NAME             WOMBATS TEST DATA
COMMENT  Gower coefficient example - Everitt (1974) p.55
N OF STIMULI    5
N OF SUBJECTS   3
MISSING     99(ALL)
MEASURE     GOWER
PRINT DATA      YES
READ MATRIX
 66.0 120.0 1.0 1.0 1.0
 72.0 130.0 2.0 2.0 1.0
 70.0 150.0 1.0 1.0 0.0
LABELS  Subject 1
Subject 2
Subject 3
GOWER weights
   1.00000
   1.00000
   1.00000
   1.00000
   1.00000
LEVELS INTE(1,2,) NOMI(3,4,) DICH(5,)
PARAMETERS  OUTPUT(2)
COMPUTE
RUN NAME             WOMBATS TEST PROG
TASK NAME            CORRELATION TEST
NO OF STIMULI         4
NO OF SUBJECTS       15
LEVELS               INTERVAL  (1 TO 4)
```

```
OUTPUT FORMAT        (1X,3F13.7)
MISSING              2.(2 , 3)  3.( 4)
PARAMETERS           OUTPUT (1)
MEASURE              CORREL
READ MATRIX
 1. 1. 3. 4.
 1. 2. 3. 3.
 2. 4. 3. 3.
 4. 3. 3. 4.
 3. 2. 3. 2.
 4. 3. 3. 4.
 3. 3. 2. 1.
 1. 1. 4. 3.
 3. 4. 3. 1.
 3. 4. 2. 1.
 1. 2. 1. 1.
 3. 3. 4. 2.
 4. 3. 2. 1.
 1. 2. 1. 2.
 2. 3. 4. 1.
COMPUTE
TASK NAME            CUBE
NO OF STIMULI        8
NO OF SUBJECTS       3
LEVELS               INTERVAL (1 TO 8)
MEASURE              DISTANCE
READ MATRIX
 0. 0. 0. 0. 1. 1. 1. 1.
 1. 1. 0. 0. 1. 1. 0. 0.
 1. 0. 1. 0. 1. 0. 1. 0.
COMPUTE
TASK NAME            TAU AND SIMILAR
NO OF STIMULI         4
NO OF SUBJECTS       15
LEVELS               INTERVAL  (1 TO 4)
MISSING              2.(2 , 3)  3.( 4)
PARAMETERS           OUTPUT(1)
INPUT FORMAT          (8F3.0)
MEASURE                GAMMA
READ MATRIX
 1. 1. 3. 4.
 1. 2. 3. 3.
 2. 4. 3. 3.
 4. 3. 3. 4.
 3. 2. 3. 2.
 4. 3. 3. 4.
 3. 3. 2. 1.
 1. 1. 4. 3.
 3. 4. 3. 1.
 3. 4. 2. 1.
 1. 2. 1. 1.
 3. 3. 4. 2.
 4. 3. 2. 1.
 1. 2. 1. 2.
 2. 3. 4. 1.
COMPUTE
TASK NAME      INDEX OF DISSIMILARITY TEST
NO OF CASES      4
NO OF VARS       4
PARAMETERS     OUTPUT(5)
INPUT FORMAT     (4F3.0)
MEASURE         ID
READ MATRIX
```

```
 58 22 41 19
 30 38 14 23
 25 44 19 22
 07 51 12 51
COMPUTE
TASK NAME            PHI
NO OF STIMULI       4
NO OF SUBJECTS        15
LEVELS               INTERVAL  (1 TO 4)
MISSING          2.(1 , 3)  3.( 4)
PARAMETERS           OUTPUT(2)
INPUT FORMAT          (8F3.0)
MEASURE          PHI
READ MATRIX
 1. 1. 3. 4.
 1. 2. 3. 3.
 2. 4. 3. 3.
 4. 3. 3. 4.
 3. 2. 3. 2.
 4. 3. 3. 4.
 3. 3. 2. 1.
 1. 1. 4. 3.
 3. 4. 3. 1.
 3. 4. 2. 1.
 1. 2. 1. 1.
 3. 3. 4. 2.
 4. 3. 2. 1.
 1. 2. 1. 2.
 2. 3. 4. 1.
COMPUTE
FINISH
```

REFERENCES

Coxon, A.P.M. (1982) The User's Guide to Multidimensional Scaling, London, Heinemann

Everitt, B.S. (1974) Cluster Analysis, London, Heinemann

Everitt, B.S. and Rabe-Hesketh, S (1966) The Analysis of Proximity Data, London, Arnold

Gower, J.C. (1971) Statistical methods for comparing different multivariate analyses of the same data, in C.R.Hodson, D.G.Kendall and P.Tăutu eds. Mathematics in the Historical and Archaeological Sciences, Edinburgh University Press, pp. 138-149.

Gower, J.C. (1971) A general coefficient of similarity and some of its properties, Biometrics, 27, 857-872.

Gower, J.C. amd Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients, Journal of Classification, 5, 5-48

Sokal, R.R. and Sneath, P.H. (1963) Principles of Numerical Taxonomy, London, Freeman