KUNST

DYADS

Bootstrapping approach to dyads

By Maria Thissen
&
Nol Bendermacher

RTOG-FSW
Radboud university Nijmegen

April 2012

1 General description

This program was developed for the analysis of dyadic data or, to be more precise, data that are a mix of individual scores and scores of dyads between individuals. An example could be an analysis of the difference in extraversion between friends as compared to that of non-friends. One could perform an F-test, where the cases are all possible dyads in a class. The dyadic variable friends / non-friends would act as a fixed factor and the difference in extraversion as the dependent variable. However, there is a problem, usually indicated by the term *multiple membership*: the cases are not independent, since one single individual (with his specific score on the extraversion scale) can be a member of several dyads, and even worse: not all individuals feature in the same number of dyads. As a consequence, the distribution of the F-statistic may deviate from the F-distribution as it is assumed by a regular F-test.

In this program we use a randomization technique to estimate the actual form of the distribution and the corresponding probability. The program generates a large number of samples by taking the given set of dyads including the dyadic variable (*friendship* in our example) but with the links to the individuals randomized. Then for each sample the dependent variable is computed (the difference in extraversion) and the F-statistic is evaluated. Finally the resulting distribution of F-statistics is used to find the probability of the F from the original data.

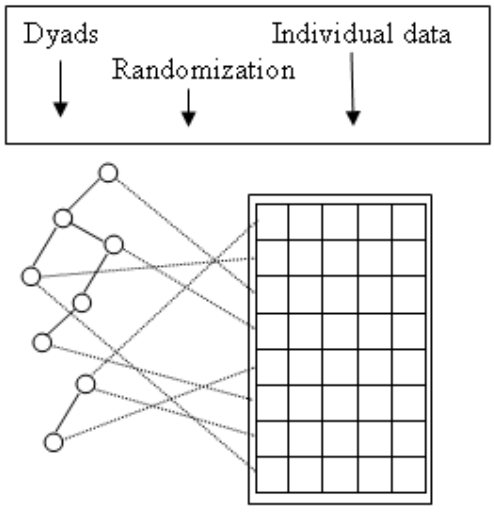

Figure 1: Randomization of individual data.

There are two data files involved: an *individual* file and a *dyadic* one.

The **individual** file contains the characteristics of the individuals or, as we will call them, cases. Each case must have an alphanumeric case identification field. This field must be the first variable in the corresponding list of variables. (It can contain up to 40 characters.) Usually the other variables will be continuous or interval data, since they will be used to construct the dependent variables in the analysis. The dependent variables will be combinations, each constructed from one individual score (v_1) of the first member of a dyad and one of the second member (v_2). V_1 and v_2 may refer to the

same individual variable. For these combinations several composition rules are available. Figure 2 gives an overview.

rule	definition
mean	$(v_1 + v_2) / 2$
sum	$v_1 + v_2$
difference	$v_1 - v_2$
reversed difference	$v_2 - v_1$
absolute difference	$\| v_1 - v_2 \|$
maximum	$\max(v_1,v_2)$
minimum	$\min(v_1,v_2)$

Figure 2: Composition rules.

It is one of these combinations that will act as the dependent variable in the analysis.

The **dyadic** file contains information on the dyads, including two case identification fields referring to the members of the dyad. These fields must match identification fields in the individual file. These identification fields must be the first variables in the corresponding list of variables. They must have the same length as the case identification field in the individual file. In addition the dyadic file contains one or more variables that together characterize the type of dyad, for instance *friends1* (member 1 nominates member 2 as a friend: 0 = no, 1 = yes), *friends2* (member 2 nominates member 1 as a friend: 0 = no, 1 = yes), *friends* (0 = none of the members nominates the other as a friend, 1 = one member nominates the other as a friend, but the other does not, 2 = both members nominate each other as a friend).

Dyads is a prototype version of a more general program. Therefore only one statistic is available for hypothesis testing: the F-statistic, using one of the dyadic variables as a fixed factor and one of the combinations as the dependent variable. The program performs a regular F-test ignoring any problems of multiple membership or distribution characteristics. But it also performs a randomization test using a bootstrapping approach: in a series of simulations it randomly links the individuals to the dyads, re-computes the aggregated variables and computes the F-statistic. This statistic will be denoted as $F_{observed}$ in the sequel. Thereby the distribution of the F-statistic is estimated under the null hypothesis that there is no relationship between the fixed factor and the dependent variable. By this approach the actual distribution of the variables is automatically taken into account, so it is no problem if the variables don't follow a F-distribution. The position of the original F-statistic in this distribution is interpreted as the probability that is used in the decision whether to accept or to reject the null hypothesis.

The number of permutations to be performed is based on three criteria that can be chosen by the user:

- a confidence level α
- a maximum **c** for the half range of this confidence interval.
- an absolute maximum $\mathbf{n_{max}}$.

The program will never perform more than $\mathbf{n_{max}}$ permutations. The actual number may be less depending on c and α:
We estimate the probability p of $F > F_{observed}$ (under the null hypothesis) by the proportion $\hat{p}$ of permutations producing an F-value greater than $F_{observed}$. So the

standard error of p can be estimated by:

$$SE(p) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

A confidence interval of level α is given by: $\hat{p} \pm Z(\alpha).SE(p)$, where $Z(\alpha)$ is the standard normal deviate with cumulative probability $\frac{1 + \alpha}{2}$.

The process of permutations will end as soon as the range of this confidence interval becomes less than 2c:

$$2Z(\alpha).SE(p) \leq 2c \quad \rightarrow \quad Z(\alpha)\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leq c \quad \rightarrow \quad n \geq \frac{Z^2(\alpha)\hat{p}(1 - \hat{p})}{c^2} \qquad (1)$$

Now the number of permutations is determined according to the following rules:

1. To find a provisional estimate of p, 1000 permutations are performed.
2. From the permutations thus far p is estimated by the proportion $\hat{p}$ of permutations producing an F-value greater than $F_{observed}$ and the permutations continue until the criterion in (1) is satisfied. To stay on the safe side in formula 1 $\hat{p}$ is replaced by $p^* = \min(\max(0.1, \hat{p}),0.9)$. After each permutation p^* is recomputed.
3. The process will always end as soon as the maximum number n_{max} of permutations is reached.

Examples:

- $p^* = 0.5$

Requirement: the 95% confidence interval must have as its range 2×0.001
$\alpha = 0.95$, $Z(\alpha) = 1.96$, $c = 0.001$
$n \geq 960400$

- $p^* = 0.05$

Requirement: the 95% confidence interval must have as its range 2×0.01
$\alpha = 0.95$, $Z(\alpha) = 1.96$, $c = 0.01$
$n \geq 3458$

The results of the analysis are stored in a listing file that borrows its name from the individual file and has ".lst" as its extension. Optionally the program also produces an SPSS-syntax file (with extension "sps") that can be used to show the distribution of the F-statistic.

2 The data

The input to *Dyads* comes from two separate file types: files with individual data and files with dyadic data. The dyadic files may be the result of the program MakeDyad[1] or otherwise they must be prepared in advance by a word processor or an editor of any kind and saved as "text only" (a flat ASCII file as it is often called).

Before the program is started one must prepare the individual and the dyadic files and retain their format. Most easily free format input is used. In that case one still must remember the order of the variables in both files. Note that in the individual files the first variable must refer to the case identification and in the dyad files the two first variables must be the two case identifications of the members of the dyad.

Each **individual file** contains one row for each individual. Such a row holds the scores on some individual variables. The first variable must be a case identification field. Note that in free format this field cannot contain spaces.

Figure 3 gives an example. The variables are (in order of appearance) the identification field, a sequence number, a group number, gender and the scores on the following 5 individual characteristics: Extraversion, Agreeability, Conscientiousness, Emotional stability and Openness.

```
....+....1....+....2....+....3....+....4....+....5....+....6..
0489282012  1 048 2    .664  -1.650  -1.661   -.311   -.688
0488282022  2 048 2  -1.445    .194    .808    .267   -.407
0483217031  3 048 1   -.771    .935   1.198    .690   -.266
0487281042  4 048 2   1.339   -.744  -1.659   -.029    .039
```

Figure 3: Four lines in an individual input file.

Each **dyadic file** contains one row for each pair of individuals that form a dyad. Such a row must start with the case identifications of the two members. Note that in free format these identification fields cannot contain spaces.

```
....+....1....+....2....+....3....+....4....+....5....+....6..
0489282012  0488282022  0
0489282012  0483217031  2
0489282012  0487281042  0
0489282012  0485160051  0
```

Figure 4: Four lines in a dyadic input file.

Figure 4 shows an example. The variables are: (1 and 2) the two identification fields, (3) the choice of the first member in the dyad towards the second. The choice indicates the type of nomination, like *friendship*, *antipathy*, *neutrality*, etc.

[1] MakeDyad is a KUNST program that reads sociometric data on individual level and writes them into a file on dyadic level.

Each single row in a file is represented by one or more lines in the file; although it will seldom happen that you need more than one line per row. The exact structure of such a row must be given by the *data definition* menu (see 6.2).

3 Files

There are five file types that are important for this program:

data files:
The files that contain the data to be analyzed: The *individual* files and the *dyadic* files. One program run can analyze several pairs of these files. So if you specify more than one set, make sure that the dyadic files are given in the same order as the individual files.

settings files:
These files are used to save the options as a user has specified them. A settings file contains all information about the analysis to be performed, including the description of the data, but not the data themselves. It is possible to have more than one settings file. By default their extension is '`.setdya`'.

listing files:
A listing file contains the main results of an analysis in a nice layout for human readers. There will be one listing file for each pair of data files. Its name will take its first part from the *individual* data file and end on '`.LST`' (or on '`1.LST`', '`2.LST`', ... and so on). You can inspect it by any editor, but in order to have an orderly layout you must view it in a small non-proportional font like Courier New 9.

SPSS-syntax files:
Depending on the chosen options the program may produce one file with SPSS commands for each pair of input files. These syntax files are meant to be input for SPSS. Their names will take their first part from the *individual* data file and end with '`.SPS`' (or '`1.SPS`', '`2.SPS`', ... and so on).

4 Installing the program on Windows

The installation of the program is very simple:

1. Copy the file “Dyads.exe” to any place on your hard disk. Optionally you may make shortcuts on the task bar and/or the desktop.

2. After the first time you have used the program, double click on the listing file (with extension “`.lst`”). Windows will ask you to select the program to be used when opening the file. Select a simple text editor like NotePad or WordPad.

3. After the first time you have saved the program settings, double click on the settings file (with extension “`.setdya`”). Windows will ask you to select the program to be used when opening the file. Select the program “Dyads.exe” or any shortcut to it.

That is all: from now on, you can start the program by double clicking the exe-file, one of its shortcuts or a settings file.

5 Starting the program

To run *Dyads* you must double click on its exe-file (Dyads.exe) or, if you have used the program before, on one of its settings files (for instance *Current.setdya*). The first thing you will see then is the main window of the program, as shown in figure 5.

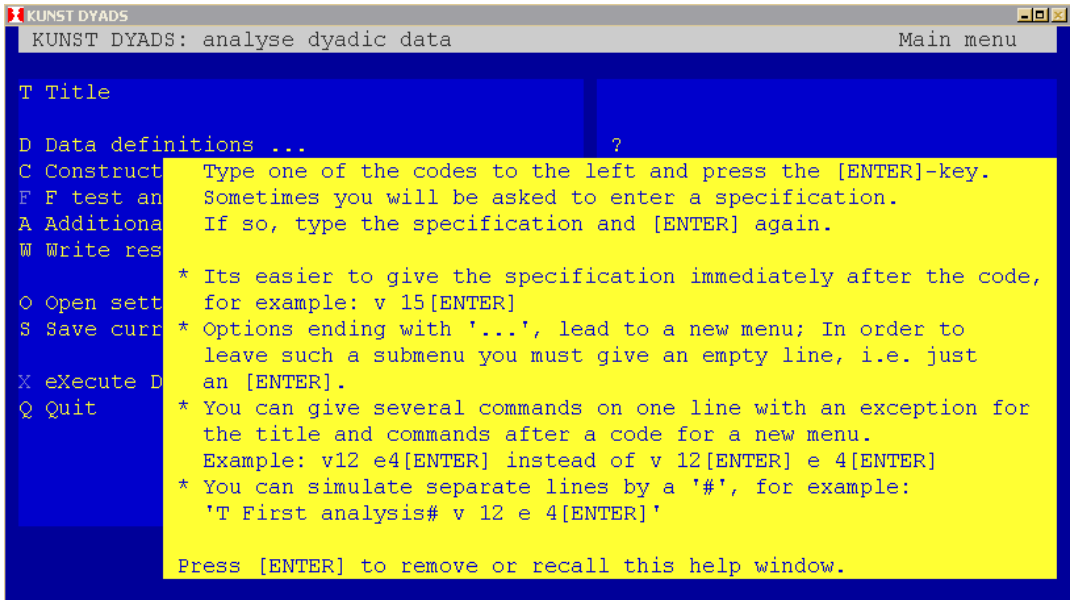

Figure 5: The main window for *Dyads*.

On the screen the light part with the text 'Type one ...' is a yellow text window. Windows like that contain hints and explanations. If you have read the text (or don't need it at all), you can press the Enter-key and the yellow window will vanish (see figure 6).

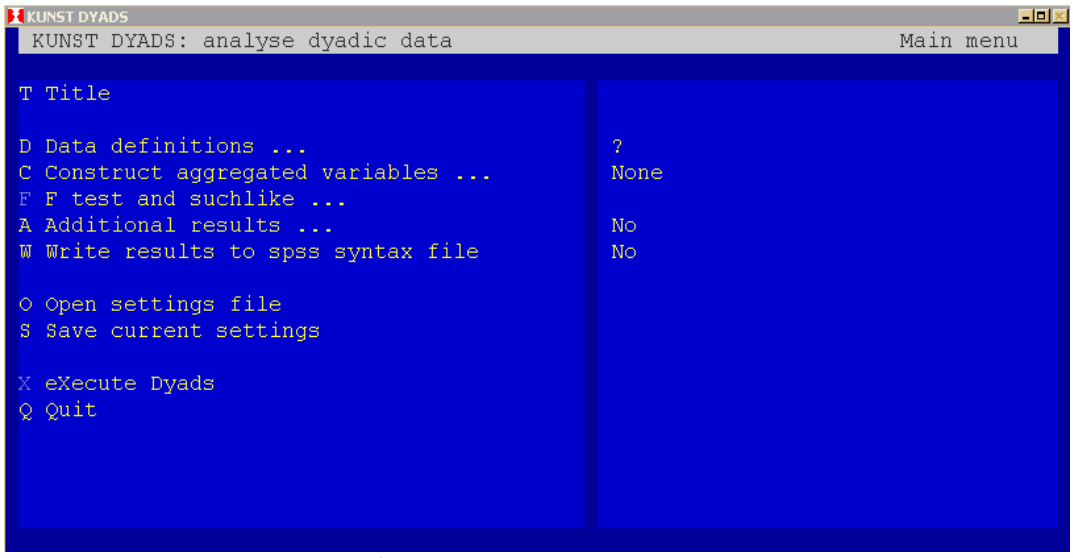

Figure 6: The main window for *Dyads* without help-window.

Now you can see the entire main window. The left part is the *main* menu. It consists of a list of options, each preceded by a one-character code. To select an option you must type the code, followed by the information you want to give and then followed by the Enter-key. From now on, we will denote the Enter-key as: Enter. You may for instance type: `T Analyzing first data`Enter to define the title that will appear as a header in the listing file. If you do not know what the meaning of an option is, you may just enter its code and Enter. There will appear a question on the screen and, if helpful, a yellow window to give you information. Press Enter to remove the yellow window and then give the specification belonging to the code or ignore the yellow window and give the specification at once. Do not repeat the code itself!

The right part of the window gives a short review of the options, as they currently are set. In the example of figure 6 you can see:

- No title line is defined.
- The data definitions are not yet specified, but must be given
- No constructed variables are defined as yet.
- An F-test cannot be used as long as the data definitions and the aggregated variables are not yet defined. (The code `F` is gray)
- No additional results will be reported
- No SPSS-syntax file will be produced

6 Menu options

6.1 The main menu

The *main* menu contains the following options:

`T Title`
This option allows you to specify a header to be used in the listing file.

`D Data definitions ...`
If you type `D`Enter the *main* menu will be replaced by the *data definitions* menu. This menu allows you to define the input data. It will be discussed in section 6.2.

`C Construct aggregated variables ...`
This option leads to a new menu that allows you to define new variables, or to modify existing aggregated variables. It will be treated in section 6.3.

`F F test and suchlike ...`
If you type `F`Enter the *F test* window will appear. By its options you choose the tests to be performed. This menu will be discussed in 6.4.

`A Additional results ...`
This option leads to a new menu, that allows you to ask for additional information in the listing file. It will be treated in 6.5.

`W Write results to SPSS syntax file`
By typing `W`Enter you switch the option from yes to no or the other way around. If it is `yes`, the F statistic will be written to a separate file along with the SPSS commands needed to read them.

`O Open settings file`
If you have ever saved the options for *Dyads* or if you received a settings file from someone else, you can retrieve the options from the settings file. If you type `O`Enter, a file-selector box will appear on the screen that allows you to select the settings file. By default settings files from *Dyads* have the extension "`.setdya`". It may be handy to save your current settings before collecting new ones. The program may remind you of that. After you have collected information from a settings file, its name will be visible on the upper right part of the main window.

`S Save current settings`
If you want to save the options and specifications that you have made so far, you can enter `S`Enter If you do so a file-selector box will appear, that allows you to specify the place and the name of the file to which the settings must be written.

`X eXecute Dyads`
If you have specified all options you can type `X`Enter to start the computations. The program will check if all obligatory options are specified and if there are no inconsistencies. If everything is right, the computations will start. If the program is correctly installed it will, when it is finished, automatically open the last (or only)

listing file it has made. If it fails to do so, you can open it yourself by any text editor like *WordPad*, *NotePad* or *Word*. In order to have a nicely outlined text, you must select a small non-proportional font like Courier New 9.

`Q Quit`
The option Q is a kind of emergency exit. If you choose it, *Dyads* will halt without performing any calculations and without producing any output files.

6.2 Definition of the data

From the main menu, you may type the option `D`Enter to enter the *data definitions* menu. After you have filled out the options in this menu you must press Enter to return to the main menu.

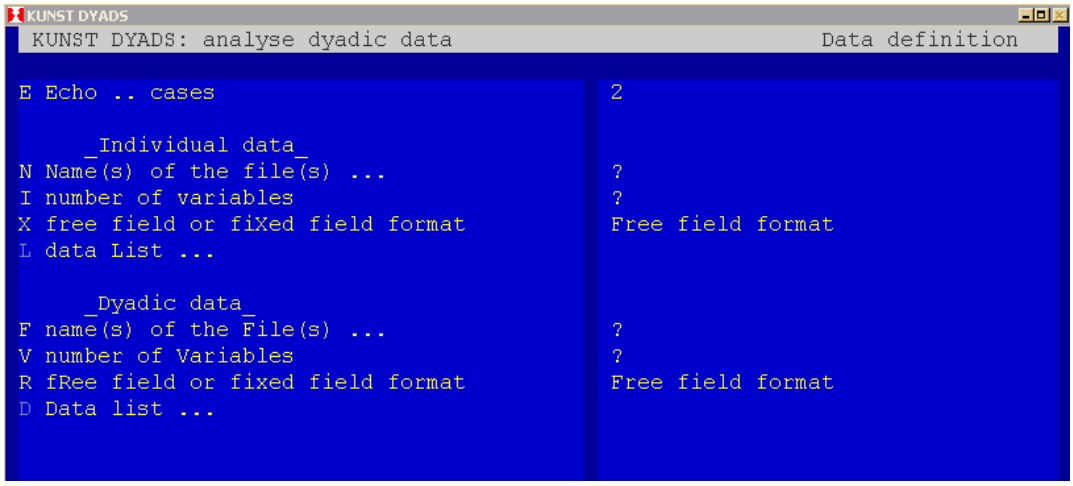

Figure 7: The *data definition* menu.

In the *data definition* menus (see figure 7) the following options can be chosen:

`E Echo first .. cases`
If you type `e ##`Enter the first ## individuals in the individual files and the first ## dyads in the dyadic files will be shown in the listing file. This may help you to check if the input specifications are correct.

6.2.1 Options pertaining to the individual data

`N Name(s) of the file(s)`
If you choose this option a file-selector box will appear that enables you to select one or more *individual* input files.

`I number of variables`
By this option you must specify the number of variables in the individual file. Note that the first variable must always be the case identification.

`X free field or fiXed field format`
This option defines where the values in a row of the individual file are to be found:

In a *free* format, the positions of the variables may be different from row to row, although there order must be always the same. The user can still choose whether line

numbers are specified or not.
In general, the use of free format is highly recommended since it is less sensitive to irregularities in the data or mistakes in the specifications.

In a *fixed* format, each row consists of the same number of lines (most probably just one) and each variable has a precisely defined position in that line. This format is especially useful if the variables are typed in fixed columns and the missing values are replaced by spaces. You must use this format also when the case identification may contain spaces.

Each time you type `X`Enter the choice switches.

`L data List ...`
If you choose this option, a new window will open with a layout that differs from the usual menu layout. Its layout depends on the chosen format type. If the input file contains one single line per case and the scores are separated by spaces, commas or tabs, you can select the free format option and skip the data list. Note that the name of the first variable is already given because it must be the case identification.

Figure 8 shows a data list for an individual file with fixed format data. Now you can type one of the indicated characters (C, T, N, L, D, M) followed by a sequence number or a range of sequence numbers and then followed by the corresponding information. The possibilities can best be clarified by some examples:

You type: `c1 1-10`Enter
Thereby you specify that the first **c**olumn (variable) be in positions 1 through 10. From now on, you can leave the code c out; it will be assumed as long as you do not select another code.

You type: `4 19`Enter
This means that variable 4 occupies position 19.

You type: `5-7 20-43`Enter
This means that variables 5, 6 and 7 occupy positions 20 through 43, so variable 5 is in position 20-27, variable 6 is in positions 28-35 and variable 7 in positions 36-43, each occupying 8 positions.

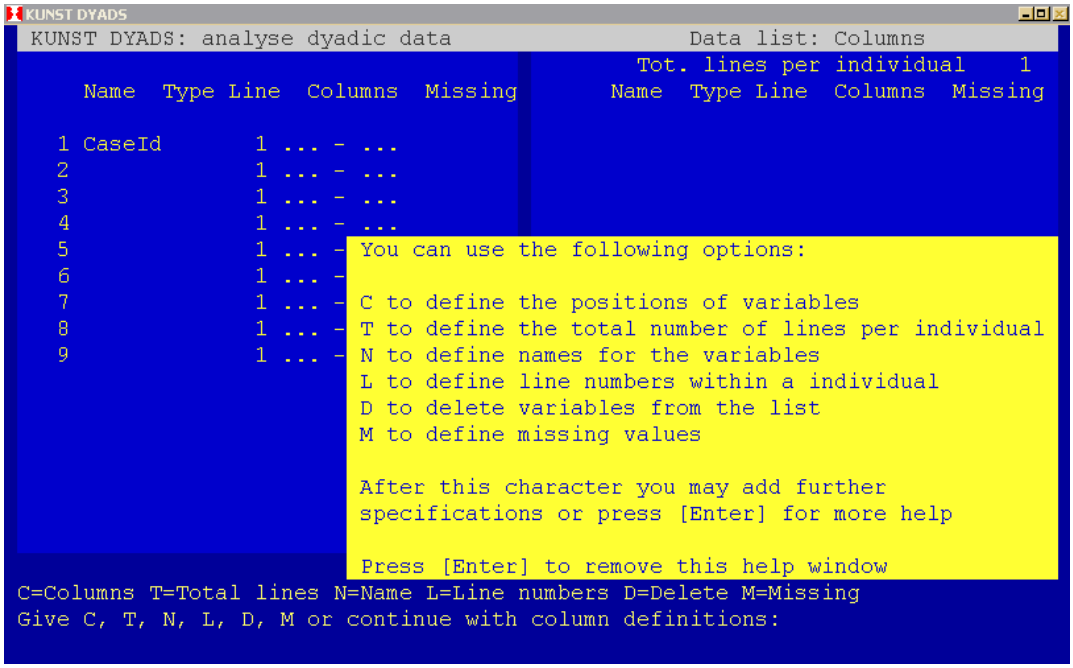

Figure 8: The data list for fixed format individual data.

You type: `L6 2`Enter

Now you have entered a different code. The code **L** indicates that you are defining line numbers. You specify that variable 6 is contained in the second line of a data row. Line numbers must be in ascending order. Therefore, the program will adjust the line numbers of variable 7 and higher to 2 if they are still 1. From now on, you may leave the code L out, until you switch to another code.

You type: `12-16 3`Enter

You specify that variables12 through 16 are contained in the third line of a data row.

You type: `N2 SeqNr`Enter

The code **N** indicates that you are specifying names. Variable 2 will receive the name 'SeqNr'. Names will be truncated to 8 characters.

You type: `5-7 Scores`Enter

Variables 5 through 7 will all be called 'Scores'.

You type: `5-7 Score1`Enter

Variables 5 through 7 will be called '`Score1`', '`Score2`' and '`Score3`'. As you see, if the name ends on a number subsequent names will have their numbers adjusted.

You type: `D5`Enter

The option **D** allows you to remove a variable from the list. All sequence numbers and all other definitions will be adjusted accordingly. In this example, you remove the fifth variable. You can also specify a range of variable numbers, for instance '`D5-7`' to remove the 5-th, 6-th and 7-th variable.

You type: `M4 9`Enter

With the option **M** you define the value to be treated as missing: If the data of variable 4 contains 9 **or any higher value** it will be treated as missing.

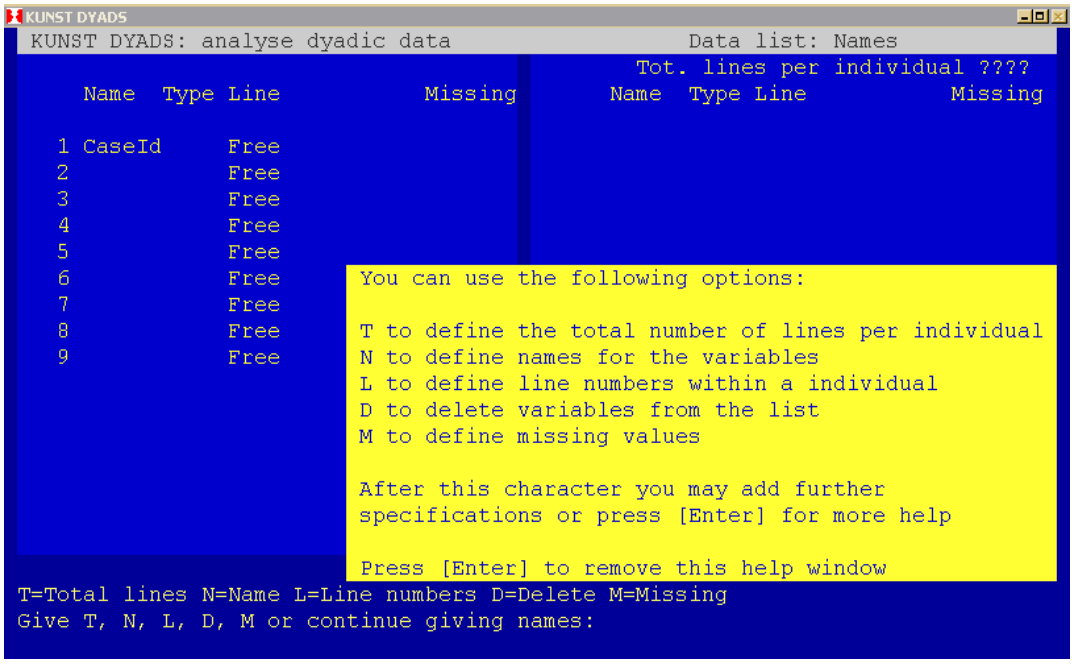

Figure 9: The data list for free format individual data.

You type: T4Enter
If you don't use the option **T**, the program will assume that the number of lines for each data row is equal to the line number of the last variable in the data list. If a row contains more lines than that, you must specify so by the option T. In this example, you specify that there are 4 lines in each row.

You type: FEnter or BEnter
If there are more than 32 variables, they will not fit at once on the data list screen. Therefore, you have the possibility to scroll **f**orward and **b**ackward with the options F and B.

If you use free format data (see figure 9), you do not need to specify anything in the data list. However, you may specify on which line of an individual data row each variable is recorded. So you can use the option L. The options P, T, N and D are also available. Their meaning and use are the same as with fixed format data (see above).

If you have finished the data list, you can go back to the *data* definition menu by entering an empty line (just Enter). If the yellow window is still visible, you must enter two empty lines (Enter Enter): one to remove the yellow window and one to return to the *data definitions* menu.

6.2.2 Options regarding the dyadic file

`F name(s) of the File(s)`
If you choose this option(see figure 7) a file-selector box will appear that enables you to select one or more dyadic input files. If you choose more than one, you must make sure that they are in the same order as the *individual* files.

`V number of Variables`
By this option you must specify the number of variables in the dyadic file. Note that

the first two variables must always refer to the case identifications of the dyad members. Moreover, only the third variable in the data list will play the role of independent variable in the analysis. Any more variables will be ignored. If there are variables between the two case identifications and the independent variable, you must use the fixed field format to specify the its precise location.

`R fRee field or fixed field format`
This option defines where the values in a row of the individual file are to be found: See the section on the *individual* file (6.2.1) for more information.

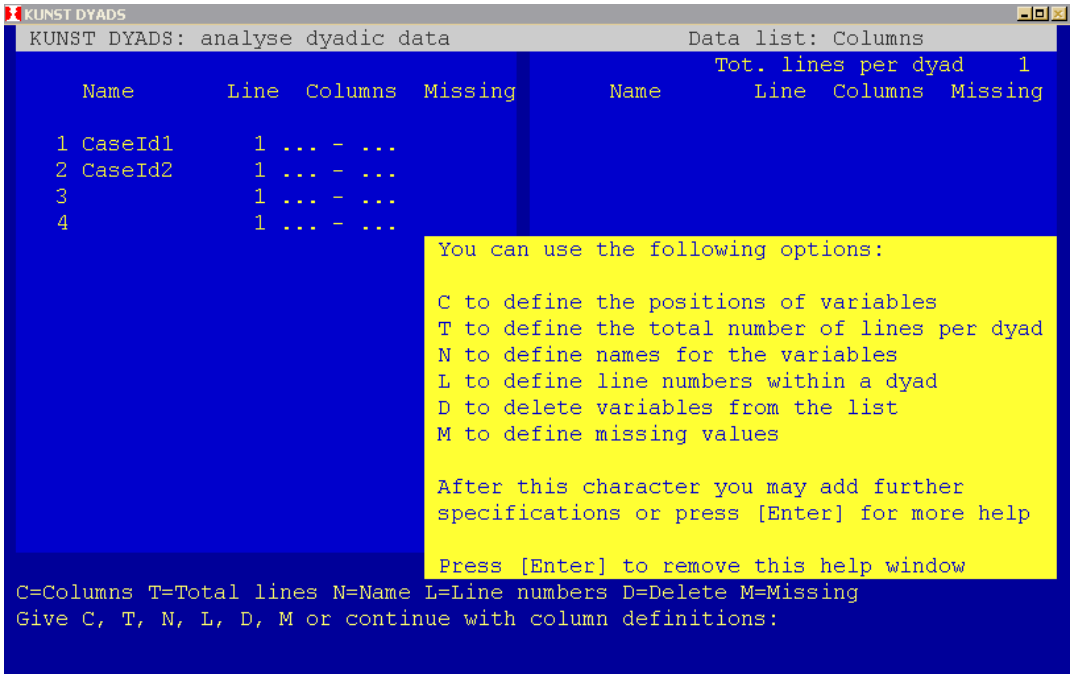

Figure 10: The data list for fixed format dyadic data.

`D Data list ...`
If you choose this option, a new window will open with a layout that differs from the usual menu layout. Its layout depends on the chosen format type. If the input file contains one single line per dyad and the scores are separated by spaces, commas or tabs, you can select the free format option and skip the data list. Note that the names of the first two variables are already given because they must be the case identifications of the two members of the dyad.

Figure 10 shows a data list of the dyadic file for fixed format data. In figure 11 the data list for free format dyadic data is shown. The options to be used are the same as in the individual file (see 6.2.1).

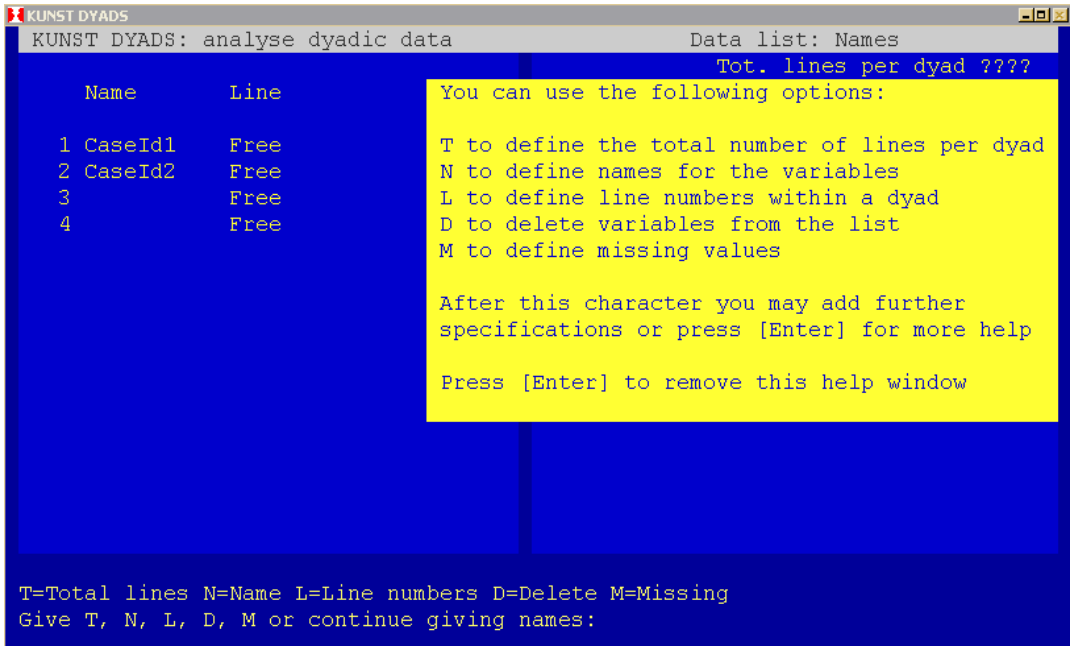

Figure 11: The data list for free format dyadic data.

6.3 Construction of aggregated variables

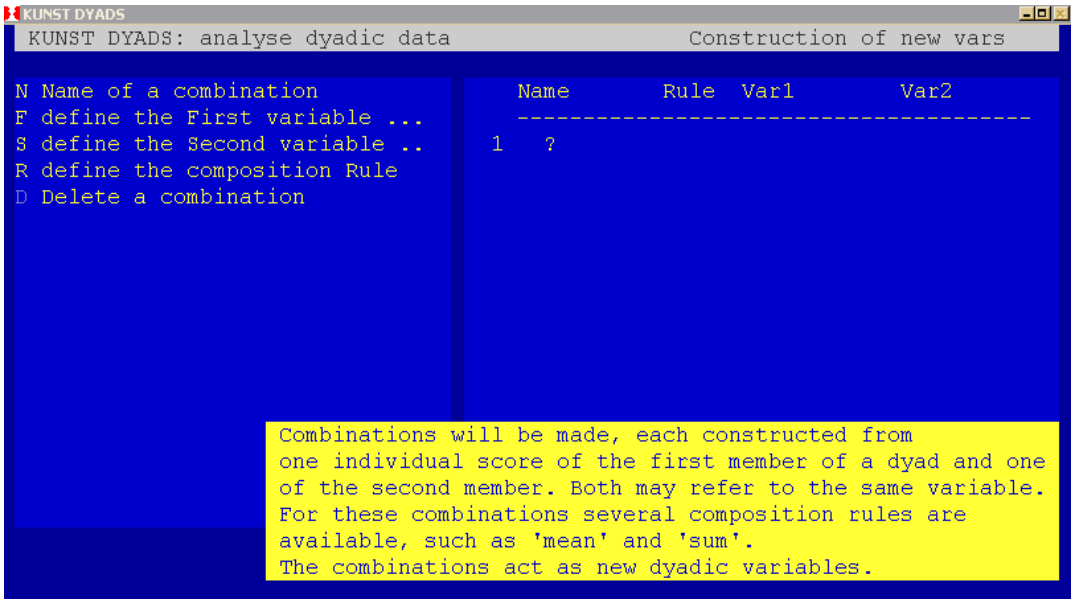

Figure 12: The *construct aggregated variables* menu

From the *main* menu, you may type the option `C`Enter to enter a submenu that handles the construction of new aggregated variables (see figure 12). After you have filled out the options in this menu you must press Enter to return to the *main* menu. Before you make new combinations make sure that the variables in the *individual* file and in the *dyadic* file are defined properly.

On the right side of the screen a list is given of the aggregates as they are defined so far. For each combination the list contains a sequence number, a name, a combination rule

and the number (and name) of the two variables from the individual data that are combined.

The aggregated variables are combinations of two individual variables to form a new variable on dyadic level.

On the right side of the screen a list is given of the aggregates as they are defined so far (see Figure). For each combination the list contains a sequence number, a name, a combination rule and the number (and name) of the two variables from the individual file that are combined.

In this menu the following options can be chosen:

`N Name of a combination`
With this option you can give a name to a new combination or change the name of an old one: just type `N` followed by a combination number and a name, like
`N 1 MeanEmot`Enter to give the first (new) combination the name *MeanEmot*.

`F define the First variable ...`
`S define the Second variable ...`
Option `F` is used to define (or redefine) the *individual* variable that belongs to the first member in the dyad. Option `S` is used to define (or redefine) the *individual* variable that belongs to the second member in the dyad. When you type `F##`Enter or `S##`Enter, where ## is the number of a combination, the screen will reveal a list of all individual variables. If the combination already exists an individual variable will already be marked by an arrow as in figure 13.

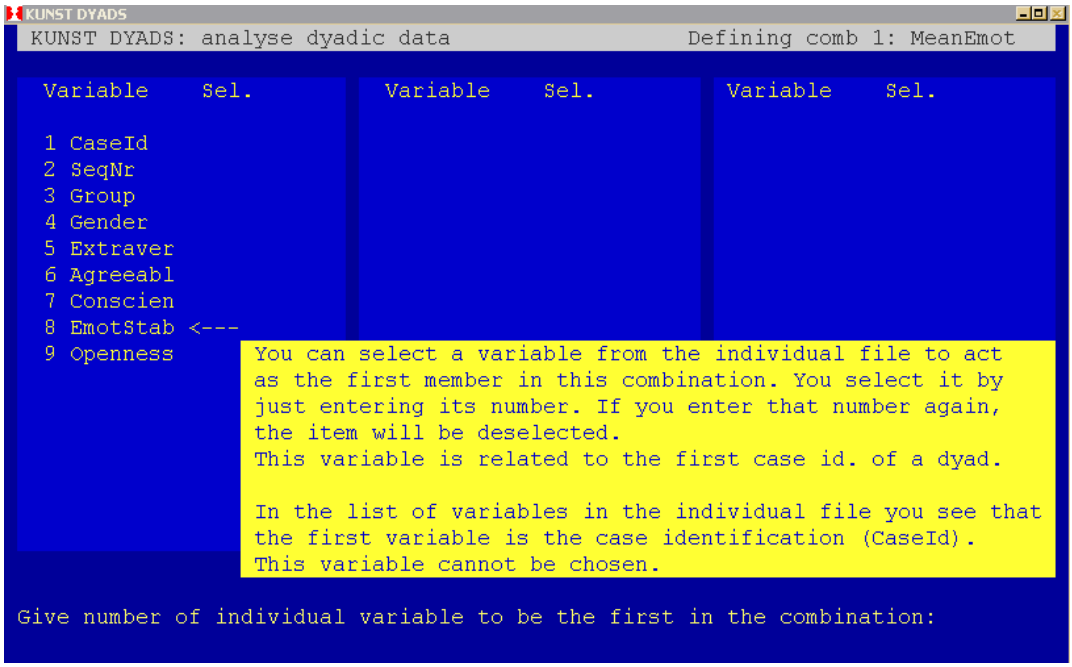

Figure 13: Definition of the first member in a combination.

Now you can choose the variable by typing its number. Note that you cannot choose the case identification.

You can leave the list of *individual* variables by entering an empty line (just Enter). If the yellow window is still visible, you must enter two empty lines (Enter Enter): one to remove the yellow window and one to return to the *construct* menu.

`R define the composition Rule`
In the *construct* menu the option **R** can be used to define the rule by which the two individual variables must be combined. The following rules are available:

m the mean of the 2 variables: `(v1 + v2)/2,`
s the sum: v1 + v2,
p the product: `v1 * v2`
d the difference: `v1 - v2`
r the reversed difference: `v2 - v1`
a the absolute difference: `v1 - v2`
h the highest value: `max(v1,v2)`
l the lowest value: `min(v1,v2)`

To define a composition rule type `R#1 @`Enter where #1 is the number of the combination to receive the rule and @ is the rule, so one of the characters above.

`D Delete a combination`
Within the construct menu the option `D` can be used to delete a combination or a range of combinations. Type `D #`Enter or `D #1-#2`Enter, where # is the sequence number of a combination and `#1-#2` a range of sequence numbers.

In figure 14 you see the *construct* menu after two combinations have been defined.

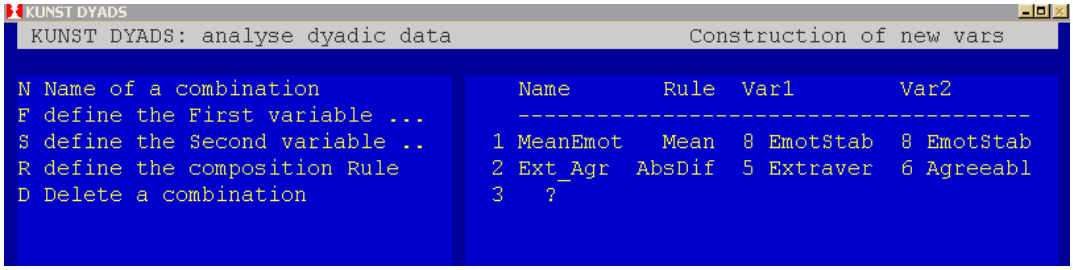

Figure 14: The *construct* menu after two combinations have been made.

As you can see the first constructed dyadic variable is called *MeanEmot* and is defined as the mean score of the dyad members on the individual variable *EmotStab*. The second constructed variable is called *Ext_Agr* and is formed by the absolute difference between the score of the first dyad member on the individual variable *Extraver* and the score of the second dyad member on the individual variable *Agreeabl*.

6.4 Significance tests

In the Main-menu the option `F` leads to a submenu that controls the tests (see figure 15). After you have filled out the options in this menu you must press Enter to return to the *main* menu. If you want the program to do anything, you must enter here a fixed factor (a dyadic variable, acting as a grouping variable) and a dependent variable (one of the aggregated variables).

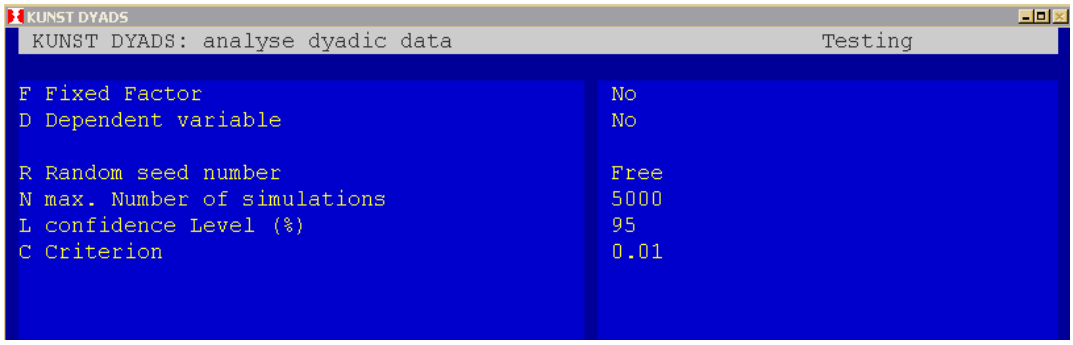

Figure 15: The *testing* menu.

The *testing* menu offers the following options:

`F Fixed Factor`
When you type `F`Enter, the screen will reveal a list of all dyadic variables (not the aggregated ones) as in figure 16. If the fixed factor was already defined, an arrow will mark the corresponding dyadic variable.

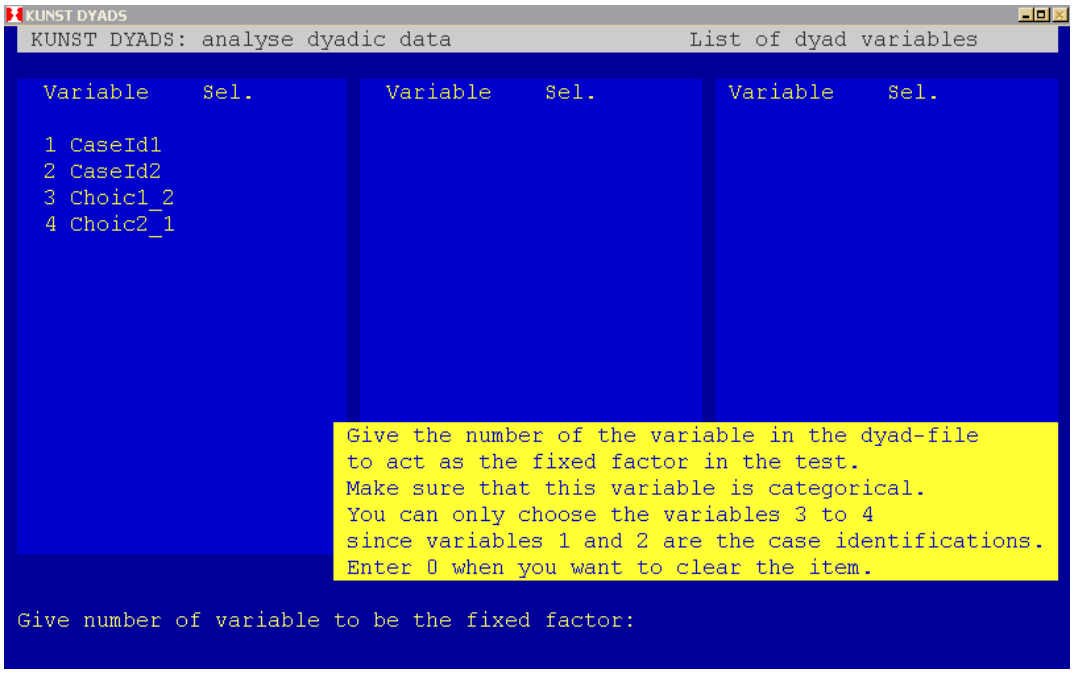

Figure 16: List of dyadic variables to define the fixed factor.

Now you can choose the variable by typing its number. Note that you cannot choose the first two variables, since they are the case identifications.

You can leave the list of dyadic variables by entering an empty line (just Enter). If the yellow window is still visible, you must enter two empty lines (Enter Enter): one to remove the yellow window and one to return to the testing menu.

`D Dependent variable`
When you type `D`Enter, the screen will reveal a list of all aggregated variables. If the dependent variable was already defined, it will be marked by an arrow as in Figure .

Now you can choose the variable by typing its number.

You can leave the list of aggregated variables by entering an empty line (just Enter). If the yellow window is still visible, you must enter two empty lines (Enter Enter): one to remove the yellow window and one to return to the *testing* menu.

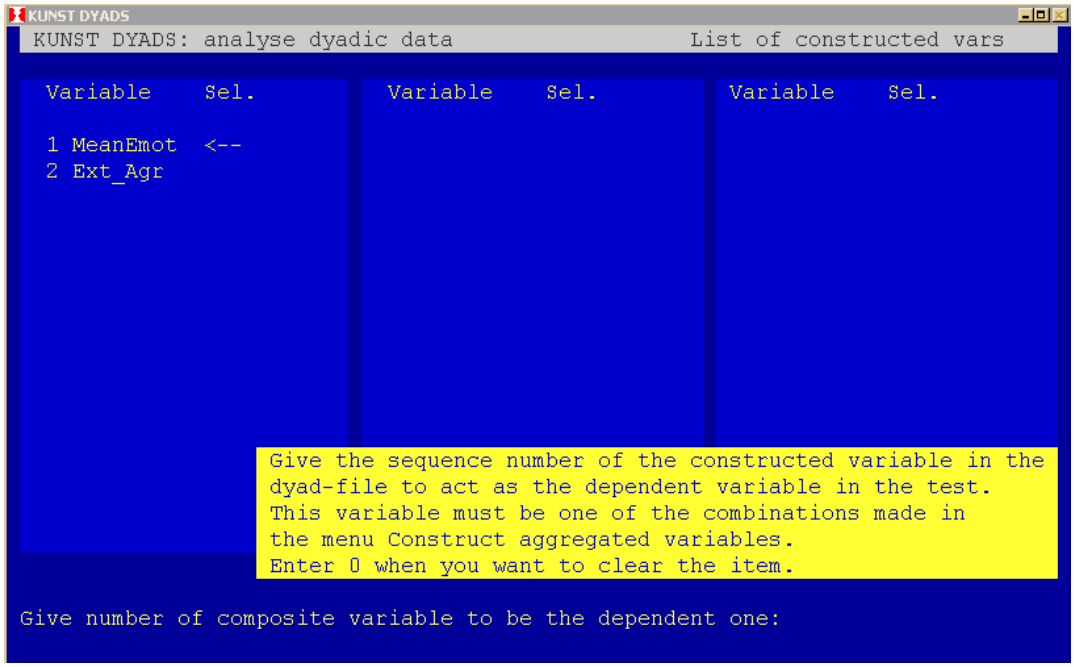

Figure 17: List of aggregated variables to define the dependent variable.

`R Random seed number`
The connections between individuals and dyads will be permuted in a random way. The randomization may start from a fixed number each time you call the program. This number is called the seed. If you type `r ##`Enter the randomization process will start with the number ##. If you type `r free`Enter, the seed number is chosen at random by the program.

The number of permutations to be performed depends on three rules, as described in section 1. The maximum number of permutations and the criterion c can be specified by the options **N** and **C**, the confidence level by option **L**.

`N max. Number of simulations`
By typing `n ##`Enter you specify the maximum number of permutations to be performed. You can use it to restrict the time needed for the analysis.

`L confidence Level (%)`
By typing `L ##`Enter you specify the level for the confidence interval around the estimated probability for the observed F-statistic. The value must be given as a percentage, i.e. a number between 0 and 100. By default it will be 95. This number will be used to determine the number of permutations to be performed as long as the number of permutations is between its minimum and its maximum.

`C Criterion`
This option is used to control the number of permutations to be performed. You may type C ##Enter where ## is a number between 0 and 1. When the range of the L-level confidence interval is less than or equal to 2 times this criterion, the simulation process

will end, unless less than 1000 permutations have been performed. A good criterion might be 0.01. Note that the effect of this criterion c on the running time is quadratic: if you divide c by x, the execution time needed is roughly multiplied by x^2.
If you do not want to control the number of permutations by this option specify
`C 0`Enter and the maximum number of permutations will be performed.

6.5 Additional results in the listing file

The option **A** in the *main* menu leads to the *additional results* menu (see figure 18). The options in this menu offer the probability to get additional information in the listing file in addition to what is given by default.

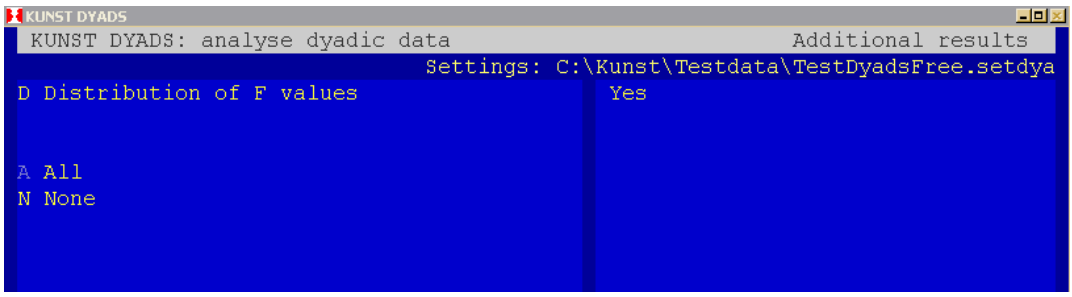

Figure 18: The additional results menu

At present this menu contains only the following option, but others may be added later:

`D Distribution of F values`
By typing `D`Enter you switch the option from `yes` to `no` or the other way around. If it is yes, the listing file will contain the distribution of the F values.

`A All`
`N None`
If you type `A`Enter, all options will be switched to `yes`.
If you type `N`Enter, all options will be switched to `no`.

7 Results

After execution of *Dyads* you will find one or two new files in your working directory:

- the listing file (.lst)
- if you asked for results in an SPSS-syntax file (W in the main menu) the SPSS-syntax file (.sps)

7.1 Results in the listing file

The listing file will contain the main results. Its precise content depends on the chosen options and the input data. The lines in the listing have a length of 80 characters or less. View (and print) the file with a small non-proportional font like Courier 9.

- The file starts with an overview of the options as they are chosen by the user:
  The definitions of the *individual* data
  The definitions of the *dyadic* data
  Other options.
- The F-statistic computed from the data, with a significance test based on the F-distribution without taking care of the problem of multiple memberships.
- A non-parametric significance test based on the distribution of the F value resulting from the bootstrapping procedure:
  - The distribution of the F-statistic
  - Summary of the distribution of the F-statistic.

7.2 Results in the SPSS-syntax file

If output to an SPSS-syntax is asked (the option W in the *main* menu) it will contain a text like the following:

```
DATA LIST FREE / F.
VAR LABELS F 'F value of the permutation'.
FORMATS F (F10.4).
BEGIN DATA
      5.9429
      0.6465
      1.3852
      ....
      ....
      ....
END DATA.

FREQUENCIES
  VARIABLES=F  /FORMAT=NOTABLE
  /STATISTICS=MINIMUM MAXIMUM SEMEAN
  /PERCENTILES= 0.95
  /HISTOGRAM
  /ORDER=ANALYSIS .
```

8 Example run

In this section an example is given with the **individual** data from the file "DyadsExampleIndividuals.dat" (see figure 19).

The columns 1 to 6 contain the case identification. They are followed by a case number, a group number and the gender. The next 5 numbers are the scores on the variables Bully, Pvic, Antisoc, Socwith and Prosoc.

Because all variables in a case are separated by blanks, the data will be read in free format.

```
....+....1....+....2....+....3....+....4....+....5....+....6..
200101  1 1 2  .76615 -1.03366   .05997 -1.58968  -.62500
200102  2 1 1  .09529  -.73601 -1.04756   .40971   .23105
200103  3 1 1 -.24014  -.14072 -1.06455 -1.60996   .36795
200104  4 1 1 -.24014  -.73601 -1.04756 -1.56629   .71840
200105  5 1 1 1.10158  -.43837  -.04303  -.17202   .40628
200106  6 1 2  .43072   .75222  -.80758  1.20822  -.12123
200107  7 1 2  .76615  1.34751  1.82585   .39723  -.61040
200108  8 1 1 -.57557  1.94280   .13961   .79025  -.61040
200109  9 1 1 -.24014  -.73601 -1.04756 -1.56629  1.28971
200112 10 1 2 -.91100 -1.03366  -.80439 -1.58500  1.55072
200113 11 1 1 -.91100  -.73601  -.80545  1.06942   .74578
200114 12 1 1 -.07242  -.43837   .26597  -.29522  -.13766
200115 13 1 2 -.24014 -1.03366  -.27027 -1.60528  1.40105
200116 14 1 1  .34686 -1.03366 -1.04756  -.63677  -.61040
200117 15 1 2 -.57557  -.73601 -1.04756 -1.56629  1.47224
200118 16 1 1 -.91100  -.73601 -1.04756  -.59934  -.26908
200120 17 1 2  .76615  -.43837   .87654 -1.58968 -1.14521
200121 18 1 2 -.57557  -.43837 -1.04756 -1.60528  1.06885
200122 19 1 2  .09529   .15693  -.78103  2.29369  -.13583
200123 20 1 2 1.10158   .45457 -1.07198  -.38724 -1.18536
200124 21 1 2 -.57557  -.73601  -.08763  -.35917  -.16504
200125 22 1 2 -.24014  -.43837  1.82585 -1.60996  -.30376
200126 23 1 2 1.77244  -.73601  1.61985  -.70696 -1.65081
200127 24 1 1  .43072  1.64516  -.78846   .63429  -.13766
200128 25 1 1 -.24014  -.14072 -1.04756   .62337  -.61040
200129 26 1 1 -.07242 -1.03366 -1.06455 -1.62399 -1.17076
200131 27 1 1 1.43701  1.94280  -.78846   .94309   .41905
200132 28 1 2  .09529  -.43837  -.80545   .08064  -.09567
200133 29 1 2 -.24014  1.64516  1.31828 -1.29180  -.61040
200134 30 1 1 -.91100  -.73601 -1.05499  -.25312  1.45398
200135 31 1 1 -.91100   .15693  -.81607 -1.01419  -.13766
200137 32 1 2 -.91100 -1.03366 -1.04756 -1.17483  -.09567
200138 33 1 2  .76615  -.73601   .45498 -1.59436 -1.16529
```

Figure 19: The individual data for the example

The **dyadic** data come from the file "DyadsExampleDyads.dat" (see figure 20). It contains two case identification variables followed by a variable indicating the dyad type.

....+....1....+....2....+..

```
200101 200120 1
200102 200134 1
200103 200118 1
200104 200118 1
200104 200129 1
200104 200131 1
200105 200108 1
200105 200128 1
200106 200132 1
200107 200108 2
200107 200131 2
200107 200137 1
200108 200125 2
200108 200126 2
200109 200113 1
200109 200125 2
200112 200114 1
200112 200115 1
200113 200116 1
200114 200135 1
200115 200117 1
200116 200126 1
200117 200121 1
200117 200134 1
200121 200124 1
200122 200132 1
200122 200137 1
200123 200133 1
200124 200132 1
200124 200138 1
200125 200133 1
200127 200131 1
200128 200135 1
```

Figure 20: The dyadic data for the example.

Figure 21 through 27 show the menus after all the options have been specified.

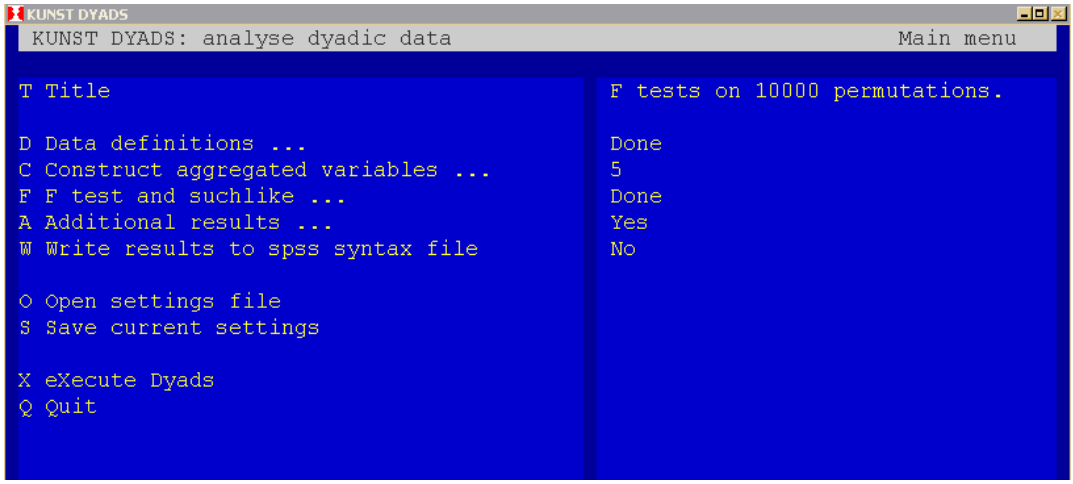

Figure 21: The Main menu for the example.

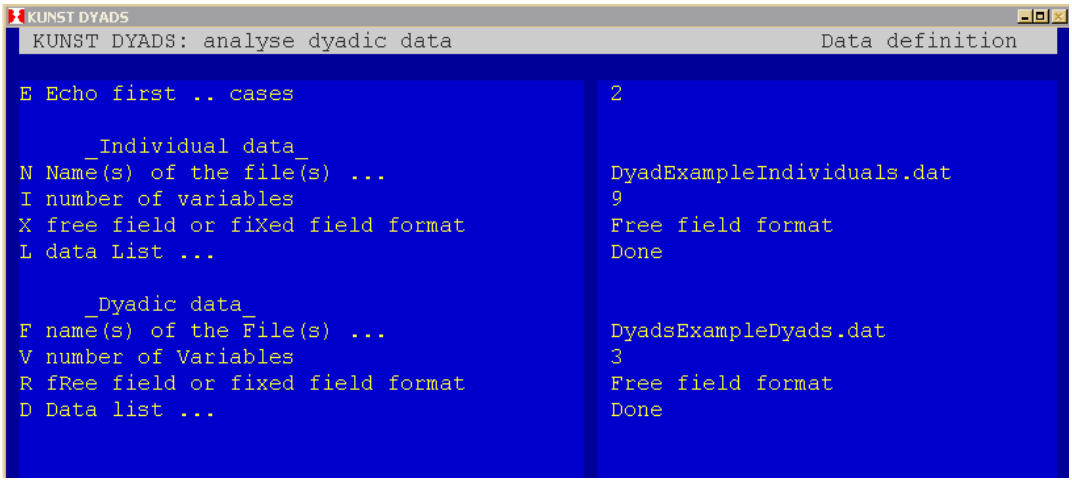

Figure 22: The Data definitions menu for the example.

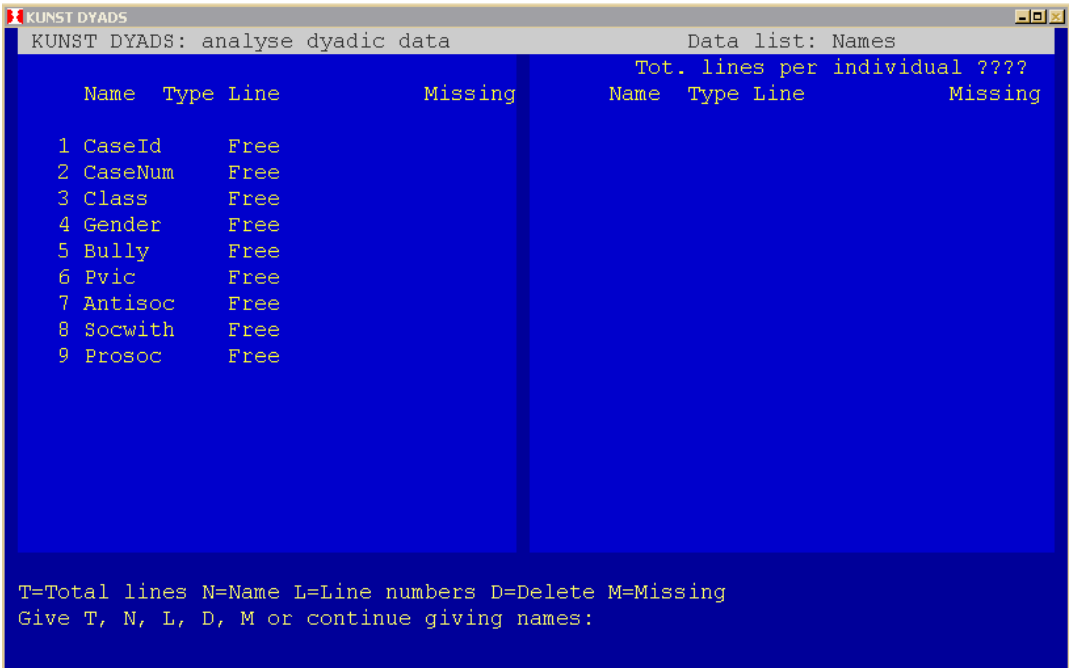

Figure 23: The Data list of the individuals for the example.

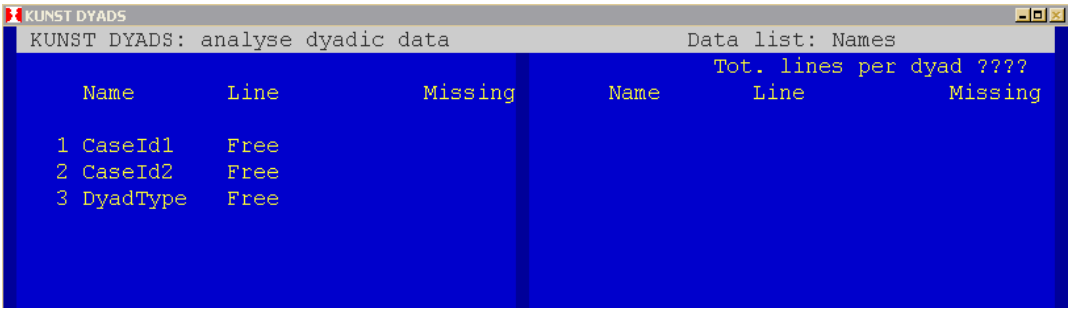

Figure 24: The Data list of the dyads for the example.

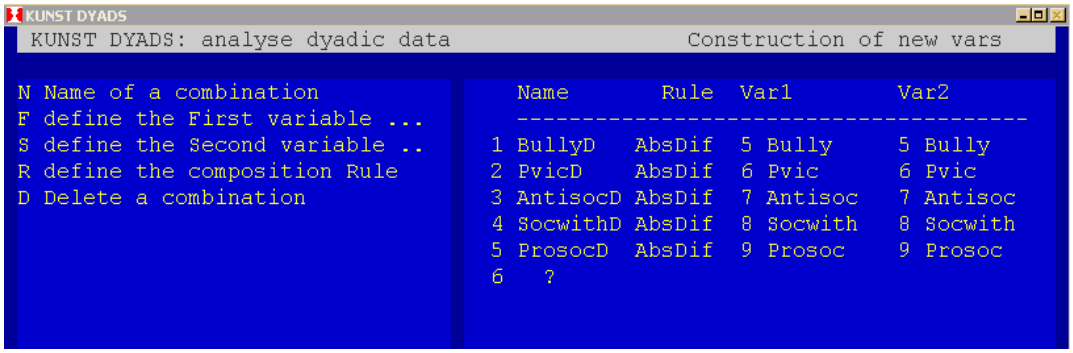

Figure 25: The construction of aggregated variables for the example.

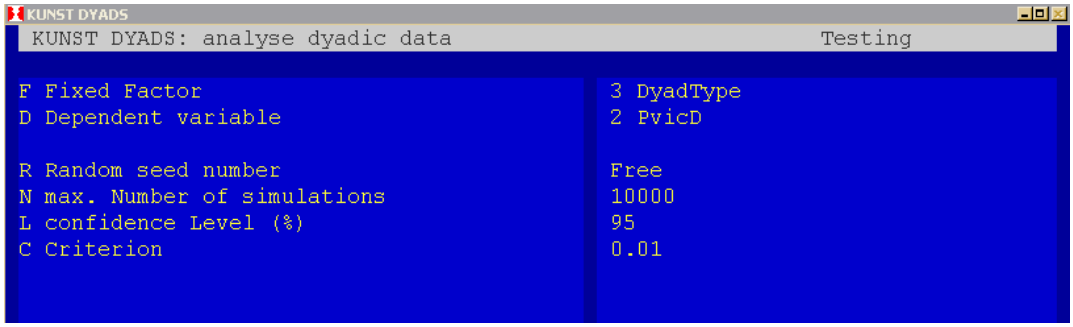

Figure 26: The testing menu for the example.

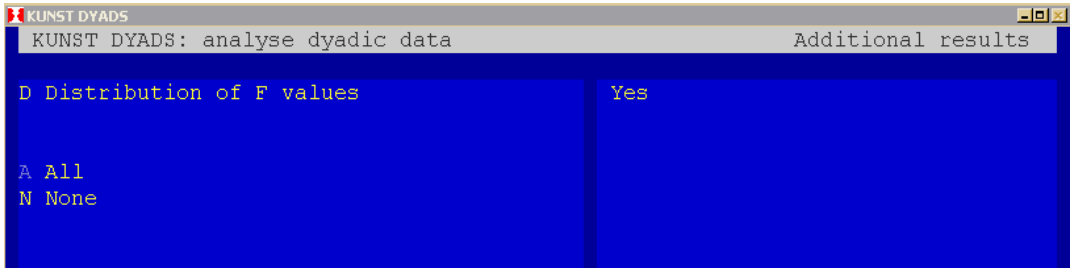

Figure 27: Additional results in the example.

After execution the program will have produced a listing file that looks as follows:

```
View (and print) this file with a small non-proportional font like Courier 9.
Lines contain 80 characters or less.
--------------------------------------------------------------------------------


* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * 17-11-2003  15:16:24

    Dyads: KUNST program for the analysis of dyadic data.
           Interactive version  0.10, June 2003

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

The operation and the accuracy of the program are not guaranteed.


F tests on 10000 permutations.
================================================================================

  1.   Summary of the user specifications and input data.

================================================================================
```

```
There are two types of input data: the individual data and the dyadic data.


1.1  The individual input data
_____________________________


The number of variables is 9.

The data are in completely free form.

The selected data are supposed to contain the following variables:

Var.   Name

    1 CaseId
    2 CaseNum
    3 Class
    4 Gender
    5 Bully
    6 Pvic
    7 Antisoc
    8 Socwith
    9 Prosoc

The individual data will be read from the following file:
  C:\Kunst\Testdata\DyadExampleIndividuals.dat


1.2  The dyadic input data
__________________________


The number of variables is 3.

The data are in completely free form.

The selected data are supposed to contain the following variables:

Var.   Name

    1 CaseId1
    2 CaseId2
    3 DyadType

The dyadic data will be read from the following file:
  C:\Kunst\Testdata\DyadsExampleDyads.dat


1.3  Other options
__________________


Aggregated variables in the dyadic data will be constructed from
characteristics in the individual data:

New variable          Rule              Indiv. var. of         Indiv. var. of
                                          CaseId1                CaseId2

 1 BullyD   is   abs. difference  of      5 Bully        and     5 Bully
 2 PvicD    is   abs. difference  of      6 Pvic         and     6 Pvic
 3 AntisocD is   abs. difference  of      7 Antisoc      and     7 Antisoc
 4 SocwithD is   abs. difference  of      8 Socwith      and     8 Socwith
 5 ProsocD  is   abs. difference  of      9 Prosoc       and     9 Prosoc

An F-statistic is to be performed on the data:
  The groups are determined by the values in variable 3 DyadType.
  This is the fixed factor.
  The dependent variable is the aggregated variable 2 PvicD.
The confidence level for the critical value of the one-tailed F-statistic
is 95%.
In a bootstrapping approach this F-statistic will be compared to 10000
permutations of the individuals.
```

```
The process of permutations may be ended if the range of the 95-level
confidence interval becomes less than 2 times the criterion of 0.01, but
a minimum of 1000 permutations will be performed.

In addition to the standard results this listing will contain:
  - Distribution of F values


1.4  Reading the individual input data
_____________________________________


Data of first individuals:

  1:                      Id: 200101
               1.000  1.000  2.000  0.766 -1.034  0.060 -1.590 -0.625
  2:                      Id: 200102
               2.000  1.000  1.000  0.095 -0.736 -1.048  0.410  0.231

The number of individual cases read is 33.


1.5  Reading the dyadic input data
_________________________________


Data of first dyads:

  1:                      Id1: 200101
                          Id2: 200120
               1.000

  2:                      Id1: 200102
                          Id2: 200134
               1.000

The number of dyads read is 33.



F tests on 10000 permutations.
================================================================================

  2.  The F-statistic for the data

================================================================================

The dependent variable is the aggregated variable 2 PvicD.
There are 2 groups according to the values in variable 3 DyadType.
The numbers of cases per group are:
   Group Value      Cases
      1     1         28
      2     2          5

The F-statistic:
                     SS      df       MS         F        p
   Model            1.680     1      1.680      2.381    0.13297
   Error           21.875    31      0.706
   Total           23.555    32



F tests on 10000 permutations.
================================================================================

  3.  Bootstrapping approach

================================================================================

In a series of maximal 10000 simulations, the dyad members are randomly linked
to the individuals, each time recomputing the aggregated variables.
The minimum number of 1000 permutations will be performed.
```

For each permutation of the individuals an F-statistic will be computed.
The randomization process is started with a randomly chosen seed number.

3.1 The distribution of the F values

	Categories of F values		Count	Cumulative proportion	
1	<	0.9	3525	0.531	
2		1.9	1223	0.715	
3		2.8	757	0.829	
4		3.8	504	0.905	
5		4.7	263	0.945	
6		5.7	148	0.967	
7		6.6	77	0.979	*
8		7.6	38	0.984	
9		8.5	30	0.989	
10		9.5	14	0.991	
11		10.4	19	0.994	
12		11.4	12	0.996	
13		12.3	5	0.996	
14		13.3	4	0.997	
15		14.2	3	0.997	
16		15.2	3	0.998	
17		16.1	0	0.998	
18		17.1	3	0.998	
19		18.0	2	0.999	
20		19.0	3	0.999	
21		>	6	1.000	

The total number of permutations is 6639

* Critical value at 95% level is 6.338

3.2 Summary of the F values

Minimum F value	0.0
Maximum F value	29.8468
F value of the data	2.381
Critical value at 95% level	6.3384
Proportion of F values higher than F data	0.2222
Maximum for the half range	0.01
Half range of the 95% level confidence interval after 6639 permutations	0.01

======= Normal end of analysis =======

9 Literature

Bendermacher, A.N.H., & Thissen-Pennings, M.C.E., (1999) *MAKEDYAD: A program to create a file of dyads*, RekenTechnische Ondersteunings Groep (RTOG), Faculty of Social Sciences, Radboud University of Nijmegen.

Thissen-Pennings, M.C.E., (1999) *SOCSTAT: A program for sociometric status*, Reken-Technische Ondersteunings Groep (RTOG), Faculty of Social Sciences, Radboud University of Nijmegen.

Content