Raters3

Estimating inter-rater agreement between 3 raters with nominal classifications

Nol Bendermacher & Pierre Souren
April 2012

1 Introduction

The program *Raters3* is primarily meant to measure inter-rater reliability in the situation where three raters classify a sample of stimuli, or as we will call them: cases, into a restricted number of nominal categories. The categories must be mutually exclusive and exhaustive. It is assumed that the raters make their classifications independently in the sense that they do not exchange any information and receive no common information except that they are presented with the same cases and possibly have received the same instruction. As an example, suppose that three raters are asked to classify 500 young birds into three subspecies. The resulting frequency matrix might be as given in table 1.

	rater 3 = 1			rater 3 = 2			rater 3 = 3		
	rater 2			rater 2			rater 2		
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
1	37	16	19	32	21	13	0	2	7
rater 1: **2**	19	11	7	30	103	38	9	11	16
3	5	7	2	10	22	11	11	13	28

Table 1: Example of an observed three-way matrix.

Raters3 computes three measures that estimate the qualities of the three raters. Pairwise combinations of these measures give measures for the pairwise inter-rater reliabilities. The most widely used measure for such pairwise inter-rater reliabilities is Cohen's kappa (Cohen 1960), but there are serious objections to that measure. These objections are overcome by the measure s.

Raters3 is based on a formal model. The model starts with **n** cases classified by three raters into **c** categories. So the basic input is a c by c by c frequency distribution that is given directly or constructed from the raw ratings by the raters. The population distribution of these categories is defined as the vector $\mathbf{V} = (V_1,V_2,...,V_c)$ where V_i is the probability that a case belongs to the i-th category. The model assumes that there is a fixed probability $\mathbf{p_r}$ that rater r makes a correct observation, where r takes the values 1, 2 and 3. The expression *correct observation* needs to be taken very strictly: it means not only that the classification is *correct* but also that it is based on a true *observation* and not just on a correct guess. In other words, a correct classification can be based on a *correct observation* but also on a *correct guess*. So the model distinguishes three types of classifications: (1) a correct observation, (2) a correct guess and (3) a wrong guess.

The products $\mathbf{s_{12}} = p_1.p_2$, $\mathbf{s_{13}} = p_1.p_3$ and $\mathbf{s_{23}} = p_2.p_3$ are the definitions of inter-rater agreement for the three pairs of raters.

If a rater r performs a correct observation, the probabilities of the categories are given by the population distribution V. However, if he does not, he may follow a different distribution **W_r** = $(W_{r1}, W_{r2}, \ldots, W_{rc})$ where W_{ri} is the conditional probability that rater r chooses the i-th category. The distinction between V and the W's is theoretically important, because V is a population characteristic, whereas the W's are properties of the classification procedure, including the raters.

Thus the complete set of parameters consists of the scalars p_1, p_2, p_3 and the vectors V, W_1, W_2 and W_3. These parameters are supposed to be statistically independent, except for the fact that the elements of V, as well as those of W_1, W_2 and W_3 add up to 1:

$$\sum_{i=1}^{c} V_i = \sum_{i=1}^{c} W_{1i} = \sum_{i=1}^{c} W_{2i} = \sum_{i=1}^{c} W_{3i} = 1.$$

Although p_r is defined as the probability of a correct observation, the probability of a correct *classification*, including a 'correct guess' may be also interesting. Once the parameters are estimated, it is straightforward to derive estimates for these probabilities also. They are denoted as p_1^+, p_2^+ and p_3^+: $p_r^+ = p_r + q_r \sum_{i=1}^{c} V_i W_{ri}$ with $q_r = 1 - p_r$.

Table 2 shows the parameter estimates from the frequencies in table 1.

$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{s}_{12}$	$\hat{s}_{13}$	$\hat{s}_{23}$	$\hat{V}$	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_3$
0.4754	0.3524	0.6692	0.1676	0.3181	0.2358	0.3805	0.2032	0.2666	0.0000
						0.3580	0.6057	0.4333	0.9698
						0.2615	0.1911	0.3001	0.0302
		Cohen's kappa:	0.1815	0.3302	0.2429				

Table 2: Parameter estimates from table 1.

From the model the joint population distribution X can be derived, where X is a c by c by c matrix, such that X_{ijk} is the expected proportion of cases that is classified in category i by the first rater, in category j by the second and in category k by the third.

The parameters can be estimated by maximizing the likelihood of the observed matrix $\hat{X}$ of observed proportions:

$$2n\left(\sum_{i=1}^{c} \sum_{j=1}^{c} \sum_{k=1}^{c} \hat{X}_{ijk} \ln\left(\frac{\hat{X}_{ijk}}{X^{*}_{ijk}} \right) \right)$$

where X^{*}_{ijk} is the reconstruction of X from the estimated parameters.

More detailed information about the formulas and algorithms involved can be found in the deocumentation of the program Raters2 and in Bendermacher & Souren (2009).

For large samples this likelihood is distributed as χ^2 with c^3-4c+1 degrees of freedom, so it can be used to test the model fit. In the example of table 1 we find $\chi^2(16) = 22.9018$ with a probability of 0.1164.

The measure s_{ij} is equivalent to Cohen's kappa for the pair i,j if $W_i = W_j = V$. In the example of table 1 the kappa values are close to the corresponding estimated s-values.

If a rater confuses certain categories, such confusion will lead to violation of the model assumption that the probabilities $p_{1,}$ p_2 and p_3 are the same for all categories. Such violations may be detected by the model test. The confusion would also make the vectors W_r dependent on the distribution of the true categories, whereas under the model assumptions they are purely characteristics of the raters.

3 Files

Four file types that are important for this program.

data files:
The files that contain the data to be analyzed. One program run can analyze several files. These files may contain all raw data or all frequency matrices.

settings files:
These files are used to save the options as they are specified by a user. 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 '`.setra3`'.

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 data file. Its name will take its first part from the data file and end on '`.LST`' (or on '`1.LST`', '`.2LST`', ... 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.

raw output files:
Depending on the chosen options the program may produce one file with 'raw' output for each input file. This raw file is meant to be input to other programs. Its name will take its first part from the data file and end with '`.OUT`' (or '`1.OUT`', '`2.OUT`', ... and so on).

4 Installing the program on Windows

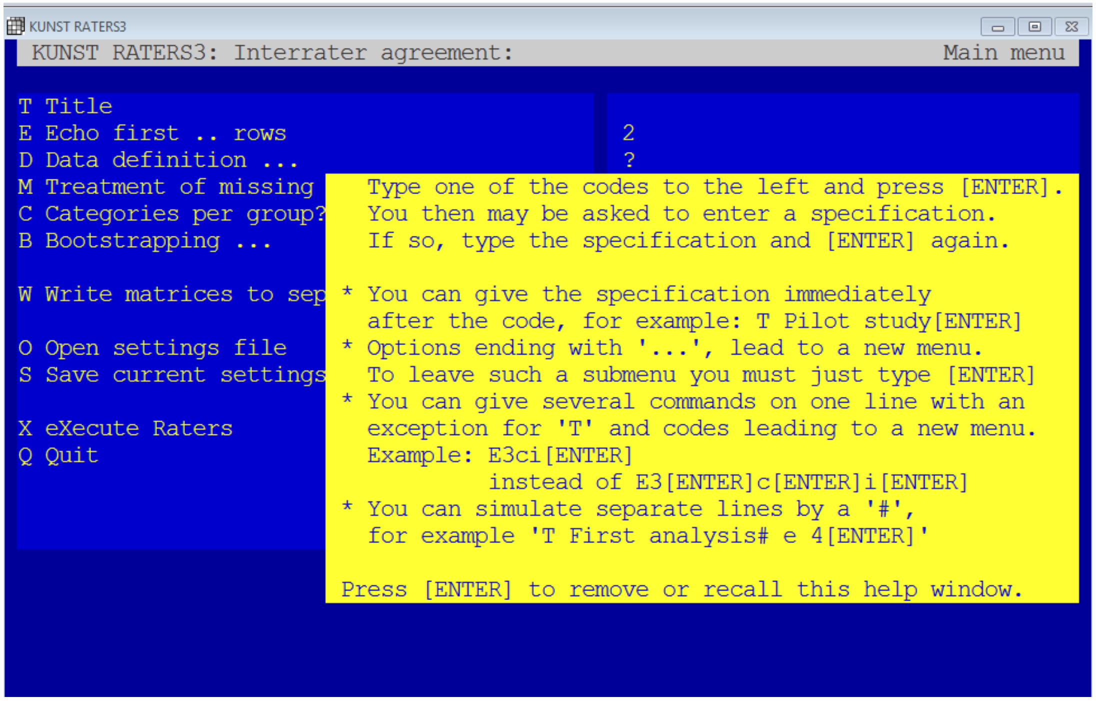
The installation of the program is very simple:

1. Copy the file *Raters3.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. Windows will ask you to select the program to be used when opening the file. Select a simple text editor like ScratchPad or WordPad.
3. After the first time you have saved the program settings, double click on the settings file. Windows will ask you to select the program to be used when opening the file. Select the program Raters3.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 with the extension '`.setra3`'.

5 Starting the program

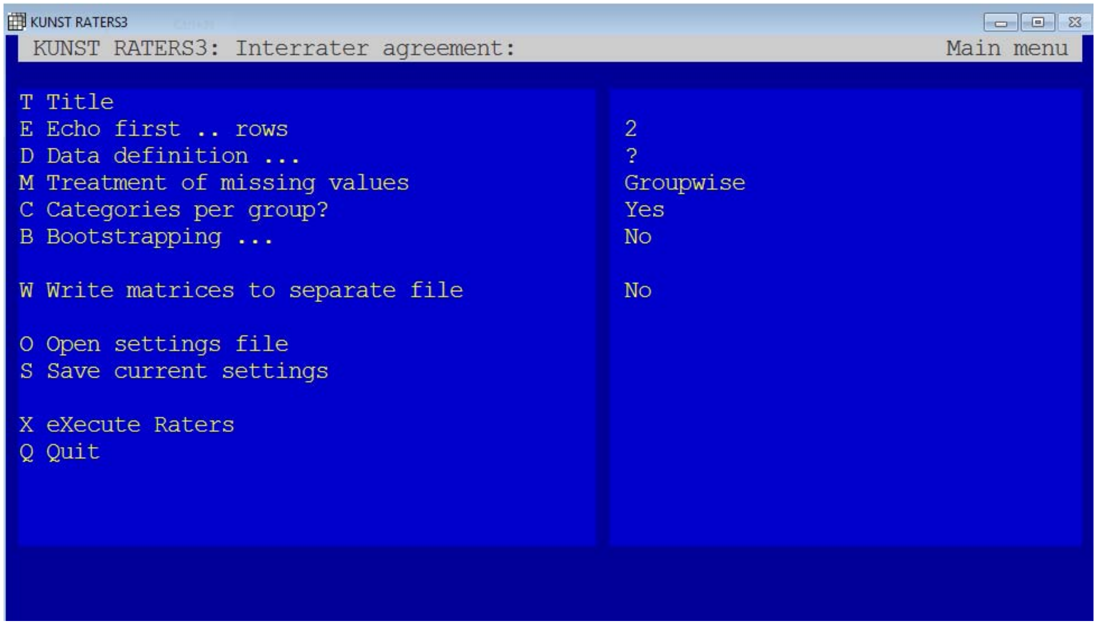
To run Raters3 you must double click on its exe-file (*Raters3.exe*) or a shortcut to it. If you have used the program before, you can also double click one of its settings files (for instance *Current.Setra3*). The first thing you will see then is the main window of the program, as shown in figure 1.



```
 KUNST RATERS3: Interrater agreement:                                Main menu

T Title
E Echo first .. rows                         2
D Data definition ...                        ?
M Treatment of missing    Type one of the codes to the left and press [ENTER].
C Categories per group?   You then may be asked to enter a specification.
B Bootstrapping ...       If so, type the specification and [ENTER] again.

W Write matrices to sep * You can give the specification immediately
                          after the code, for example: T Pilot study[ENTER]
O Open settings file    * Options ending with '...', lead to a new menu.
S Save current settings   To leave such a submenu you must just type [ENTER]
                        * You can give several commands on one line with an
X eXecute Raters          exception for 'T' and codes leading to a new menu.
Q Quit                    Example: E3ci[ENTER]
                                   instead of E3[ENTER]c[ENTER]i[ENTER]
                        * You can simulate separate lines by a '#',
                          for example 'T First analysis# e 4[ENTER]'

                        Press [ENTER] to remove or recall this help window.
```

Figure 1: The main window for Raters3.

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 2).



```
KUNST RATERS3
KUNST RATERS3: Interrater agreement:                    Main menu

T Title
E Echo first .. rows                      2
D Data definition ...                     ?
M Treatment of missing values             Groupwise
C Categories per group?                   Yes
B Bootstrapping ...                       No

W Write matrices to separate file         No

O Open settings file
S Save current settings

X eXecute Raters
Q Quit
```

Figure 2: The main window for Raters3 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 session``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 more 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!

Options that end with three periods refer two submenus.
Options may be disabled because they can not be combined with choices made for other options or because they need some other information that is not given yet. These options have their option character and possibly there information on the right side of the screen in pale blue.

The right part of the window gives a short review of the options as they currently are set. In the example of figure 2 you can see:

- No header line is defined.
- In case of raw input the first two cases will be shown in the listing file.
- The input data still have to be defined (hence the question mark).
- Rows with missing values are excluded groupwise.
- The set of categories will be defined per group of raters.
- No bootstrapping will take place.
- No raw output files will be produced.

6 Options in the main menu

The main menu contains the following options:

`T Title`

This option allows you to specify a header line to be used in the listing file.

`E Echo first .. rows`

By this option you can ask the program to show in the listing file the first input cases of each input file. Such an 'echo' helps to check if the data are correct. The option applies only to raw input data (see next section). You can enter a number or the word '`all`'.

`D Data definition ...`

If you type `D`[`Enter`] the main menu will be replaced by the data definition menu. This menu allows you to define the input data. It will be discussed in the next section.

`M Treatment of missing values`

This option is used to define how missing values must be treated. A missing value is an input value used to indicate that the true value is unknown. Missing values are only allowed if the input consists of raw data, i.e. scores assigned by raters to individual cases. The three available options are:

Groupwise: If for a case a score is missing (so it contains a missing value), the case will be ignored in all tables in the group to which this rater belongs. How groups of raters are formed is explained in the next section.

Listwise: If for a case any score is missing the case will be ignored in all tables for the current input file.

Triadwise: In the construction of any table all cases will be counted that have non-missing scores for the three raters involved.

You can circle through these options by repeatedly entering `M`[`Enter`] . You can also type several `M`'s and one single Enter: `MM`[`Enter`].

`C Categories per group`

With raw data, Raters3 will check the data matrix to find out which categories occur in the data for each table to be made. When a table is prepared, there are two possibilities:

Categories per pair: The table will contain all categories that are used by any of the three raters.

Categories per group: The table will contain all categories that are used by any of the raters in the group to which the three raters belong. How groups of raters are formed is explained in the next section.

Each time you type `C`[`Enter`] the choice switches.

`B Bootstrapping`

Bootstrapping is a technique used to investigate the distribution over samples of the parameters and the model fit criterion. It is used to estimate standard errors and confidence intervals for the parameters and to assess the model fit. Raters3 will always try to estimate the standard errors from the (observed) information matrix, but sometimes this technique fails. Moreover, even if standard errors can be estimated from the information matrix, the bootstrapping option may be useful:

- the estimated standard errors may be biased when parameters have estimates close to their theoretical boundaries.
- with skewed distributions the standard errors cannot be used to estimate confidence intervals.

- The model test is based on the likelihood ratio test, assuming that this criterion has a χ^2 distribution, but that assumption holds only for large samples and expected frequencies that are not too small (say 5 or more). The bootstrapping procedure gives a model test based on the same criterion but without any assumption about its distribution.

If you type `B` `Enter`, the main menu will be replaced by the bootstrapping menu. This menu allows you to define the options that pertain to the bootstrapping procedure. It will be described in section 8.

`W Write matrices to separate file`
With this option you can ask the program to write the frequency distributions to separate files. This option will only work if the input consists of raw data. For each input file one output file will be produced, that borrows its name from the input file, except that the name will end on '`.out`' or '`1.out`', '`2.out`' and so on. Each matrix will start with an empty line, followed by one line for each row. Each number will occupy 6 positions. Each time you type `W` `Enter` the option switches from `Yes` to `No` or back.

`O Open settings file`
If you have ever saved the options for Raters2 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 Raters2 have the extension '`.Setra2`'. It may be convenient to save the current settings before collecting new ones. After you have collected information from a settings file, the name of that settings file will be visible on the upper right part of the main window.

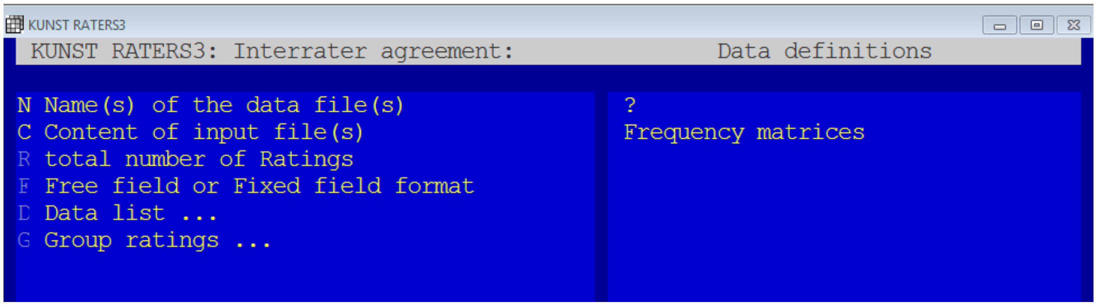
`S Save current settings`
If you want to save the options and specifications that you have made so far, you can enter `S` `Enter`. If you do so, a file selector box will appear, that allows you to specify the place and the name of the file to which the settings must be written.

`X eXecute Raters3`
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*, *ScratchPad* 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, Raters3 will halt without performing any calculations and without producing any output files.

7 Options in the data definitions menu

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.



KUNST RATERS3

KUNST RATERS3: Interrater agreement: Data definitions

```
N Name(s) of the data file(s)          ?
C Content of input file(s)             Frequency matrices
R total number of Ratings
F Free field or Fixed field format
D Data list ...
G Group ratings ...
```

Figure 3: The data definitions menu.

In the data definitions menu (see figure 3) the following options can be chosen:

`N Name(s) of the data file(s)`
If you choose this option a file selector box will appear that enables you to select one or more input files. All files must contain the same type of data: either a raw data matrix or a collection of frequency tables, depending on the option `C`.

`C Content of input file(s)`
This option allows you to choose from two possible types of input: raw data, i.e. a matrix of cases by raters, or frequency tables. Each time you type `C`[`Enter`] the choice switches.

`R Total number of Ratings`
This option is only available with raw data (see the option `C`). It is needed to define the total number of ratings, i.e. columns in the raw data. Type "`R` ##[`Enter`]", where ## is the number of ratings or type "`R`[`Enter`]" (a yellow window appears) and then "##[`Enter`]", where again ## is the number of ratings (the yellow window disappears).

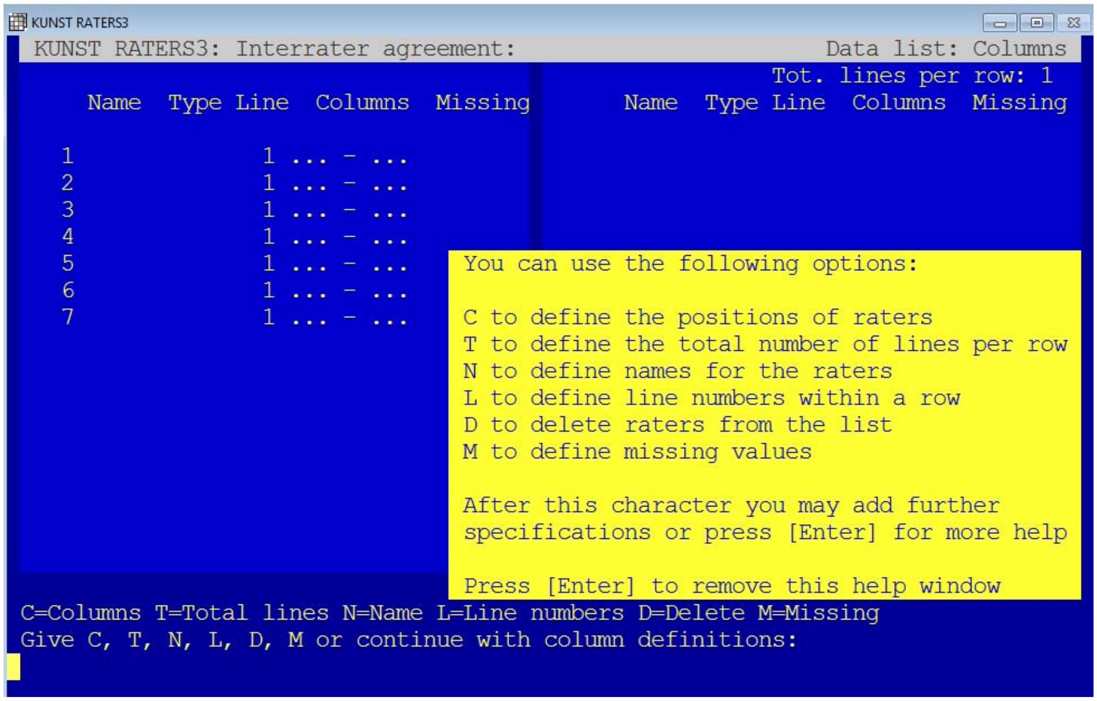
`F Free field or Fixed field format`
This option is only available with raw data (see option `C`). It defines where the values in a case are to be found.

In a fixed format, each row consists of the same number of lines (most probably just one) and each value has a precisely defined position in that line. This format is especially useful if the values are typed one after another without intervening spaces or commas.

In a free format, the positions of the values may be different from case to case, 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.
Each time you type `F`[`Enter`] the choice switches.

`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 (free or fixed field).



```
 KUNST RATERS3: Interrater agreement:                    Data list: Columns
                                                     Tot. lines per row: 1
    Name   Type Line  Columns  Missing       Name   Type Line  Columns  Missing

  1              1 ... - ...
  2              1 ... - ...
  3              1 ... - ...
  4              1 ... - ...
  5              1 ... - ...    You can use the following options:
  6              1 ... - ...
  7              1 ... - ...    C to define the positions of raters
                                T to define the total number of lines per row
                                N to define names for the raters
                                L to define line numbers within a row
                                D to delete raters from the list
                                M to define missing values

                                After this character you may add further
                                specifications or press [Enter] for more help

                                Press [Enter] to remove this help window
C=Columns T=Total lines N=Name L=Line numbers D=Delete M=Missing
Give C, T, N, L, D, M or continue with column definitions:
```

Figure 4:   The data list for fixed format data.

Figure 4 shows a data list for 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. If the list is too large to fit on one screen the list of characters will also contain ‘B’ and ‘F’ to scroll **b**ackward of **f**orward.

The possibilities can best be clarified by some examples:

You type: `c1 5-7Enter`
Thereby you specify that the first **c**olumn (rater) is in positions 5 through 7.
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: `2-3 11-14Enter`
This means that raters 2 and 3 occupy positions 11 through 14, so rater 2 is in position 11-12 and rater 3 in positions 13-14.

You type: `4 16Enter`
This means that rater 4 occupies position 16.

You type: `L5 2Enter`
Now you have entered a different code. The code **L** indicates that you are defining line numbers. You specify that rater 5 is contained in the second line of a case. Line numbers must be in ascending order. Therefore, the program will adjust the line numbers of raters 5 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: `N1 Freud`Enter
The code **`N`** indicates that you are specifying names. Rater 1 will receive the name '`Freud`'. Names will be truncated to 8 characters.

You type: `2-3 Piaget`Enter
Raters 2 through 3 will both be called '`Piaget`'.

You type: `4-6 Rater4`Enter
Raters 4 through 6 will be called '`Rater4`', '`Rater5 and` '`Rater6`'. As you see, if the name ends on a number subsequent names will have their numbers incremented.

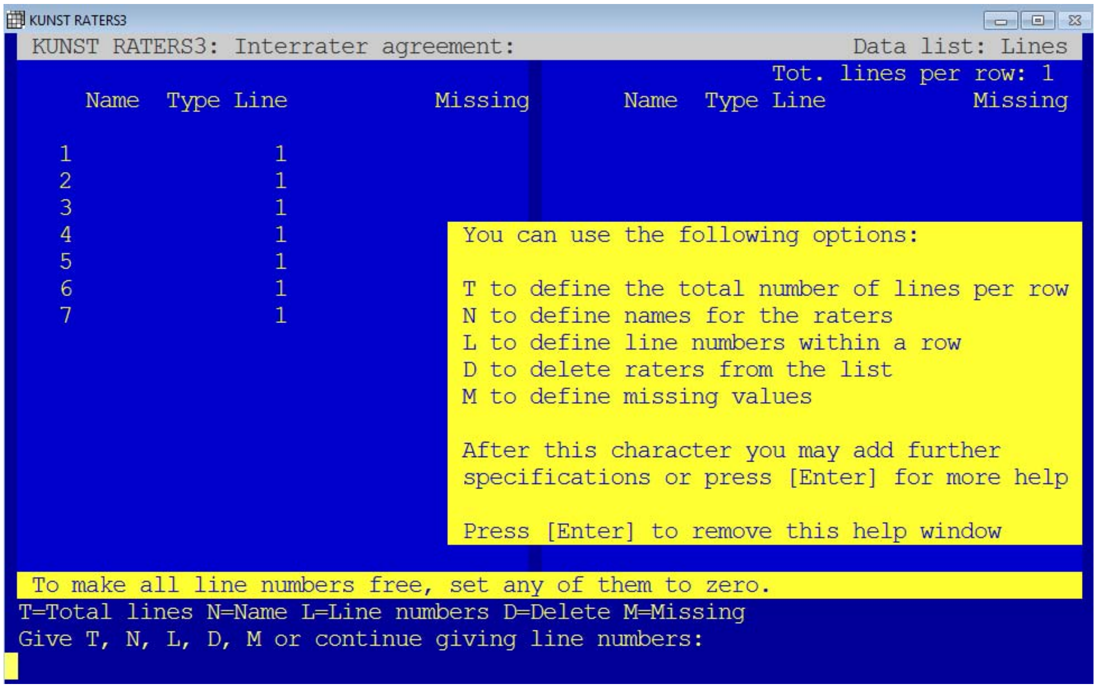
You type: `M1-5 99`Enter
The code **`M`** is used to define missing values. If for any rater a value is found that is equal to its missing value **or greater**, the program will treat it as a missing value. In this example, you define the value 99 as a missing value for raters 1 through 5. These upper limits must be positive numbers.

You type: `D4`Enter
The option **`D`** allows you to remove a rater from the list. All sequence numbers and all other definitions will be adjusted accordingly. In this example, you remove the fourth rater. You can also specify a range of rater numbers, for instance '`D4-5`' to remove the raters 4 and 5.

You type: `T4`Enter
If you don't use the option **`T`**, the program will assume that for each case the number of lines is equal to the line number of the last rater in the data list. If there are more lines in a case, you must specify so by the option `T`. In this example, you specify that there are 4 lines for each case.

You type: FEnter or BEnter

If there are more than 32 raters, they will not fit at once on the data list screen. Therefore, you have the possibility to scroll forward and backward with the options F and B.



```
 KUNST RATERS3: Interrater agreement:                        Data list: Lines
                                                          Tot. lines per row: 1
     Name  Type Line            Missing        Name  Type Line            Missing

   1               1
   2               1
   3               1
   4               1                  You can use the following options:
   5               1
   6               1                  T to define the total number of lines per row
   7               1                  N to define names for the raters
                                      L to define line numbers within a row
                                      D to delete raters from the list
                                      M to define missing values

                                      After this character you may add further
                                      specifications or press [Enter] for more help

                                      Press [Enter] to remove this help window

 To make all line numbers free, set any of them to zero.
T=Total lines N=Name L=Line numbers D=Delete M=Missing
Give T, N, L, D, M or continue giving line numbers:
```

Figure 5: The data list for free format data.

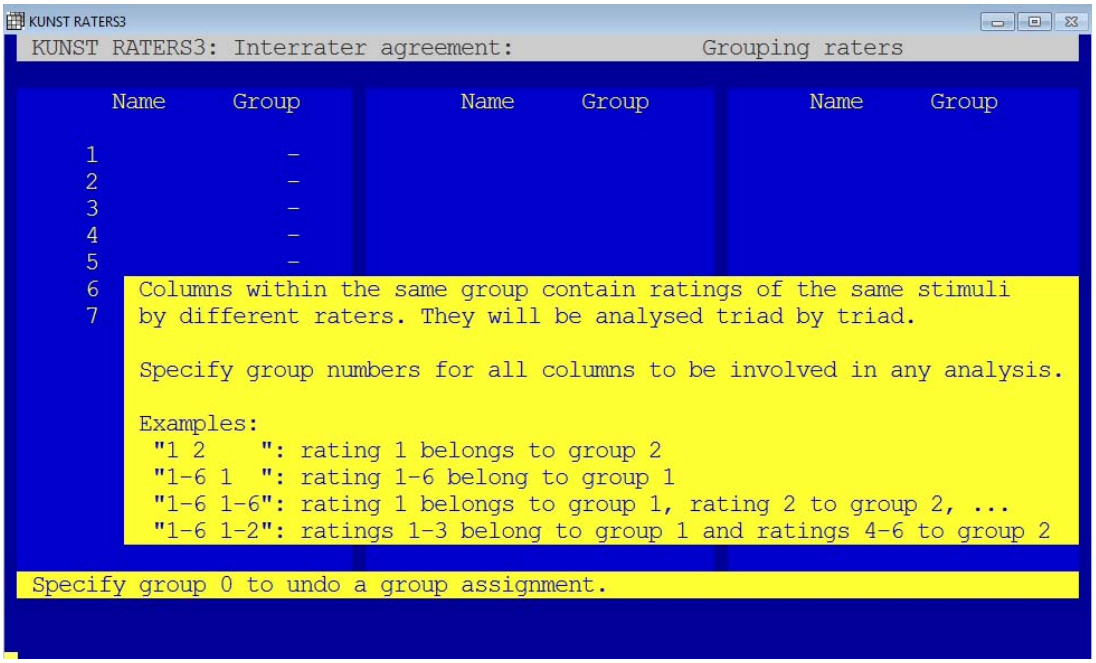
If you use free format data (see figure 5), you do not need to specify anything in the data list. However, you may specify on which line of a row each rater is recorded. So you can use the option L. The options T, N, D and M are also available. Their meaning and use are the same as with fixed format data (see above).

If you have finished the data list, you can go back to the data 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 menu.

G Group ratings

If the input to the program consists of raw data, the program will build three-dimensional frequency distributions for combinations of columns (raters). Therefore, you may classify the raters in groups. Ratings3 will compute a frequency table for each triad of raters **within** a group, but not for triads of raters from different groups. The way groups are defined also can influence the definition of the categories that are to be used in the frequency tables (see option C in the main menu).

The partitioning can be defined by choosing the option **G** from the data menu. When you type GEnter the grouping window will appear. Figure 6 gives an example.



```
KUNST RATERS3: Interrater agreement:            Grouping raters

      Name      Group         Name      Group         Name      Group

    1              -
    2              -
    3              -
    4              -
    5              -
    6   Columns within the same group contain ratings of the same stimuli
    7   by different raters. They will be analysed triad by triad.

        Specify group numbers for all columns to be involved in any analysis.

        Examples:
         "1 2    ": rating 1 belongs to group 2
         "1-6 1  ": rating 1-6 belong to group 1
         "1-6 1-6": rating 1 belongs to group 1, rating 2 to group 2, ...
         "1-6 1-2": ratings 1-3 belong to group 1 and ratings 4-6 to group 2

 Specify group 0 to undo a group assignment.
```

Figure 6: The grouping window.

Now you can assign group numbers to the raters. Raters with the same number will form a group.
To assign a group number to a single rater, type the sequence number of the rater and the group number, for instance '`1 1`Enter' to specify that rater 1 belongs to group 1.
To assign a group number to a series of adjacent raters, type the range of sequence numbers of the raters and the group number, for instance '`1-6 1`Enter' to specify that raters 1 through 6 all belong to group 1.
To assign subsequent group numbers to adjacent raters specify the range of sequence numbers and the range of group numbers. If both ranges contain the same number of values, the groups are assigned one by one. Specify for instance '`1-3 1-3`Enter' to indicate that rater one belongs to group 1, rater 2 to group 2 and so on. If the range of group numbers is smaller than that of the sequence numbers, the group numbers are distributed over the raters. If, for instance, you type '`1-6 1-3`Enter' raters 1 and 2 will belong to group 1, raters 3 and 4 to group 2 and raters 5 and 6 to group 3.

If you have finished the group definitions, you can go back to the data 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 menu.

8 The bootstrapping menu

For the computation of confidence intervals around the parameter estimates, one may make use of the estimated standard errors from what is called the observed information matrix. But if any of the parameters is a boundary extreme, the information matrix cannot be computed. Moreover, the estimation of the confidence intervals by the standard errors ad the model fit based on the likelihood ratio test are only correct if the assumption of a normal distribution is justified.

The way out of these problems is a technique called bootstrapping. From the parameter estimates the joint population distribution X can be estimated, where X is a c by c by c matrix, such that X_{ijk} is the probability that a case is classified in category i by the first rater, in category j by the second and in category k by the third. In the bootstrapping procedure a large number of samples of size n are drawn from this estimate $\hat{X}$. For each of these tables the estimation procedure is performed, resulting in an observed distribution over the samples for each of the parameter estimates. These distributions then can be used to estimate confidence intervals.

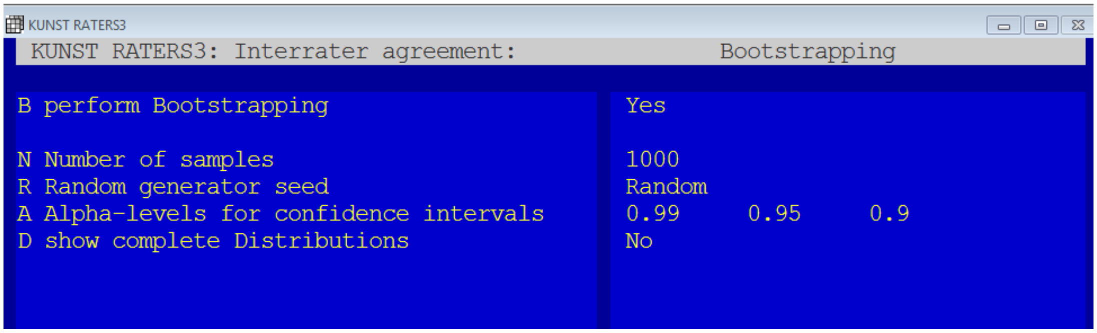
Raters3 will produce two types of intervals:

- Symmetrical intervals: intervals chosen such that the parameter estimate is (as much as possible) in the center of the interval.
- Shortest intervals: intervals chosen such that the difference between the two bounds is as small as possible. If there are several equally short intervals, the program will select the most symmetrical of them.

If bootstrapping is performed, the program will also compute the model fit criterion for the generated samples and estimate the corresponding probability. The correctness of this probability does not depend on the form of the distribution of the criterion.

Raters3 will also compute the proportions of samples for each order of the estimated parameters p_1, p_2 and p_3.

If you choose the option '`B Bootstrapping ...`' from the main menu, the bootstrapping menu appears (see figure 7).



```
KUNST RATERS3: Interrater agreement:              Bootstrapping

B perform Bootstrapping                     Yes

N Number of samples                         1000
R Random generator seed                     Random
A Alpha-levels for confidence intervals     0.99      0.95      0.9
D show complete Distributions               No
```

Figure 7: The bootstrapping menu.

This menu contains the following options:

`B performing Bootstrapping`
By this option you determine whether bootstrapping must be performed or not. It must be noted that the bootstrapping procedure may be very time consuming.
Each time you type `B`Enter the choice switches from `Yes` to `No` or back.

`N Number of samples`
This option is used to specify the number of samples to be drawn. Just type '`N ##`Enter', where ## is the number of samples to be drawn. The more samples the better the parameter distributions can be estimated, but also the more time it will take!

`R Random generator seed`
The sampling procedure uses a random number generator. By typing '`R ##Enter`', you may specify any integer starting value. If the starting value is positive, the generator will always start with that number. That means that you will get exactly the same results each time you perform the same series of analyses. If you specify zero or a negative number, the starting point of the generator will itself be random! As a consequence results of repeated analyses may be slightly different.

`A Alpha levels for confidence intervals`
This option allows you to specify up to three alpha-levels for confidence intervals. If you specify '`A ##1 ##2 ##3 Enter`" where ##1, ##2 and ##3 are values between 0 and 1, the program will produce the corresponding 3 intervals for each parameter. If you specify only two or one values only two or one intervals will be computed.
Example: If you type '`A 0.99 0.98 0.95Enter`', 99%, 98% and 95% intervals will be computed.

`D show complete Distributions`
With this option you can ask to show in the listing file not only the confidence intervals, but also the complete distributions of the parameter estimates. Each time you type `DEnter` the choice switches from `Yes` to `No` or back.

If you are ready with the bootstrapping menu, you can go back to the main 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 main menu.

9 The input data

Raters3 accepts two types of input: raw data and two-dimensional frequency tables. In one program run however all data must be of the same type. In each run, there may be several input files.

9.1 Raw data

A raw data file consists of a number of rows, corresponding to individual cases (a case is a stimulus to be classified) and a number of columns, corresponding to the raters. Each case will occupy one or more lines in the file. If a free format is chosen, numbers must be separated by spaces or tabs. If a fixed format is chosen all cases must have exactly the same format, so for each score the line number and the position within that line is completely fixed. The fixed format is very rigid and more sensitive to errors in the data. Therefore, one should use it only if there are no blanks or tabs between the values.

9.2 Frequency tables

The input may consist of three-way frequency tables. If so, they can be coded in two ways:

- The information of a table starts with a line that contains only one number, specifying the number of categories (`c`) in the table. After that first line follow `c` sub-tables, each containing `c` lines with `c` frequencies per line. The first sub-table refers to the first category for rater 3, the second to the second category for rater 3 and so on.

```
3

37  16  19
19  11   7
 5   7   2

32  21  13
30 103  38
10  22  11

 0   2   7
 9  11  16
11  13  28
```

```
Example 2:
3
37  16  19
19  11   7
 5   7   2

32  21  13
30 103  38
10  22  11

 0   2   7
 9  11  16
11  13  28
```

```
Example 3:
Rater 3 = 1
37  16  19
19  11   7
 5   7   2
Rater 3 = 2
32  21  13
30 103  38
10  22  11
Rater 3 = 3
 0   2   7
 9  11  16
11  13  28
```

Figure 8: Three equivalent examples of a frequency table as input.

- The information of a table starts immediately with a line containing the first row of the first subtable. The program will count the values in this first line and take it as the number of categories (c).

Figure 8 gives three equivalent examples for the input of the data as shown in table 1. Both types of definitions may be mixed, some tables being specified in the first and others in the second way. In both forms, text lines may be inserted before and between the submatrices and between the rows of the matrices. Any line that contains non-numerical characters (that is other than ‘+’, ‘-‘, '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '.' or a tab-character) or is completely empty, will be interpreted as a comment. If it is not empty, its content will be shown in the listing file. Note that a line like ‘-----‘ or ‘.....’ will not be recognized as a comment line and will cause an i/o-error.

10 Results in the listing file

A listing file contains the results of the analyses of the data in the corresponding data file. Its exact content depends on the specifications.

- The file starts with a short definition of the parameters and an overview of the options as they are chosen by the user.
- If the input file contains raw data, the first input rows may be echoed, dependent on the option *Echo* in the main menu. If the input file contains frequency tables any comment inserted before or between the input rows will be shown.
- Then follow the results for the frequency matrices, one by one.
- The frequency matrix as it is read from the input file or constructed from the raw input data will be shown (see figure 9).

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

   3.  Frequency matrix 1:

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

 ....................................
 Rater 3 has value 1:
 ....................................

               Rater 2
 Rater 1

           1        2        3
   1     37.0000  16.0000  19.0000
   2     19.0000  11.0000   7.0000
   3      5.0000   7.0000   2.0000


 ...................................
 Rater 3 has value 2:
 ...................................

              Rater 2
 Rater 1

         1        2        3

   1     32.0000  21.0000  13.0000
   2     30.0000 103.0000  38.0000
   3     10.0000  22.0000  11.0000


 ...................................
 Rater 3 has value 3:
 ...................................

              Rater 2
 Rater 1

         1        2        3

   1      0.0000   2.0000   7.0000
   2      9.0000  11.0000  16.0000
   3     11.0000  13.0000  28.0000

Estimated standard errors:
      se(p1) =      0.0495
      se(p2) =      0.0447
      se(p3) =      0.0555

        se(V)         se(W1)       se(W2)        se(W3)
       --------  ------------  ------------   -----------
        0.0364        0.0435        0.0330        0.0518
        0.0372        0.0346        0.0303        0.0649
        0.0490        0.0514        0.0457        0.0789
```

Figure 9: An input table reported in the listing file.

- Then follow the estimates of the parameters, as shown in figure 10: first the estimated p- and s-values and, as a bonus the values of Cohen's kappa for the three subtables, followed by estimated proportions of correct classifications and the estimated V- and W-vectors.
- Next come the estimated standard errors as they are computed from the information matrix. If any of the parameter estimates are boundary values, these standard errors cannot be computed. Moreover, there may be numerical problems that prevent this computation.

```
Minimization by DFP took 54 iterations.



 p1 (Rater 1 ) =  0.4754
 p2 (Rater 2 ) =  0.3524
 p3 (Rater 3 ) =  0.6692

 s12 = p1.p2 =  0.1676
 s13 = p1.p3 =  0.3181
 s23 = p2.p3 =  0.2358
 Kappa(1,2) =  0.1815
 Kappa(1,3) =  0.3302
 Kappa(2,3) =  0.2429

 p1+ = proportion of correct classifications by Rater 1  : 0.6559
 p2+ = proportion of correct classifications by Rater 2  : 0.5694
 p3+ = proportion of correct classifications by Rater 3  : 0.7866

       V    W1(Rater 1)  W2(Rater 2)   W3(Rater 3)
     ------ -----------  -----------   -----------
     0.3805     0.2032      0.2666      0.0000
     0.3580     0.6057      0.4333      0.9698
     0.2615     0.1911      0.3001      0.0302


Estimated standard errors:
      se(p1) =       0.0495
      se(p2) =       0.0447
      se(p3) =       0.0555

        se(V)          se(W1)       se(W2)        se(W3)
       --------  ------------  ------------   -----------
        0.0364        0.0435        0.0330        0.0518
        0.0372        0.0346        0.0303        0.0649
        0.0490        0.0514        0.0457        0.0789
```

Figure 10: The estimated parameters reported in the listing file.

- The listing file will also show the frequency matrix as reconstructed from the parameter estimates. Figure 11 gives an example. The similarity of this matrix to the observed frequency matrix gives an impression of the model fit and allows the user to trace possible model violations.

```
X.n = Frequencies based on parameter estimates:
 _______________________________________________



 ....................................
 Rater 3 has value 1:
 ....................................

              Rater 2
 Rater 1

          1        2        3

   1     38.9054  20.7907  14.3990
   2     21.2424  11.3517   7.8619
   3      6.7002   3.5805   2.4798


 ....................................
 Rater 3 has value 2:
 ....................................

              Rater 2
 Rater 1

          1        2        3

   1     22.6858  23.1804  13.0192
   2     36.7506  98.1638  38.3734
   3     10.4471  19.7363  17.8435


 ....................................
```

```
Rater 3 has value 3:
...................................

              Rater 2
Rater 1

          1        2        3

  1      2.2479   3.0870   5.4274
  2      5.4333   8.9854  15.8219
  3      8.9550  14.5097  28.0207
```

Figure 11: The frequency matrix based on the parameter estimates.

- The model fit is evaluated more precisely by a χ^2-test based on the likelihood ratio. Small values for the corresponding probability are indications that the model assumptions are violated. Figure 12 gives an example.

```
Model test based on likelihood ratio:
 -------------------------------------

 chi square         =          22.9018
 degrees of freedom =          16.
 probability        =           0.11641
```

Figure 12: Model test.

- After the results of the model test, the listing file contains for each rater a table, where the rows refer to the true categories of the cases and the columns to the categories chosen by the rater. Cell (i,j) contains the proportion of cases, that belong to category i and for which the rater chooses category j. The table is completely based on the estimated parameters. Figure 13 gives an example.

```
 True categories and predicted score distribution of rater 1 (Rater 1)


 True        Predicted
 categories   1      2      3     | Total
 ---------------------------------|-------
          1 0.2215 0.1209 0.0381 | 0.3805
          2 0.0382 0.2840 0.0359 | 0.3580
          3 0.0279 0.0831 0.1505 | 0.2615
 ---------------------------------|-------
      Total: 0.2875 0.4880 0.2245 | 1.0000
   Observed: 0.2875 0.4880 0.2245 | 1.0000


 True categories and predicted score distribution of rater 2 (Rater 2)


 True        Predicted
 categories   1      2      3     | Total
 ---------------------------------|-------
          1 0.1998 0.1068 0.0739 | 0.3805
          2 0.0618 0.2266 0.0696 | 0.3580
          3 0.0451 0.0734 0.1430 | 0.2615
 ---------------------------------|-------
      Total: 0.3067 0.4068 0.2865 | 1.0000
   Observed: 0.3067 0.4068 0.2865 | 1.0000


 True categories and predicted score distribution of rater 3 (Rater 3)


 True        Predicted
```

```
categories   1      2      3     | Total
---------------------------------|-------
         1 0.2546 0.1221 0.0038 | 0.3805
         2 0.0000 0.3544 0.0036 | 0.3580
         3 0.0000 0.0839 0.1776 | 0.2615
---------------------------------|-------
     Total: 0.2546 0.5604 0.1850 | 1.0000
  Observed: 0.3067 0.4068 0.2865 | 1.0000
```

Figure 13: Distribution of true scores and ratings for a single rater.

- The listing file also contains a summarizing table of the basic probabilities as they are estimated from the parameters. In this table the ratings of the three raters are divided into three categories: correct observations, correct guesses and wrong guesses. The table shows the probabilities for the 27 combinations of these categories. Figure 14 gives an example.

```
Overview of probabilities for the three raters:
 _______________________________________________


 good  = correct observation
 lucky = correct guess
 wrong = wrong guess


 ....................................
 Rater 3 makes a correct observation:
 ....................................
               Rater 2
 Rater 1
        |  good    lucky   wrong   |   total
 ---------------------------------------------
 good :  |  0.1121  0.0690  0.1370  |   0.3181
 lucky:  |  0.0426  0.0267  0.0488  |   0.1208
 wrong:  |  0.0811  0.0488  0.1024  |   0.2302
 ---------------------------------------------
 total:  |  0.2358  0.1452  0.2881  |   0.6692

 ....................................
 Rater 3 makes a correct guess:
 ....................................
               Rater 2
 Rater 1
        |  good    lucky   wrong   |   total
 ---------------------------------------------
 good :  |  0.0197  0.0121  0.0240  |   0.0558
 lucky:  |  0.0075  0.0047  0.0086  |   0.0212
 wrong:  |  0.0142  0.0086  0.0180  |   0.0404
 ---------------------------------------------
 total:  |  0.0414  0.0255  0.0506  |   0.1175

 ....................................
 Rater 3 makes a wrong guess:
 ....................................
               Rater 2
 Rater 1
        |  good    lucky   wrong   |   total
 ---------------------------------------------
 good :  |  0.0357  0.0220  0.0437  |   0.1014
 lucky:  |  0.0136  0.0085  0.0156  |   0.0385
 wrong:  |  0.0259  0.0156  0.0326  |   0.0734
 ---------------------------------------------
 total:  |  0.0752  0.0463  0.0919  |   0.2134
```

Figure 14: Overview of probabilities for the three raters.

- After this three-way overview the listing contains three summarizing tables for the three pairs of raters, as shown in figure 15.

```
Overview of probabilities for raters 1 and 2:
 _____________________________________________


 good  = correct observation
 lucky = correct guess
 wrong = wrong guess

            rater 2
 rater 1
        |  good    lucky   wrong   |   total
 ---------------------------------------------
 good :  |  0.1676  0.1031  0.2047  |   0.4754
 lucky:  |  0.0636  0.0399  0.0729  |   0.1805
 wrong:  |  0.1213  0.0698  0.1530  |   0.3441
 ---------------------------------------------
 total:  |  0.3524  0.2170  0.4306  |   1.0000




 Overview of probabilities for raters 1 and 3:
 _____________________________________________


 good  = correct observation
 lucky = correct guess
 wrong = wrong guess

            rater 3
 rater 1
        |  good    lucky   wrong   |   total
 ---------------------------------------------
 good :  |  0.3181  0.0558  0.1014  |   0.4754
 lucky:  |  0.1208  0.0333  0.0230  |   0.1805
 wrong:  |  0.2302  0.0249  0.0890  |   0.3441
 ---------------------------------------------
 total:  |  0.6692  0.1175  0.2134  |   1.0000




 Overview of probabilities for raters 2 and 3:
 _____________________________________________


 good  = correct observation
 lucky = correct guess
 wrong = wrong guess

            rater 3
 rater 2
        |  good    lucky   wrong   |   total
 ---------------------------------------------
 good :  |  0.2358  0.0414  0.0752  |   0.3524
 lucky:  |  0.1452  0.0178  0.0390  |   0.2170
 wrong:  |  0.2881  0.0433  0.0991  |   0.4306
 ---------------------------------------------
 total:  |  0.6692  0.1175  0.2134  |   1.0000
```

Figure 15: Overview of probabilities for the three pairs of raters.

- If the bootstrapping option is chosen, the listing will be concluded by the results of that procedure. It should be remembered that these results are based on randomization. So, if you repeat an analysis, the results from this part will vary from run to run, unless you have chosen a fixed random generator seed (see section 8).

This part of the listing file starts with the symmetric confidence intervals, as illustrated by figure 16. The central column of this table contains the parameter estimates. To the left and to the right of this central column, the lower and upper bounds are given for the confidence intervals, with the smallest ranges nearest to the central column. The last column contains the standard errors as they are estimated by the bootstrapping procedure. If the bootstrapping is based on a sufficiently large number of samples (say 1000), one may have more trust in these standard errors than in those computed from the observed information matrix.

```
Symmetric confidence intervals and standard errors:

 --------------------------------------------------------------------------------
         99.00%   95.00%   90.00%   Estimated 90.00%  95.00%   99.00%   Standard
                                    value                               error
 --------------------------------------------------------------------------------
 p1:     0.334500 0.370500 0.387500 0.475407 0.563500 0.580500 0.616500 0.052304
 p2:     0.227500 0.261667 0.277500 0.352445 0.427500 0.442667 0.477500 0.045623
 p3:     0.339500 0.543000 0.569167 0.669173 0.771167 0.796000 0.990500 0.073204
 p1p2:   0.085500 0.107500 0.118500 0.167555 0.209500 0.227500 0.249500 0.028536
 p1p3:   0.229500 0.253500 0.263269 0.318129 0.374269 0.383500 0.407500 0.035585
 p2p3:   0.148500 0.170000 0.183500 0.235847 0.289500 0.301000 0.322500 0.033934
 p1+:    0.566500 0.589167 0.598250 0.655938 0.711250 0.721167 0.744500 0.034566
 p2+:    0.484500 0.509500 0.519500 0.569401 0.617500 0.628500 0.654500 0.030517
 p3+:    0.573500 0.687500 0.705833 0.786655 0.866833 0.885500 0.990500 0.054131
 V( 1):  0.256500 0.295900 0.314700 0.380505 0.447700 0.466900 0.504500 0.044078
 V( 2):  0.169500 0.254833 0.273167 0.358003 0.443167 0.462833 0.547500 0.057600
 V( 3):  0.167500 0.189500 0.202786 0.261492 0.319786 0.333500 0.355500 0.036886
 W1( 1): 0.077500 0.109500 0.126300 0.203188 0.281300 0.297500 0.329500 0.048041
 W1( 2): 0.445500 0.485500 0.512250 0.605748 0.699250 0.725500 0.765500 0.058688
 W1( 3): 0.088500 0.114000 0.131500 0.191064 0.252500 0.269000 0.294500 0.038446
 W2( 1): 0.169500 0.197500 0.207500 0.266585 0.325500 0.336500 0.363500 0.036508
 W2( 2): 0.331500 0.359500 0.372250 0.433314 0.496250 0.508500 0.535500 0.039903
 W2( 3): 0.216500 0.235500 0.245500 0.300101 0.353500 0.366500 0.384500 0.032731
 W3( 1): 0.000000 0.000000 0.000000 0.000000 0.120500 0.157500 0.361500 0.069638
 W3( 2): 0.424500 0.778500 0.832500 0.969793 1.000000 1.000000 1.000000 0.099010
 W3( 3): 0.000000 0.000000 0.000000 0.030207 0.111500 0.141500 0.208500 0.051557
```

Figure 16: Symmetric confidence intervals and standard errors from bootstrapping.

In the example of figure 16 the 90% confidence interval for s_{12} is (0.118500, 0.209500), the 95% interval is (0.107500, 0.227500) and the 99% interval is (0.085500, 0.249500). These intervals are symmetric as far as possible: if a boundary exceeds the theoretical minimum or maximum, it is set to that extreme.

- After the symmetric confidence intervals, also shortest intervals are computed. If there are more than one possible shortest intervals of equal length, the one is chosen in which the parameter estimate is most central. Figure 17 gives an example.

```
 Shortest confidence intervals and standard errors:

 --------------------------------------------------------------------------------
         99.00%   95.00%   90.00%   Estimated 90.00%  95.00%   99.00%   Standard
                                    value                               error
 --------------------------------------------------------------------------------
 p1:     0.334500 0.395500 0.409516 0.475407 0.574488 0.599500 0.604500 0.052304
```

```
p2:      0.239500 0.276500 0.298515 0.352445 0.443485 0.450500 0.483500 0.045623
p3:      0.668500 0.531500 0.550503 0.669173 0.745493 0.767500 0.999500 0.073204
p1p2:    0.089500 0.125500 0.132500 0.167555 0.218500 0.232500 0.251500 0.028536
p1p3:    0.228500 0.254500 0.255507 0.318129 0.369496 0.383500 0.396500 0.035585
p2p3:    0.150500 0.181500 0.181504 0.235847 0.287495 0.307500 0.322500 0.033934
p1+:     0.565500 0.594503 0.606518 0.655938 0.716488 0.725494 0.739500 0.034566
p2+:     0.484500 0.522500 0.532531 0.569401 0.628469 0.636500 0.651500 0.030517
p3+:     0.785500 0.679500 0.696514 0.786655 0.851486 0.869500 0.999500 0.054131
V( 1):   0.238500 0.288500 0.305500 0.380505 0.437500 0.454500 0.475500 0.044078
V( 2):   0.237500 0.257500 0.290500 0.358003 0.457500 0.463500 0.558500 0.057600
V( 3):   0.167500 0.189500 0.195506 0.261492 0.312487 0.333500 0.352500 0.036886
W1( 1): 0.079500 0.120500 0.132518 0.203188 0.285486 0.305500 0.329500 0.048041
W1( 2): 0.442500 0.485500 0.511500 0.605748 0.698500 0.722500 0.757500 0.058688
W1( 3): 0.095500 0.106500 0.128500 0.191064 0.250500 0.256500 0.297500 0.038446
W2( 1): 0.181500 0.196500 0.212500 0.266585 0.329500 0.334500 0.364500 0.036508
W2( 2): 0.330500 0.358500 0.362505 0.433314 0.485491 0.506500 0.527500 0.039903
W2( 3): 0.222500 0.234555 0.249500 0.300101 0.357500 0.365473 0.384500 0.032731
W3( 1): 0.000500 0.000500 0.000500 0.000000 0.120500 0.157500 0.361500 0.069638
W3( 2): 0.968500 0.968500 0.968500 0.969793 0.999500 0.999500 0.999500 0.099010
W3( 3): 0.000500 0.000500 0.000500 0.030207 0.111500 0.143484 0.208500 0.051557
```

Figure 17: Shortest confidence intervals from the bootstrapping procedure.

- After the confidence intervals the model test, based on the bootstrapping procedure is reported: the probability is the proportion of samples with a χ^2-value larger than or equal to the value found for the original data. Figure 18 shows the output.

```
Model test:
 -------------------------------
 chi square  =      22.9018
 probability =       0.1480

 Probability(p1 >= p2 >= p3) = 0.0000
 Probability(p1 >= p3 >= p2) = 0.0240
 Probability(p2 >= p1 >= p3) = 0.0010
 Probability(p2 >= p3 >= p1) = 0.0000
 Probability(p3 >= p1 >= p2) = 0.9390
 Probability(p3 >= p2 >= p1) = 0.0360

 Probability(p1+ >= p2+ >= p3+) = 0.0000
 Probability(p1+ >= p3+ >= p2+) = 0.0200
 Probability(p2+ >= p1+ >= p3+) = 0.0000
 Probability(p2+ >= p3+ >= p1+) = 0.0000
 Probability(p3+ >= p1+ >= p2+) = 0.9530
 Probability(p3+ >= p2+ >= p1+) = 0.0270
```

Figure 18: Model test based on bootstrapping.

- The program reports also the estimated probabilities of the different orders of the parameter estimates $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_3$. These estimates are the corresponding proportions of samples in the bootstrapping procedure.
- Optionally the listing file contains the complete distribution of the parameters and the statistics p_1^+ and p_2^+ from the bootstrapping procedures. For each parameter and statistic, the range of possible values is divided in 1000 categories and for each category the number of occurrences is computed. Note that this table will take a lot of space. Figure 18 shows a small and shortened part of such a report.

```
 Distributions based on bootstrapping:
 -------------------------------------



 Value         p1    p2    p3    p1p2  p1p3  p2p3  p1+   p2+   p3+   V(1)
 -------------------------------------------------------------------------
  Distributions based on bootstrapping:

 -------------------------------------

(1000 samples)




 Value         p1    p2    p3    p1p2  p1p3  p2p3  p1+   p2+   p3+   V(1)
 -------------------------------------------------------------------------
 0.000500 :     0     0     0     0     0     0     0     0     0     1
 .....
 .....
 0.102500 :     0     0     0     1     0     0     0     0     0     0
 0.103500 :     0     0     0     1     0     0     0     0     0     0
 0.104500 :     0     0     0     0     0     0     0     0     0     0
 0.105500 :     0     0     0     0     0     0     0     0     0     0
 0.106500 :     0     0     0     2     0     0     0     0     0     0
 0.107500 :     0     0     0     2     0     0     0     0     0     0
 0.108500 :     0     0     0     3     0     0     0     0     0     0
 0.109500 :     0     0     0     1     0     0     0     0     0     0
 .....
 .....
 0.165500 :     0     0     0    14     0     2     0     0     0     0
 0.166500 :     0     0     0    16     0     0     0     0     0     0
 0.167500 :     0     0     0    25     0     2     0     0     0     0
 0.168500 :     0     0     0    13     0     1     0     0     0     0
 0.169500 :     0     0     0    22     0     1     0     0     0     0
 0.170500 :     0     0     0    13     0     0     0     0     0     0
 0.171500 :     0     0     0    10     0     0     0     0     0     0
 0.172500 :     0     0     0     9     0     1     0     0     0     0
 0.173500 :     0     0     0    16     0     2     0     0     0     0
 0.174500 :     0     0     0    19     0     1     0     0     0     0
 0.175500 :     0     0     0    10     0     2     0     0     0     0
 0.176500 :     0     0     0    14     0     1     0     0     0     0
 0.177500 :     0     0     0    14     0     0     0     0     0     0
 0.178500 :     0     0     0     9     0     0     0     0     0     0
 0.179500 :     0     0     0    19     0     2     0     0     0     0
 0.180500 :     0     0     0    13     0     1     0     0     0     0
 0.181500 :     0     1     0    23     0     3     0     0     0     0
 ....
```

Figure 19: Part of the report of the distributions of the parameters.

11 The raw output file

If the input to Raters3 consists of raw data, the user has the possibility to save the computed frequency matrices in a separate file, that borrows its name from the input file, possibly followed by a sequence number, but ending on `'.out'`, or '`1.out`', '`2.out`' and so on. The matrices are stored without any layout except a single empty line before their first row. Each row is written as one line and each number will use 6 positions. Figure 20 contains an example of the content of such a file.

```
    37    16    19
    19    11     7
     5     7     2
    32    21    13
    30   103    38
    10    22    11
     0     2     7
     9    11    16
     11     13     28
```

Figure 20: Content of a raw output file.

12 References

Bendermacher, Nol and Souren, Pierre. (2009), Beyond kappa: Estimating Inter-Rater Agreement with Nominal Classifications, *Journal of Modern Applied Statistical Methods, Vol. 8, No 1, 110-121*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Raters2.doc at: http://www.ru.nl/socialewetenschappen/rtog/software/statistische/kunst/raters2/