R with Rcmdr: BASIC INSTRUCTIONS

Contents

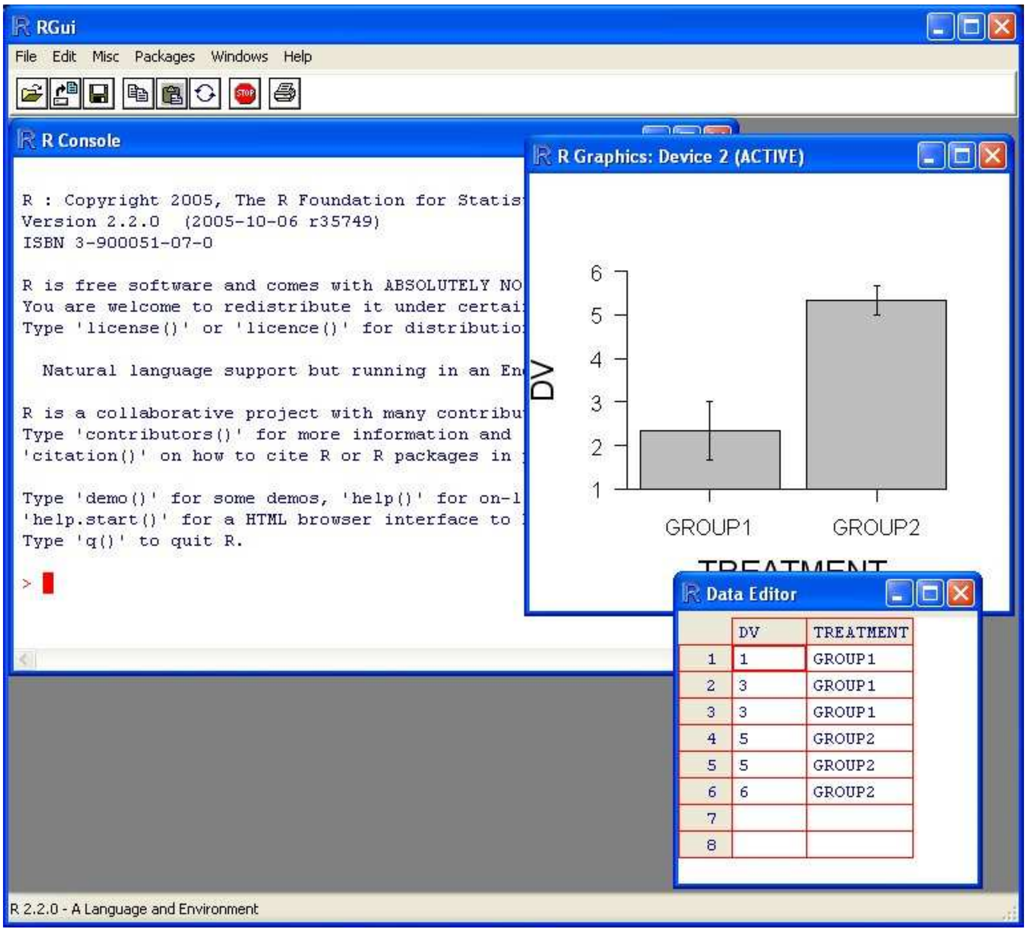


Figure 1: RGui - R 2.4.1 for windows. When running Rcmdr, the **R Console** window is rarely examined. All graphs produced by Rcmdr will appear in a **R Graphics** window within **RGui**. The **Data Editor** window is a spreadsheet called from **Rcmdr** that can be used to create and modify data sets. Note, that both the **R Graphics** and **Data Editor** windows are not initially present - they only appear as required.

By Murray Logan

1 RUNNING & INSTALLATION R UNDER WINDOWS

1.1 Running R and Rcmdr from CD

1.1.1 To load up R

1. Goto the directory `rw2000/bin/`

2. Run the executable file `Rgui.exe`
   This will start R. Note that R itself is a command driven program, the menus are provided by an add-in package called `Rcmdr` (see section 2).

1.1.2 To load up Rcmdr

1. Select the **Packages** menu (from the **Rgui** window)
2. Select the **Load packages..** submenu
   The *Select one* window will appear from which you need to select **Rcmdr** and click the **OK** button
   This will load up the `Rcmdr` package and a new window will appear (see figure 2)

1.2 Installing from CD

To install R and all the packages used on the CD (including `Rcmdr`) onto your own computer:

1. Run the file called **install.bat** that is in the top(root) directory of the CD
2. Follow the prompts and allow it to install in the default position
3. Once it has installed, a menu and desktop icon will be included
4. The **install.bat** will then automatically install all the packages into their correct locations
5. R and `Rcmdr` can then be run locally (without the CD) by the same instructions as in sections 1.1.1 and 1.1.2 respectively.

1.3 Downloading from R web page

R/Rcmdr can also be downloaded from Murray's web page

- `http://users.monash.edu.au/downloads`.

This location also contains the Eworksheets as well as other resources. Occasionally, if a bug is identified in R/Rcmdr or the Eworksheeets, corrected versions may be posted on this site.

R and Rcmdr as well as other packages used in this course can also be downloaded directly from the Comprehensive R Archive Network (CRAN). Windows versions can be downloaded from:

- `http://cran.r-project.org/bin/windows`.

Whilst R, `Rcmdr` and all of the other packages required can be downloaded from the above site, some menus and dialog boxes of the official `Rcmdr` package have been added and/or modified by Murray Logan to better suit BIO3011 students. As a result, some of the procedures documented in this manual are not available with the standard `Rcmdr` download.

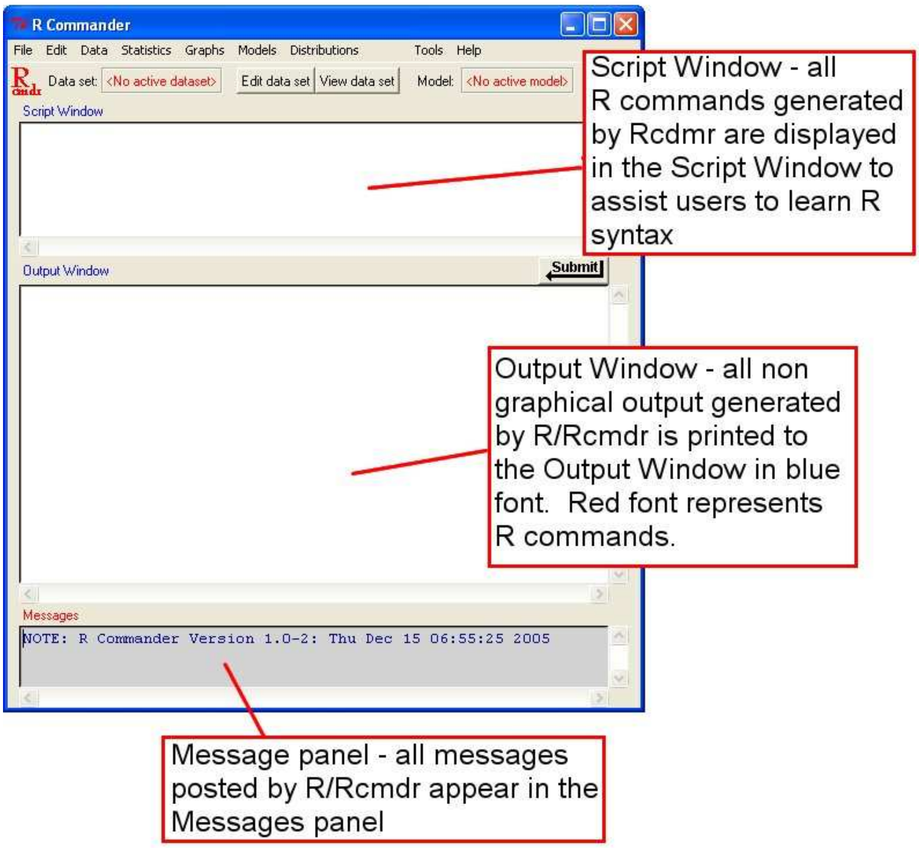


Figure 2: Rcmdr GUI

2 Rcmdr

Although R itself is a command driven statistical package, in recognition of the difficulty most students experience while learning to use command driven software, a package (`Rcmdr`) has recently been included that enables most basic statistical procedures to be performed using a graphical user interface (menus, buttons, boxes, etc).

To enable easy use of R (and `Rcmdr`), some additional procedures have been developed for `Rcmdr` by Murray Logan. These procedures extend the capacity and coverage of `Rcmdr` to include all topics and procedures relevant to BIO3011.

Hereafter, all procedures will relate to `Rcmdr` (the Rgui window) unless otherwise specified.

3 Data files

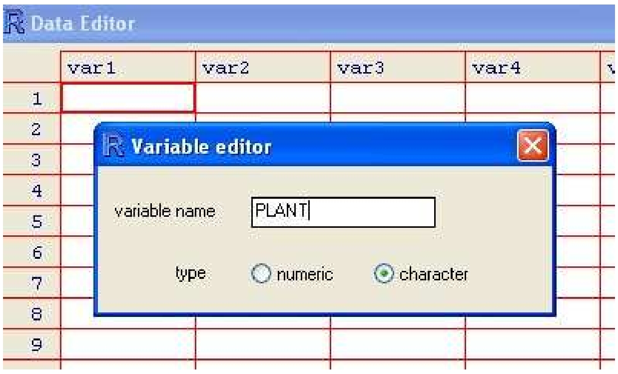
It is possible to have multiple data sets open at any time. As a result, each data set must be given a unique name by which it can be referred to and identified with.

3.1 Generating a new dataset using the R spreadsheet

Note that the spreadsheet offered by R is at this stage very rudimentary and offers only very limited editing facilities. R users usually use command-line procedures for data entry and dataset creation. Consequently, it is generally recommended that for serious data entry, a package such as excel should be used. The data can then be imported into R (see section 3.2.3).

1. Select the **Data** menu
2. Select the **New data set...** submenu
   *The **New Data Set** dialog box will appear.*

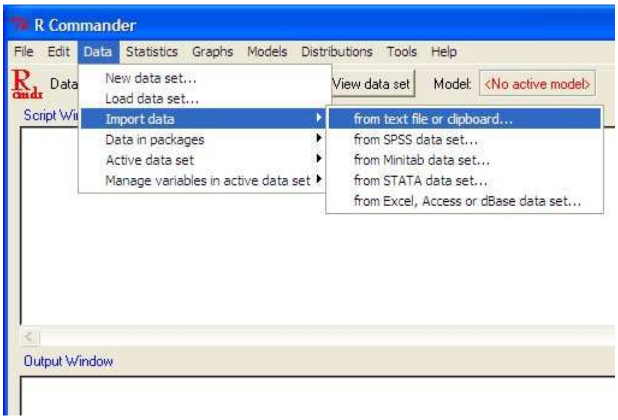
3. Enter a name for the data set
4. Click the **OK** button
   *The **R Data Editor Window** (R's graphical spreadsheet) window will appear within **RGui**. Switch control to **Rgui** using either **Alt-tab** or the Windows navigation buttons.*



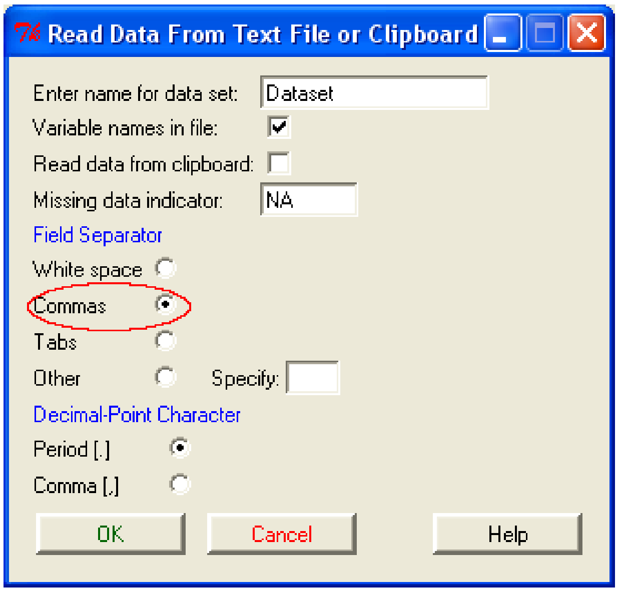
5. Clicking on a column heading and selecting **Change Name** from the resulting pop-up menu enables variable names to be customized.
6. Data are added by entering values in the cells
7. Close the **R Data Editor Window** window and the dataset will be created. You will notice that the **Data set panel** now displays the name of the newly generated dataset.

3.2 Opening an existing data file

3.2.1 Comma delimited text files - CSV



1. Select the **Data** menu
2. Select the **Import data..** submenu
3. Select the **from text file or clipboard..** submenu
   *The **Read Data from text file or clipboard** dialog box will appear.*



4. Provide a name for the imported data set (can be any name does not have to match the name of the file being imported). This is the name used to access the data once imported.
5. Select **Commas** as the Field Separator. This specifies how the columns are delimited (separated).
6. Click the **OK** button.
   *The **Read Data from Text File** dialog box will appear.*
7. Locate and select the file you wish to import (for BIO3011 these will always have the ***.csv** file extension). The data should now be ready to use.
8. To view the data set, click on the **View data set** button from the main R commander window.

The data are arranged in rows and columns - each row contains the data for one replicate unit. The top line of the file consists of variable names (i.e. names of each column). Each column represents a variable, and column names can consist of any number of characters (e.g. WEIGHT or LENGTH or NUMBER), however they must each begin with a letter rather than a number and cannot contain the following characters (_ , $ % ˆ & # *). Missing data are represented by a full stop (.). R will ignore these in the analysis. Make sure you distinguish missing values, where you have no data, from zeros, where you have data but the value was zero.

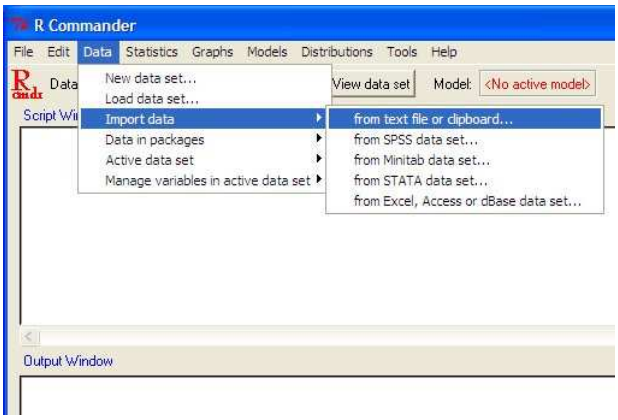
3.2.2 SYSTAT or SPSS files

1. Select the **Data** menu
2. Select the **Import data..** submenu
3. Select the **from SYSTAT data set..** or **from SPSS data set..** submenu
   *The **Import SYSTAT data set** or Import SPSS data set dialog box will appear.*
4. Enter a unique name to be assigned to the imported data set. Remember that while this can be any name (and doesn't necessarily need to be the same as the name of the imported file), a name that describes the data set is recommended.
5. Keep any other default options and click the **OK** button
6. Locate the file you wish to import and click the OK button.
   The data should now be ready to use.
7. To view the data set, click on the **View data set** button from the main **Rgui** window.

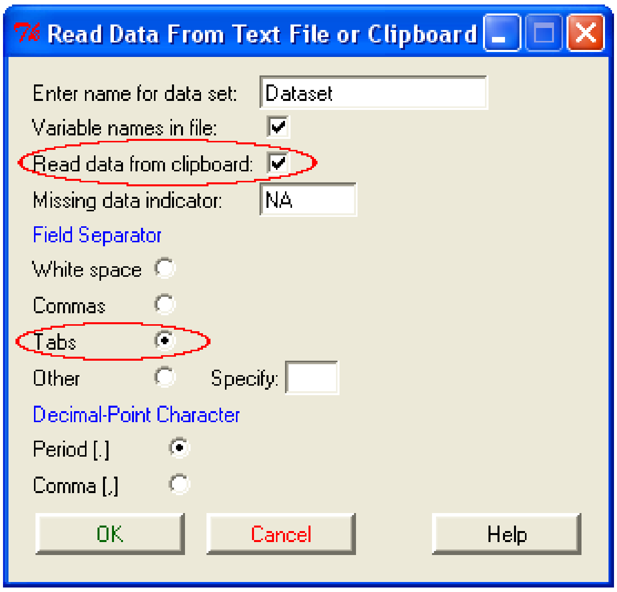
3.2.3 Excel files

At this stage, R does not support the native excel format. However, an excel sheet can be saved (in excel) as a `comma delimited text file (*.csv)`. This can then be imported directly into R (see section 3.2.1)

3.3 Importing from the clipboard



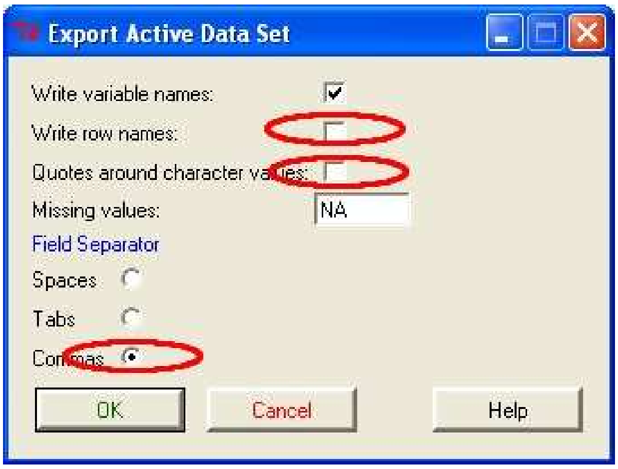
1. Select the **Data** menu
2. Select the **Import data..** submenu
3. Select the **from text file or clipboard..** submenu
   *The **Read Data from text file or clipboard** dialog box will appear.*



4. Provide a name for the imported data set (can be any name). This is the name used to access the data once imported.
5. Check the **Read data from clipboard** checkbox.
6. Select **Tabs** as the Field Separator. This specifies how the columns are delimited (separated). Most programs place text onto the clipboard in tab delimeted format.
7. Click the **OK** button. The data should now be ready to use.
8. To view the data set, click on the **View data set** button from the main R commander window.

3.4 Saving a data file

1. Select the **Data** menu
2. Select the **Active data set..** submenu
3. Select the **Export active data set..** submenu
   *The **Export Active Data Set** dialog box will appear.*
4. UN-check the **Quotes around character values**.
5. Select **Commas** as the Field Separator. This specifies how the columns are delimited (separated).
6. Click the **OK** button.
   *The **Export Data from text file** dialog box will appear.*
7. Supply a filename and path for the output file (for BIO3011 always use a ***.csv** file extension). The data should now be saved.



3.5 Examining and editing data files

Viewing

1. Click the **View data set** to view a data set
   A window containing the data set will appear. Note that the data in this window cannot be edited, only viewed.

Editing

1. Click the **Edit data set**
   *The **R Data Editor Window** dialog box will appear.*
2. Make any alterations to the spreadsheet (note that it is a fairly primitive spreadsheet)
3. Click the **Quit** button. The changes are now made. Note that this only alters the data in memory, not in the original file. To apply the changes to the file, save the data set using the instructions in section 3.4.

3.6 Data Transforms

1. Select the **Data** menu
2. Select the **Manage variables in active data set..** submenu
3. Select the **Compute new variable..** submenu
   *The **Compute New Variable** dialog box will appear.*
4. Enter the name of a new variable (should be a unique name) in the **New variable name** box
5. Enter an transformation expression (see Table 1) in the **Expression to compute** box
6. Click the **OK** button
   A new variable (containing the transformed data) should now have been added to the data set. Confirm this by viewing the data set (see section 3.5).

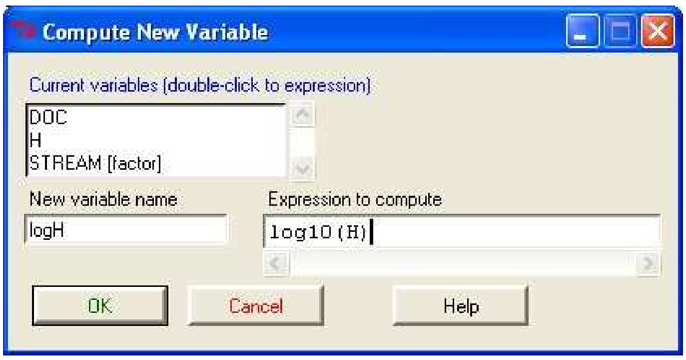
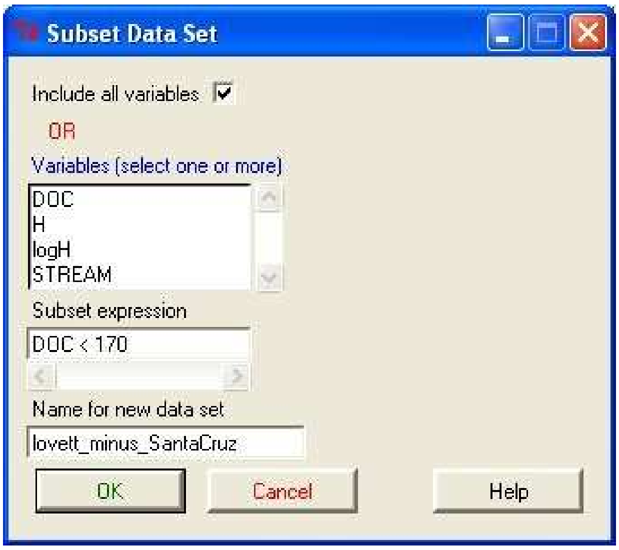


Table 1 Common data transformations

Nature of data	Transformation	R Expression
Measurements (lengths, weights, etc)	$\log_e$	`log(`**VAR**`)`
Measurements (lengths, weights, etc)	$\log_{10}$	`log(`**VAR**`,10)`
Counts (number of individuals, etc)	$\sqrt{}$	`sqrt(`**VAR**`)`
Percentages (data must be proportions)	$\arcsin$	`asin(sqrt(`**VAR**`))`
	scale (mean=0,unit variance)	`scale(`**VAR**`)`

where **VAR** is the name of the vector (variable) whose values are to be transformed.

3.7 Selecting subsets or subgroups of data



1. Select the **Data** menu
2. Select the **Active data set..** submenu
3. Select the **Subset active data set..** submenu
   *The **Subset Data Set** dialog box will appear.*
4. If all the variables are to be retained, ensure that the **Include all variables** check-box is checked. Otherwise, select the variables to retain from the Variables box.
5. Enter a subset expression (see Table 2) in the **Subset Expression** box
6. Enter a name for the subset data set into the **Name for new data set** box (for example, subsetDataSet). This should be a unique name that enables the data set and its contents to be easily recognized for future use.
7. Click the **OK** button
   A new data set (containing only the defined subset of the data) should now have been created. Confirm this by viewing the data set (see section 3.5).

Table 2 Listing or referencing subsets of the data

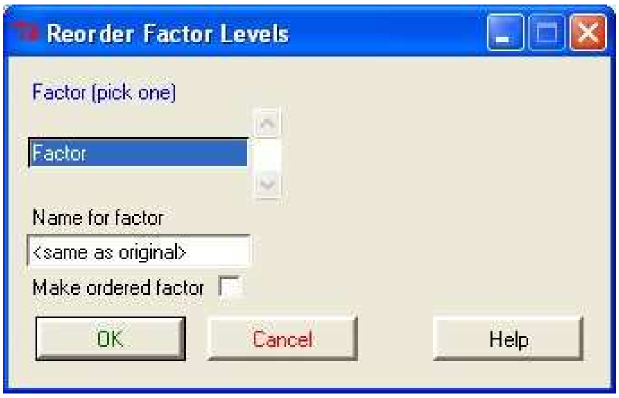
Selection	**Command**
Values of **Var** less than 50	`Var<50`
The first 10 values in **Var**	`Var[1:10]`
The 20th to the 50th value of **Var**	`Var[20:50]`
Only those entries whose values of **Var** are **High**	**Var**==’High’

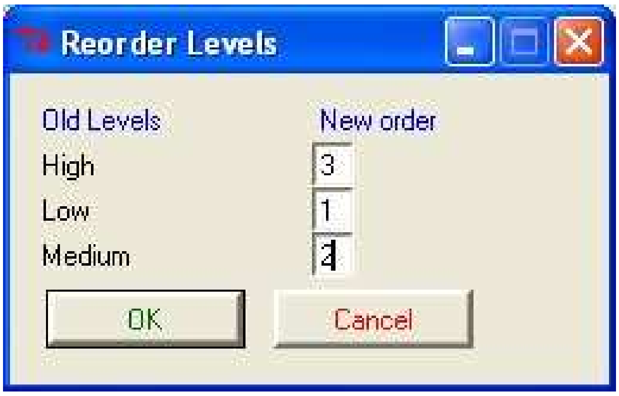
3.8 Reordering factor levels

Consider the following data set. There are three levels of the categorical variable (Factor) and they appear in alphabetical order. In fact even if they were entered in an alternative order, when R (or any other statistical software) compiles the list of the levels of the categorical variable in memory, by default the levels are placed in alphabetical order. While the order of factor levels is not important for statistical analyses, sometimes when generating graphs it is more preferable to have the levels ordered differently. For example, it is more preferable for a graph that summarizes the data in table 3 to order the levels of Factor as `Low`, `Medium`, `High` rather than alphabetically (`High`, `Low`, `Medium`). Table 3 Listing or referencing subsets of the data

DV	**Factor**
16	High
12	High
1	Low
3	Low
5	Medium
7	Medium

1. Select the **Data** menu
2. Select the **Manage variables in active data set..** submenu
3. Select the **Reorder factor levels..** submenu
   *The **Reorder Factor Levels** dialog box will appear.*
4. Select the categorical (factor) variable whose levels you wish to reorder from the Factor box.
5. Click the **OK** button
   You will be warned that the variable already exists, this is OK, press the **Yes** button
   The Reorder Levels dialog box will appear.
6. The current order of the levels in the factor will be presented. Using the entry boxes, provide a new order. A 1 indicates the first in the order.
7. Click the **OK** button
   The factor will be reordered. Note that this only affects how R internally considers the ordering of factor levels. It will not visibly alter the data set or file in any way.



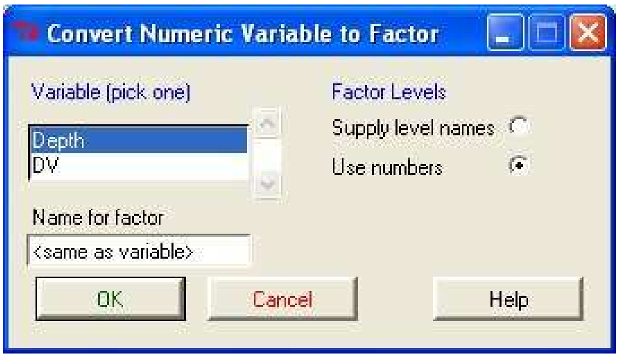


3.9 Converting numeric variable to a factor

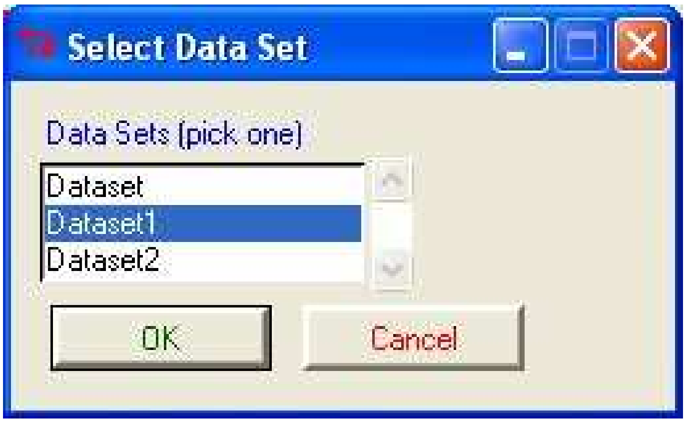
Generally, factors (categories) are entered as words. When this is the case R automatically recognizes the variable as a factor and therefore a categorical (rather than continuous) variable. However, occasionally the levels of a categorical variable may be numbers. For example, you might have a categorical variable to depict the water depth at which samples were collected. Samples may have been collected at 0, 5, 10 and 15 meters below sea level. In this case, your factor levels are 0, 5, 10, and 15. However, as these are numbers (rather than words), R will not automatically consider the variable as a category. It is possible, however, to convert such a numeric variable into a factor variable.
Table 4 Listing or referencing subsets of the data

DV	Depth
16	0
12	0
1	5
3	5
5	10
7	10

1. Select the **Data** menu
2. Select the **Manage variables in active data set..** submenu
3. Select the **Convert numeric variable to factor..** submenu
   *The **Convert Numeric Variable to Factor** dialog box will appear.*
4. Select the variable to be converted into a factor from the Variable box.
5. Select the **Use numbers** option
6. Click the **OK** button
   You will be warned that the variable already exists, this is OK, press the **Yes** button
   The variable will be converted into a factor. Note that this only affects how R internally perceives the variable type. It will not visibly alter the data set or file in any way.



3.10 Switching between different loaded data sets

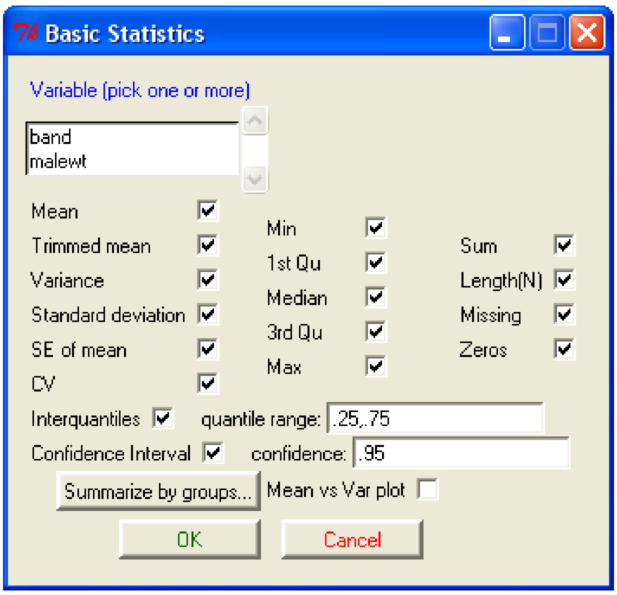


1. Click on the **Data set** display panel in the RGui window
   The Select Data Set dialog box will be displayed
2. Select the required data set
3. Click **OK**

4 Summary Statistics

4.1 Univariate

1. Select the **Statistics** menu
2. Select the **Summaries..** submenu
3. Select the **Basic statistics..** submenu
   *The **Basic Statistics** dialog box will appear.*
4. Enter a name for to call the resulting table of summary statistics - the default name is usually fine.
5. Select the variable(s) to summarize from the Variable box
6. Select the required statistics
7. Click the **OK** button
   A table containing the statistics will appear in the output window.



4.2 Bivariate

Follow the steps outlined in section 4.1 above. In addition, click the **Summarize by groups..** button and select a grouping variable. A table containing the statistics will appear in the output window. Note, it is not possible at this stage to summarize the statistics by groups for multiple variables at a time!

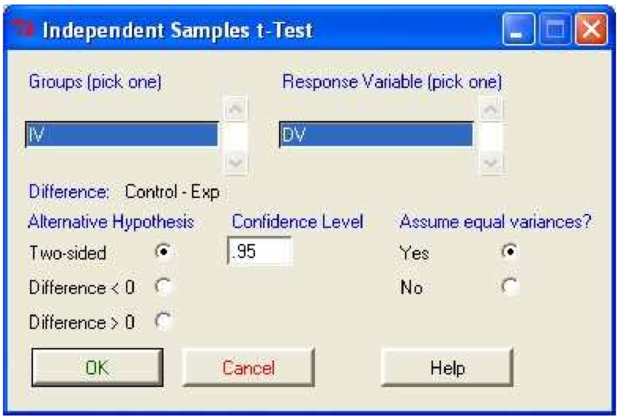
5 Two sample tests

Table 5 Example of the general format of data for two sample tests

DV	IV
1	Exp
2	Exp
3	Exp
3	Control
5	Control
8	Control

5.1 Independent t-tests

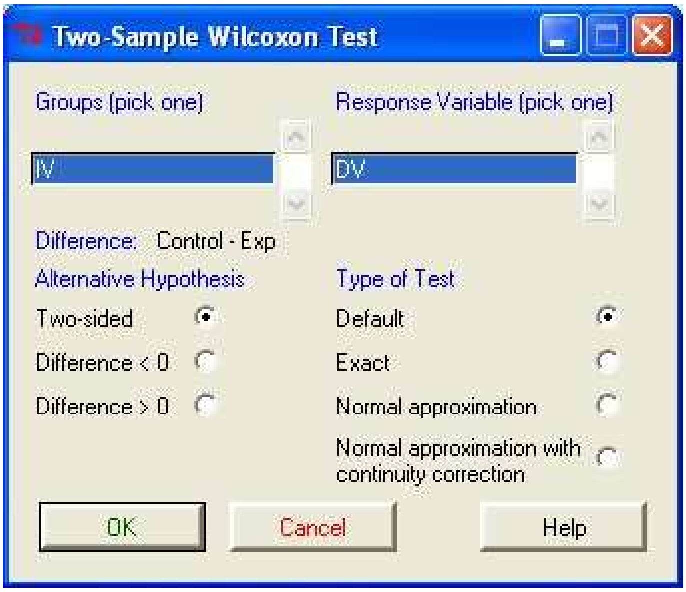
1. Select the **Statistics** menu
2. Select the **Means..** submenu
3. Select the **Independent samples t-test..** submenu
   *The **Independent Samples t-Test** dialog box will appear.*



4. Select the grouping (categorical) variable from the box. Note, this variable must contain two groups. Once selected, a label will appear to inform you of which groups are being compared (and the direction of the comparison).
5. Select the response (dependent) variable from the Response box.
6. For pooled variance t-test select the **Yes** option for Assume equal variances?, otherwise select **No**
7. Click the **OK** button
   The results will appear in the output window.

5.2 Mann-Whitney-Wilcoxon test

1. Select the **Statistics** menu
2. Select the **Non-parametric tests..** submenu
3. Select the **Two sample Wilcoxon test..** submenu
   The ***Two-Samples Wilcoxon Test*** *dialog box will appear.*
4. Select the grouping (categorical) variable from the Groups box. Note, this variable must contain two groups. Once selected, a label will appear to inform you of which groups are being compared (and the direction of the comparison).
5. Select the response (dependent) variable from the Response Variable box.
6. Click the **OK** button
   The results will appear in the output window.

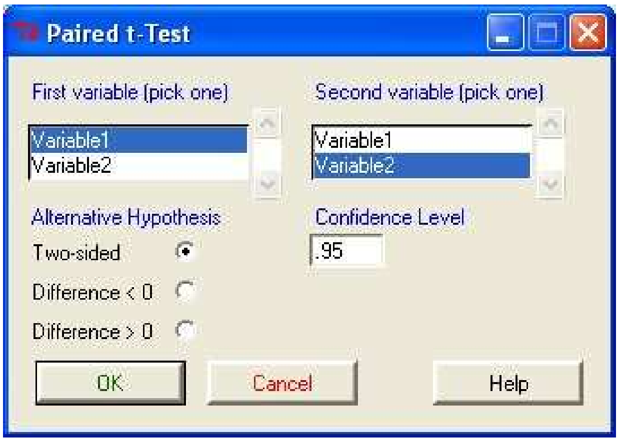


5.3 Paired t-test

Table 6 Example of the general format of data for paired t-test

Variable1	Variable2
1	2
2	4
3	3
3	4
5	7
8	10

1. Select the **Statistics** menu
2. Select the **Means..** submenu
3. Select the **Paired t-test..** submenu
   The ***Paired t-Test*** *dialog box will appear.*
4. Select one of the paired variables from the First variable box.
5. Select the other of the paired variables from the Second variable box.
6. Click the **OK** button
   The results will appear in the output window.



6 Correlations and Regression

Table 7 Example of the general format of data for correlation and regression. Note that the distinction is that in regression, one variable is identified as potentially dependent on the other, whilst in correlation, the direction or existance of causality is not implied.

a) Correlation

Variable1	Variable2
1	2
2	4
3	3
3	4
5	7
8	10

b) Regression

DV	IV
1	2
2	4
3	3
3	4
5	7
8	10

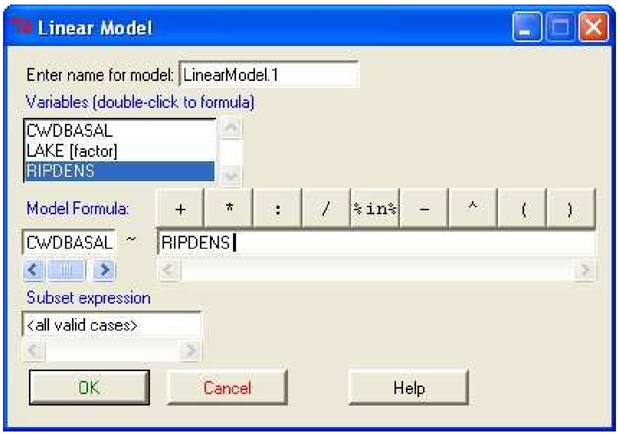
6.1 Correlation

1. Select the **Statistics** menu
2. Select the **Summaries..** submenu
3. Select the **Correlation..** submenu
   *The **Correlation** dialog box will appear.*
4. Select the variables to be correlated from the Variables box. To select multiple variables, hold the CNTRL key while making selection.
5. Select the appropriate correlation type (Pearson is default).
6. Click the **OK** button
   The results will appear in the output window. If two variables were selected, the full correlation output (including t-test and confidence intervals) is generated. If more than two variables are selected, a matrix of correlation coefficients and a matrix of associated probabilities (uncorrected) are generated.

6.2 Simple linear Regression

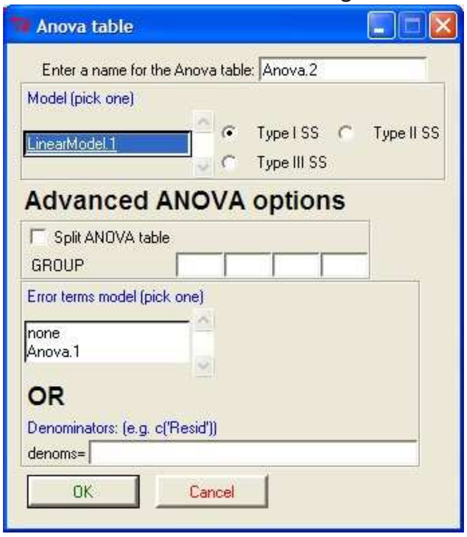
Note that it is also possible to follow the steps for ANOVA in section 7.1

1. Select the **Statistics** menu
2. Select the **Fit models..** submenu
3. Select the **Linear model..** submenu
   *The **Linear Model** dialog box will appear.*
4. Enter a name for the model output in the **Name for model** box. This can be any name but should be informative enough to remind you of what statistic was performed
5. Double click on the dependent variable in the Variables box. This will add the dependent variable to the text box on the left hand side of the ˜ under Model formula
6. Double click on the indepedent variable (the predictor variable) in the Variables box. This will add the predictor variable to the text box on the right hand side of the under Model formula
7. Click the **OK** button
   The summary of the results will appear in the output window.



6.2.1 Regression ANOVA table

1. Select the **Models** menu
2. Select the **Hypothesis tests..** submenu
3. Select the **ANOVA table..** submenu
   *The **Anova table** dialog box will appear.*



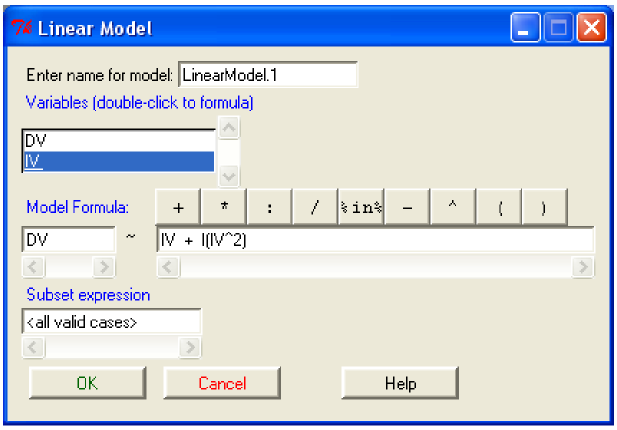
4. Select the model for which the ANOVA table is to be generated - this is the name you provided when you performed the Regression analysis
5. Click the **OK** button
   The regression ANOVA table will appear in the R Commander output window.

6.2.2 Regression diagnostics

1. Select the **Models** menu
2. Select the **Graphs..** submenu
3. Select the **Basic diagnostic plots..** submenu
   A set of four diagnostic plots will appear in a graphical window.

6.3 Polynomial Regression

1. Select the **Statistics** menu
2. Select the **Fit models..** submenu
3. Select the **Linear model..** submenu
   *The **Linear Model** dialog box will appear.*
4. Enter a name for the model output in the **Name for model** box. This can be any name but should be informative enough to remind you of what statistic was performed
5. Double click on the dependent variable in the Variables box. This will add the dependent variable to the text box on the left hand side of the ~ under Model formula
6. Double click on the indepedent variable (the predictor variable) in the Variables box. This will add the predictor variable to the text box on the right hand side of the under Model formula. So far this is a first order polynomial.



7. To add the second order component, add a plus (+) sign to the right hand side then include the independent variable followed by a hat (ˆ) sign and a 2 (see figure). Finally, enclose the independent variable, hat and 2 with a set of brackets preceded by an I. The I represents a function that preserves the polynomial component.

8. similarly to add higher order (3, 4,...) polynomial terms, follow the step above, using the powers 3, 4, etc.

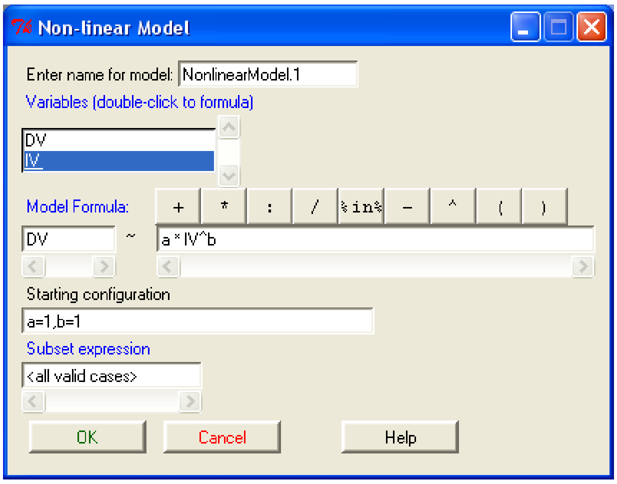
9. Click the **OK** button
   The summary of the results will appear in the output window.

6.4 Nonlinear Regression

1. Select the **Statistics** menu

2. Select the **Fit models..** submenu

3. Select the **Nonlinear model..** submenu
   *The **Non-linear Model** dialog box will appear.*



4. Enter a name for the model output in the **Name for model** box. This can be any name but should be informative enough to remind you of what statistic was performed

5. Double click on the dependent variable in the Variables box. This will add the dependent variable to the text box on the left hand side of the ˜ under Model formula

6. Construct the appropriate model on the right hand side of the under Model formula. For unknown parameters (constants), provide single letters (these must not be the names of any existing variables within the data set)

7. You must also define a starting configuration. This is a comma separated list of initial estimates for the unknown parameters. The non linear modeling process will progressively modify these estimates until the model best fits the data

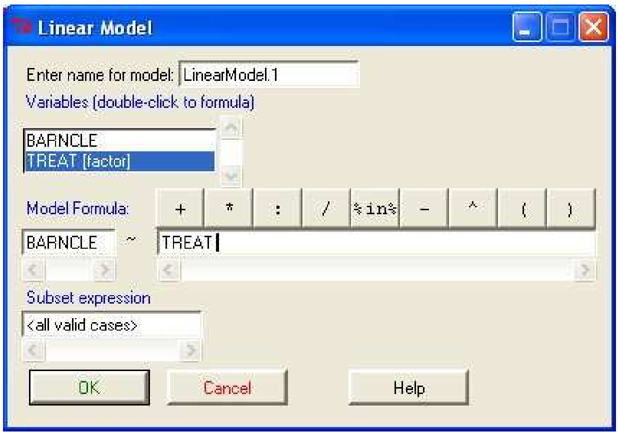
8. Click the **OK** button
   The summary of the results will appear in the output window.

7 ANOVA

See table 5 for an example of the data format for single factor ANOVA.

7.1 Single factor ANOVA

1. Select the **Statistics** menu
2. Select the **Fit models..** submenu
3. Select the **Linear model..** submenu
   *The **Linear Model** dialog box will appear.*
4. Enter a name for the model output in the **Name for model** box. This should be a unique name that enables the resulting model and its contents to be easily recognized for future use.
5. Double click on the dependent variable in the Variables box. This will add the dependent variable to the text box on the left hand side of the ˜ under Model formula
6. Double click on the categorical variable (the factor variable) in the Variables box. This will add the categorical variable to the text box on the right hand side of the under Model formula
7. Click **OK**
   A summary of the ANOVA results will appear in the output window.



7.1.1 ANOVA table

Follow the steps outlined in section 6.2.1.

7.1.2 ANOVA diagnostics

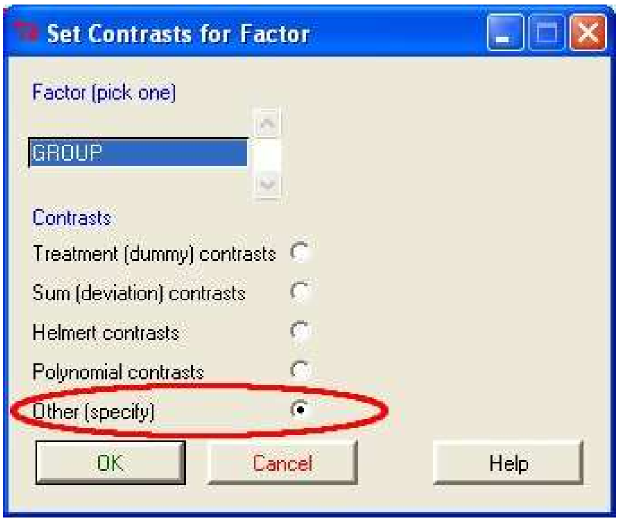
Follow the steps outlined in section 6.2.2

7.2 Post-Hoc Tukey's test

1. Select the **Models** menu
2. Select the **Hypothesis tests..** submenu
3. Select the **Tukeys test..** submenu
   *The **Tukey's test** dialog box will appear.*
4. Select the categorical (factorial) variable from the Factor list.
5. Click the **OK** button
   The Tukey's tests will appear in the output window. The tests are labeled a little strangely. Each test (row name) gives the factor name and level minus a different level of that factor name. For example, if the factorial variable was called `TREAT` and there were three levels of this factor (`High`, `Low`, & `Medium`) then one of the tests (rows) might be labeled as `TREATHigh-TREATLow`.

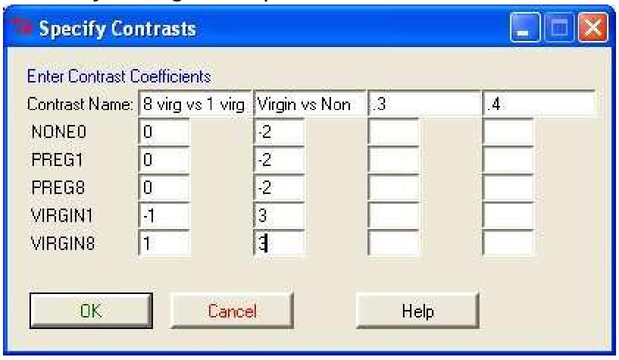
7.3 Planned Comparisons

As implied by the name (Planned comparisons), these are specific comparisons that planned at the design stage. Consequently planned comparisons (contrasts) are defined prior to fitting the linear model (running the ANOVA).

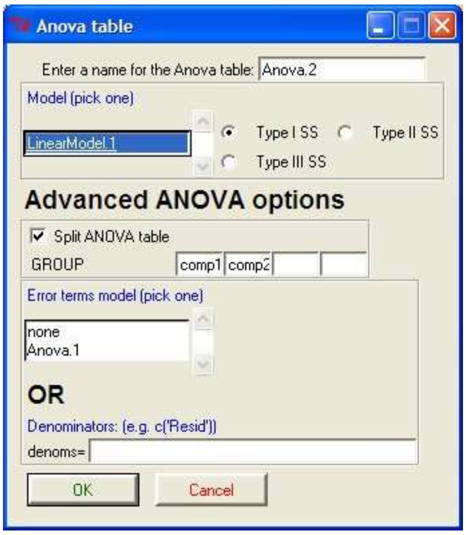


1. Select the **Data** menu
2. Select the **Manage variables in active data set..** submenu
3. Select the **Define contrasts for a factor..** submenu
   *The **Set Contrasts For Factor** dialog box will appear.*
4. Select the categorical (factorial) variable from the Factor list.
5. Select the **Other (specify)** option.
   *The **Specify Contrasts** dialog box will appear.*

A matrix will be initiated with the levels of the categorical variable used as the row names. There will be n-1 columns (where n is the number of levels in the categorical variable), reflecting the maximum number of planned comparisons allowable. It is possible to define (n-1) planned comparisons, although, it is not necessary to define this maximum number of comparisons. For example, you can decide to define only a single comparison.



6. Enter a name for each comparison you intend to define in the **Contrast Name:** box(es)
7. Enter the contrast coefficients in each column.
8. Click the **OK** button
   If the defined contrasts are orthogonal (independent) the full matrix of contrasts will be displayed in the R Commander Output Window, otherwise you will be returned to the **Set Contrasts For Factor** dialog box for another attempt.



9. Fit the linear model according to the steps outlined in section 7.1
10. To examine the ANOVA table that includes the planned comparisons (contrasts)
    (a) Select the **Models** menu
    (b) Select the **Hypothesis tests..** submenu
    (c) Select the **ANOVA table..** submenu
    *The **Anova table** dialog box will appear.*
    (d) Click the **Split ANOVA table** check button. A table listing the factor(s) in the model and the contrast names that were defined when the contrasts for the factor(s) were defined will be listed. Note that there will be n-1 (where n is the number of groups) defined comparisons.
    (e) Delete the text in the table for the comparisons that you are not interested in. The text for required comparisons can be modified if necessary
    (f) Click **OK**
    A summary of the ANOVA results will appear in the Output Window.

7.4 Factorial ANOVA

1. Select the **Statistics** menu
2. Select the **Fit models..** submenu
3. Select the **Linear model..** submenu
   *The **Linear Model** dialog box will appear.*
4. Enter a name for the model output in the **Name for model** box. This should be a unique name that enables the resulting model and its contents to be easily recognized for future use.
5. Double click on the dependent variable in the Variables box. This will add the dependent variable to the text box on the left hand side of the ~ under Model formula
6. Double click on a categorical variable (the factor variable) in the Variables box. This will add the categorical variable to the text box on the right hand side of the under Model formula
7. Click on the * button. This symbol means 'crossed' and is an abbreviated way of meaning include the two terms either side of this symbol plus their interaction.
8. Double click on the other categorical variable (factor variable) in the Variables box. This will add the categorical variable to the Model formula
9. Click **OK**
   A summary of the ANOVA results will appear in the output window.

Table 8 Example of the general format of factorial data

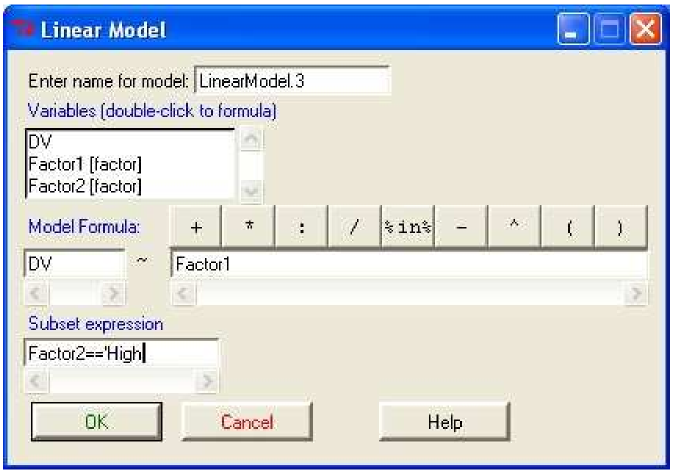
DV	Factor1	Factor2
10	Big	High
5	Medium	High
3	Small	High
11	Big	High
7	Medium	High
1	Small	High
7	Big	Low
5	Medium	Low
4	Small	Low
5	Big	Low
8	Medium	Low
6	Small	Low

7.5 Simple main effects

Following a factorial ANOVA with a significant interaction, it is usual to attempt to examine the simple main effects. That is explore the effect of one of the factors for each level of the other factor(s). There are a number of steps involved in this procedure.

1. **Perform global ANOVA** - the fully factorial ANOVA (see section 7.4)
2. **Analyze the effect of one factor for each level of the other factor(s)** - for example, for the data set in table 8 we might decide to analyze the effects of `Factor1` separately for each level of `Factor 2`

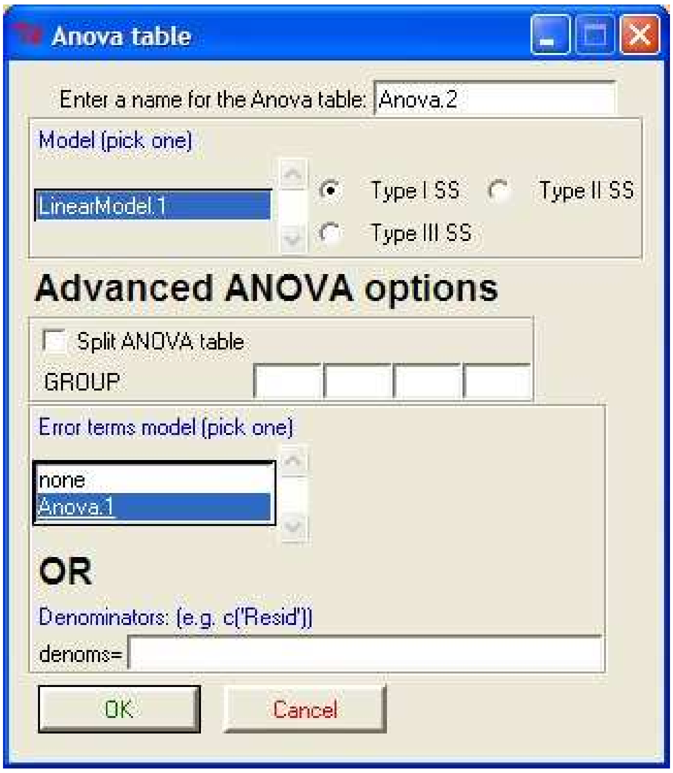
(a) Select the **Statistics** menu
(b) Select the **Fit models..** submenu
(c) Select the **Linear model..** submenu
*The **Linear Model** dialog box will appear.*
(d) Enter a name for the model output in the **Name for model** box. This should be a unique name that enables the resulting model and its contents to be easily recognized for future use.
(e) If you have just performed the fully factorial ANOVA prior to this step, then the previous model will already be setup. Retain this model the main factor you wish to explore (`Factor1` in the example in table 8) and remove the other factor(s) from the model (just delete the words including the * sign).



(f) Use the Subset Expression box to indicate one level of the other factor (in this case `Factor2`) in a similar way to described in section 2. In our example, to analyze the effects of `Factor1` on `DV` for just the `High` level of `Factor2`, the statement in the Subset Expression box would be: `Factor2 == 'High'`
(g) Click **OK**
A summary of the ANOVA results will appear in the output window.

3. **View the ANOVA with the correct residual term**

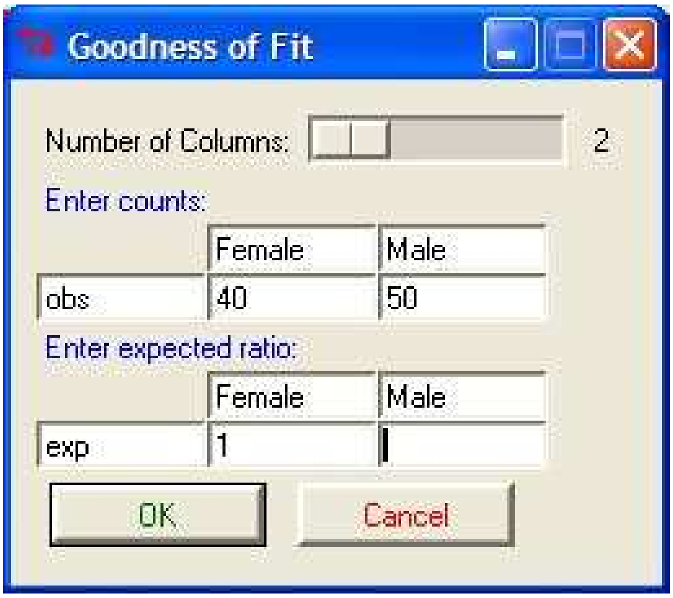
(a) Select the **Models** menu
(b) Select the **Hypothesis tests..** submenu
(c) Select the **ANOVA table..** submenu
*The **Anova table** dialog box will appear.*
(d) Select the model for which the ANOVA table is to be generated - this is the name you provided when you performed the above analysis
(e) Select the model for which the fully factorial ANOVA (Global) - this is the name you provided when you performed the fully factorial analysis and is used to provide the correct error term for the simple main effects.
(f) Click the **OK** button
The simple main effects ANOVA table will appear in the R Commander output window.



8 Analysis of frequencies

8.1 Goodness of fit test

1. Select the **Statistics** menu
2. Select the **Summaries..** submenu
3. Select the **Goodness of fit test** submenu *The **Goodness of fit test** dialog box will appear.*
4. Specify the number of categories with the **Number of columns** slider and enter the counts manually in the Enter counts table. Note, that the column titles by default are 1, 2... These can be changed to more meaningful names by editing the entries (e.g Male & Female).
5. Enter the expected frequencies or frequency ratio in the Enter expected ratio table
6. Click **OK**
   A table of observed and expected values as well as a goodness-of-fit test (Chisq) will appear in the output window.

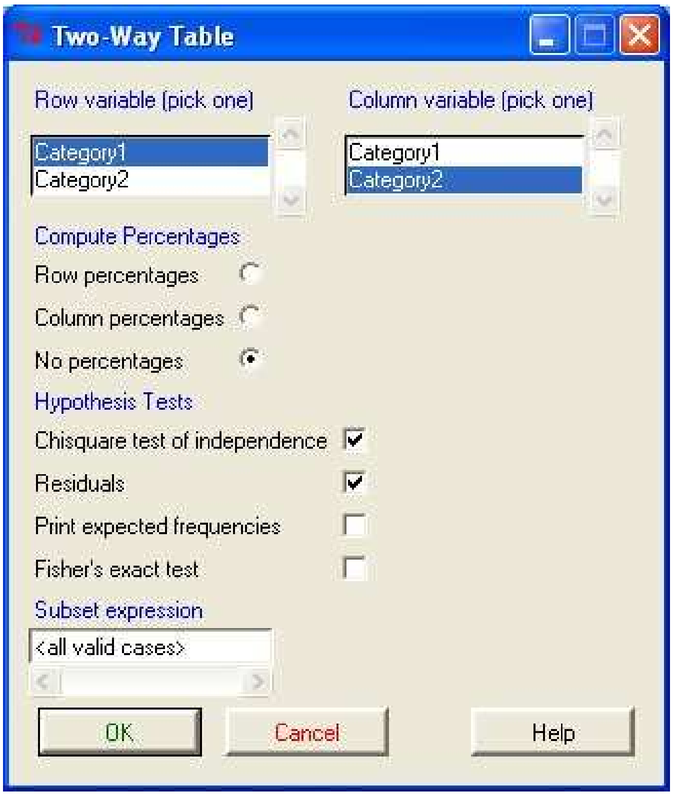


8.2 Contingency tables - un-compiled counts

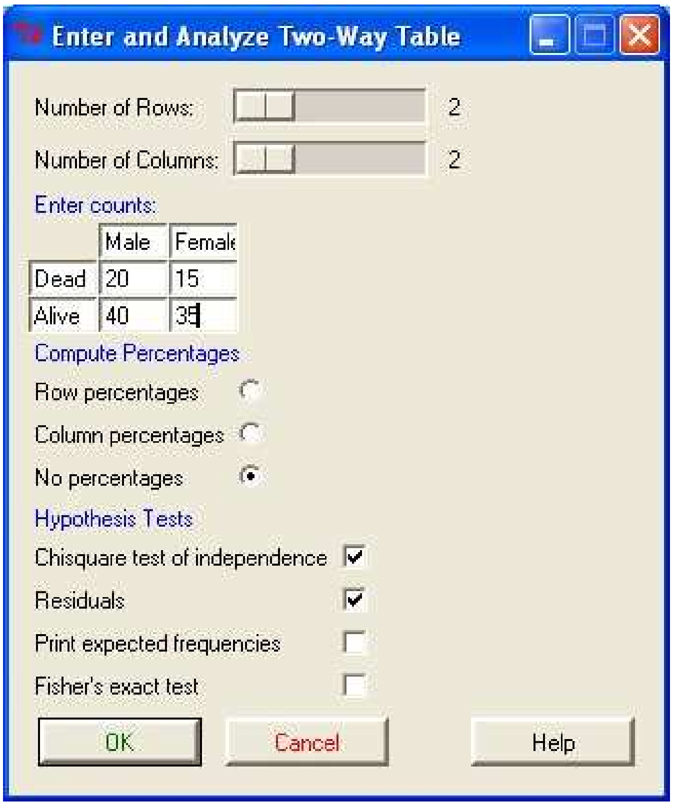
Table 9 Example of the general format of un-compiled frequency data

Category1	Category2
Male	Dead
Female	Dead
Male	Dead
Female	Alive
..	..

1. Select the **Statistics** menu
2. Select the **Contingency Tables..** submenu
3. Select the **Two-way table** submenu *The **Two-way table** dialog box will appear.*
4. Select the row and column variables (it doesn't matter which variable is row and which is column) from the list boxes
5. Select the **Chisquare test of independence** and **Residuals** options under Hypothesis Tests
6. Click **OK**
   A table of observed values, the Pearson's Chi-squared test output and a table of residuals will appear in the output window.



8.3 Contingency tables - pre-compiled counts



1. Select the **Statistics** menu
2. Select the **Contingency Tables..** submenu
3. Select the **Enter and analyze two-way table** submenu *The **Enter and analyze two-way table** dialog box will appear.*
4. Specify the number of rows and columns with the corresponding sliders. Note, it does not matter which of the two categorical variables you use as the row and which as the column variable
5. Enter the counts and the variable categories manually in the Enter counts table.
6. Select the **Chisquare test of independence** and **Residuals** options under Hypothesis Tests
7. Click **OK**
   A table of observed values, the Pearson's Chi-squared test output and a table of residuals will appear in the output window.

9 Multivariate analysis

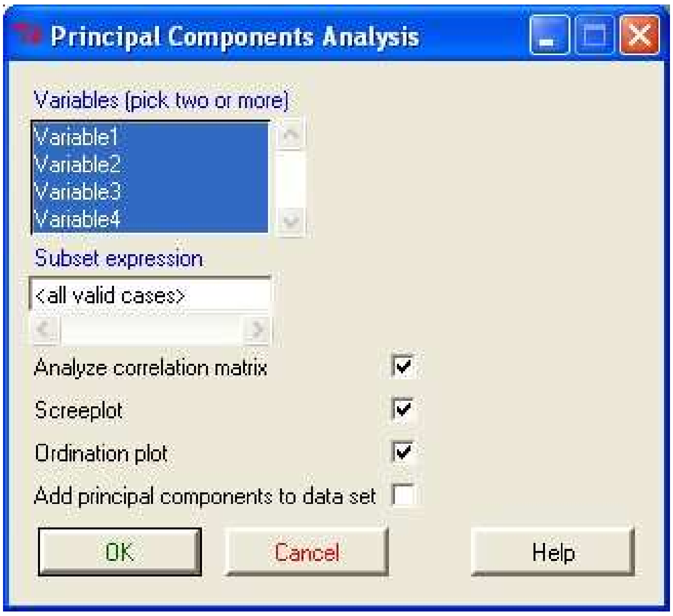
Table 10 Example of the general format of data for multivariate analysis in R

	Variable1	**Variable2**	**Variable3**	**Variable4**
Site1	1	0	5	34
Site2	0	0	8	21
Site3	7	9	3	17
Site4	9	12	5	6
Site5	9	12	5	6

9.1 PCA - Principal components analysis

Need to have variables (e.g. species) in columns and samples (e.g. sites) in rows

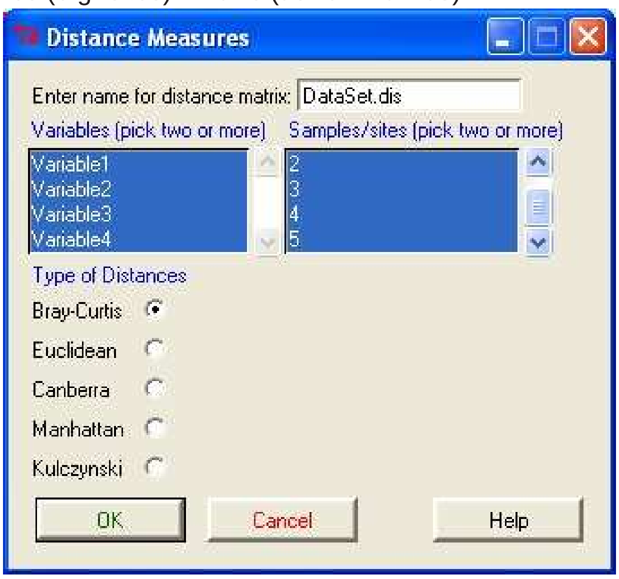
1. Select the **Statistics** menu
2. Select the **Dimensional analysis..** submenu
3. Select the **Principal-components analysis** submenu *The **Principal-components analysis** dialog box will appear.*
4. Select the variables from the Variables list
5. Select **Analyze correlation matrix** to base calculations on a correlation matrix, otherwise covariance matrix is used
6. Select **Screeplot**
7. Select **Ordination**
8. Click **OK**
   The component loadings and component variances will appear in the output window. In the **R Console** window of **RGui** will be prompt you to hit the `<Return>` key to cycle through two figures to be drawn on the newly created **Graph** window within **(**RGui. The first of these figures is a screeplot, and the second is the ordination plot.



9.2 Distance measures

Need to have variables (e.g. species) in columns and samples (e.g. sites) in rows (as **row names**)

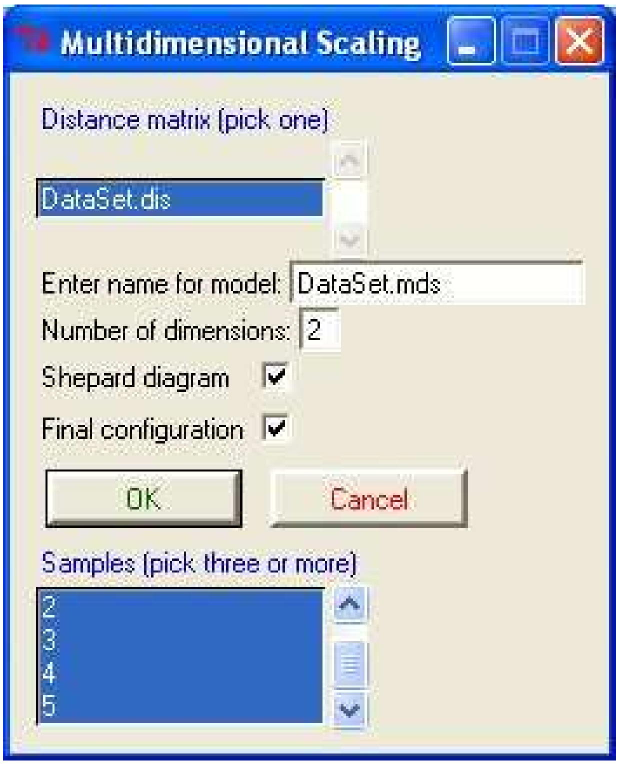
1. Select the **Statistics** menu
2. Select the **Dimensional analysis..** submenu
3. Select the **Distance measures** submenu *The **Distance measures** dialog box will appear.*
4. Enter a name for the resulting distance matrix. By default, Rcmdr will append the suffix `.dis` to the name of the currently active data set.
5. Select the variables from the Variables list
6. Select the type of distance measure from the blue-Type of Distance list of options
7. Click **OK**
   The distance matrix (rectangular) will appear in the output window.



9.3 MDS - Multidimensional Scaling

Need to provide a rectangular dissimilarity matrix (in `dist` format - the format output from the Distance() procedure, see section 9.2)

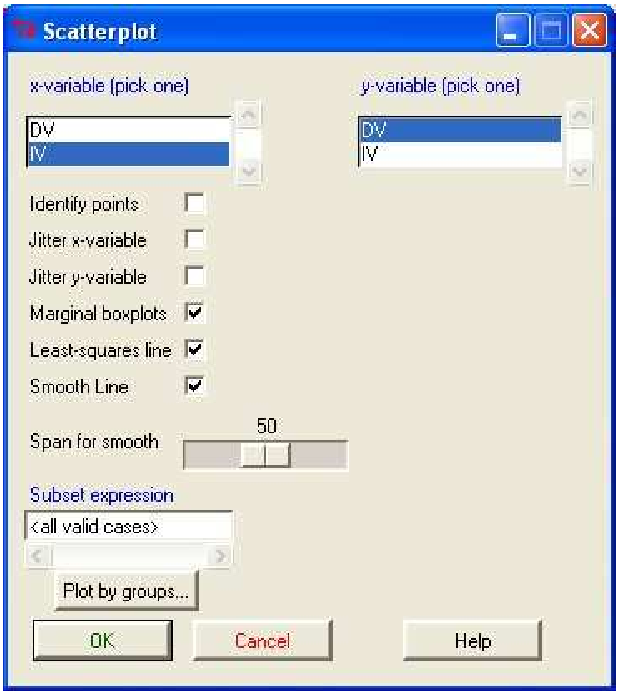
1. Select the **Statistics** menu
2. Select the **Dimensional analysis..** submenu
3. Select the **Multidimensional scaling** submenu
   *The **Multidimensional scaling** dialog box will appear.*
4. Select a distance matrix from the Distance matrix listbox//An additional listbox will be added to the bottom of the *Multidimensional scaling* dialog box. This lists the variables in the distance matrix. Select 3 or more to include in the MDS.
5. Enter a name for the resulting output (scores, and stress value) in the **Enter name for model:** box
6. Select the samples from the Samples list
7. Select **Shepard** to include a Shepard diagram
8. Select **Configuration** to include the final configuration plot
9. Click **OK**
   The final coordinates and stress value (as a percentage) will appear in the output window. In the **R Console** window of **RGui** will be prompt you to hit the `<Return>` key to cycle through two figures to be drawn on the newly created **Graph** window within **(**RGui. The first of these figures is a Shepard diagram, and the second is the configuration plot.



10 Graphs

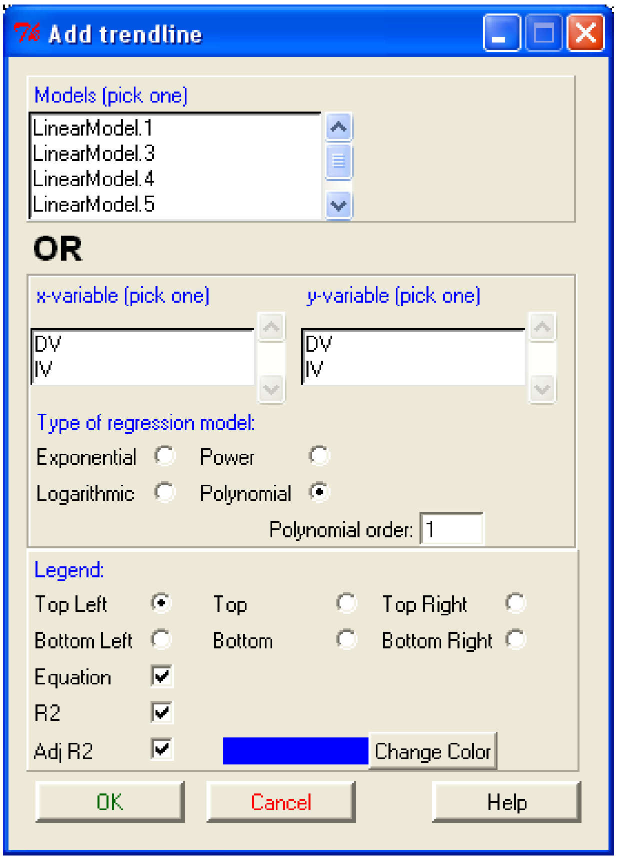
10.1 Scatterplots

1. Select the **Graphs** menu
2. Select the **Scatterplot..** submenu
   *The **Scatterplot** dialog box will appear.*
3. Select one of the variables (usually a independent or predictor variable) from the x-variables list box
4. Select another variable (usually a dependent or response variable) from the y-variable list box
5. Select the **Marginal boxplots** option to include boxplots in the margins
6. Select the **Least-squares line** option to include a regression line of best fit through the data
7. Select the **Smooth line** option to include a lowess smoother through the data
8. Click **OK**
   A scatterplot will appear in a graphical window of **RGui**.



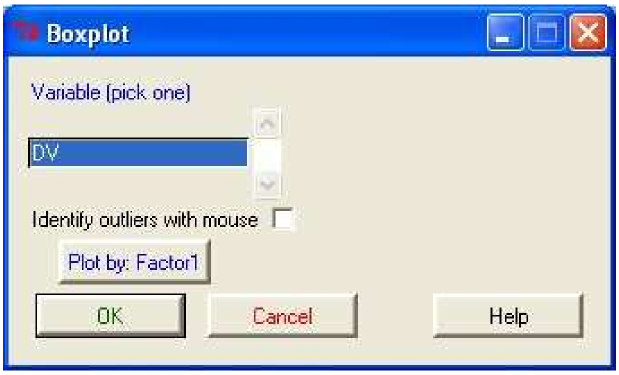
10.1.1 Trend lines

1. Select the **Graphs** menu
2. Select the **Trend lines..** submenu
   *The **Add trendline** dialog box will appear.*
3. Trend lines can either be constructed from a fitted model or from one of the generic line types
   (a) To construct a trend line from a fitted model, select the model from the Models box
   (b) To construct a trend from a generic line type (exponential, power, logarithmic and polynomial);
      i. Select the independent variable from the x-variables list box
      ii. Select the dependent variable from the y-variables list box
      iii. Select the type of regression model from the Type of regression model options
4. Indicate what to include in the legend, what color and where the legend should be located within the plot area
5. Click **OK**
   The trend line and legend will be added to the current scatterplot in a graphical window of **RGui**.

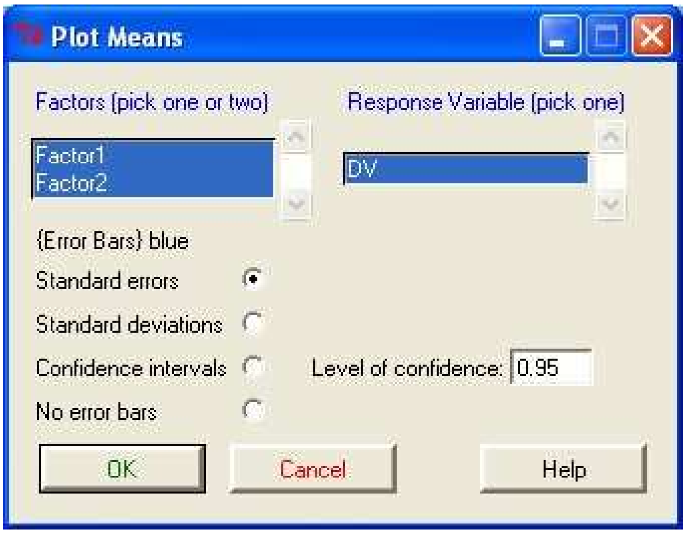


10.2 Boxplots

1. Select the **Graphs** menu
2. Select the **Boxplot..** submenu
   *The **Boxplot** dialog box will appear.*
3. Select the dependent variable from the Variable list
4. To generate separate boxplots according to the levels of a categorical variable, click the **Plot by groups..** button , select a categorical variable (factor) from the list of Groups variable and click the **OK** button
5. Click **OK**
   A boxplot will appear in a graphical window of **RGui**.



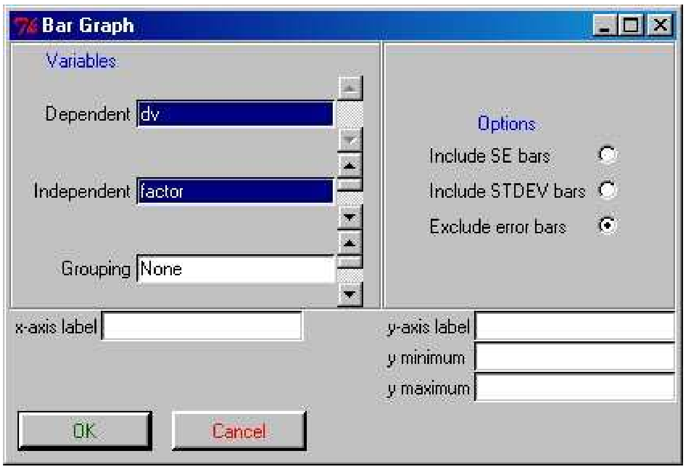
10.3 Interaction plots



1. Select the **Graphs** menu
2. Select the **Plot of means..** submenu*The **Plot Means** dialog box will appear.*
3. Select the categorical (factorial) variable(s) from the Factors list
4. Select the dependent variable from the Response list
5. Select the type of error bars from the Error Bars options
6. Click **OK**
   A interaction plot will appear in a graphical window of **RGui**.

10.4 Bargraphs

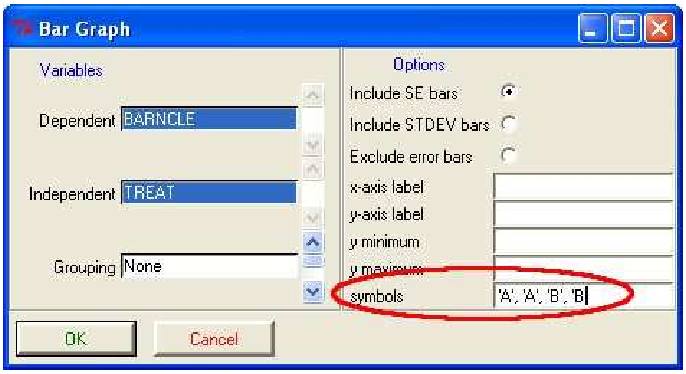
1. Select the **Graphs** menu
2. Select the **Bargraph..** submenu
   *The **Bar Graph** dialog box will appear.*
3. Select one of the dependent variable from the **Dependent** list box
4. Select one of the categorical (independent) variable from the **Independent** list box
5. For a two-factor bargraph, select another categorical (independent) variable from the **Grouping** list box. To avoid clutter and confusion, in a two-factor bargraph it is best to have the categorical variable with the greater number of levels as the independent (x-axis) variable and the other variable as the grouping variable.
6. Select the type of error bars from the Options
7. It is also possible to set the x and y labels as well as the upper and lower limits of the y axis.
8. Click **OK**
   A bargraph will appear in a graphical window.



10.5 Symbols on bargraphs

From the Bar Graph dialog box

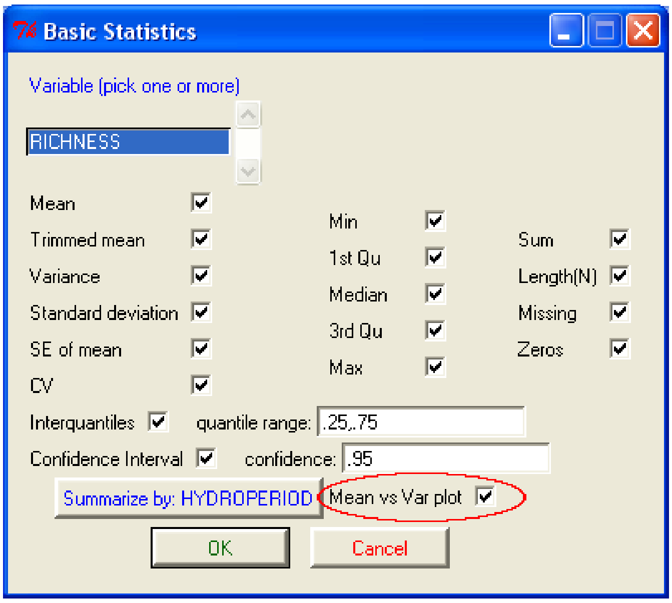
1. Select the variables and settings according to section 10.4
2. Enter a series of symbols that are to appear above the error bars of each bar. The following formatting rules are important:
   (a) Each symbol should be surrounded by a set of quotation marks (e.g. 'A')
   (b) Symbols should be listed with commas separating each symbol (e.g. 'A','B','A')
   (c) There should be as many symbols as there are bars on the graph. If less symbols are required than there are bars, then blank symbols (' ') are used for those bars not requiring a symbol
   (d) The following are all valid symbol definitions for a factor with 4 groups:
   'A','A','B','B'
   'A', ' ', ' ','B'
   ' ', ' ', '*', '**'



3. Define the series of symbols such that common symbols signify non-significant comparisons and differences between symbols signify significant differences.
4. Click the **OK** button
   A bargraph will appear in a graphical window.

10.6 Plot of mean versus variance

1. Select the **Statistics** menu
2. Select the **Summaries..** submenu
3. Select the **Basic statistics..** submenu
   *The **Basic Statistics** dialog box will appear.*
4. Select a dependent variable from the **Dependent** list box
5. Select the at least the Mean and Variance options
6. Click the **Mean vs Var plot** checkbox
7. Click **Summarize by groups**
   The Groups dialog box will be displayed
8. Select the categorical variable from the Groups variable list
9. Click **OK** in the *Groups dialog box*
10. Click **OK**
    Along with a table of statistics, a plot of Mean vs Variance will appear in an RGui Graphics Window.



11 Saving results

11.1 Graphs

11.1.1 Copying

1. Right-click on the graph
2. Select either **copy as metafile** (if intending to modify/edit the graph after it is pasted into another program) or **copy as bitmap** (if don't intend to modify the graph after it is pasted into another program)
3. Switch control to the other program using either **Alt-tab** or the Windows navigation buttons and paste the graph

11.1.2 Saving

1. Click on the graph to be saved. This will alter the **RGui** menus and buttons
2. From the **RGui** menus, select the **File** menu
3. Select the **Save as..** submenu
4. Select either the **JPEG 100% quality** submenu (if not intending to modify the graph after it is pasted into another program) or the **METAFILE** submenu (if intending to modify/edit the graph after it is pasted into another program
5. Use the *Save As* dialog box to provide a filename and path for the graph.
6. Click the **OK** button. The graph will then be saved.

11.2 Results

11.2.1 Copying

To copy and paste results from the **Rcmdr** output window

1. Highlight the results that you are interested in copying
2. From the **Rcmdr** menus, select the **Edit** menu
3. Select the **Copy** submenu
4. Switch control to the other program using either **Alt-tab** or the Windows navigation buttons and paste the graph

Note that you can also copy highlighted text by pressing the **Alt-c** key combination.

11.2.2 Saving

To save all of the results in the **Rcmdr** output window to a file

1. Select the **File** menu
2. Select the **Save output as...** submenu
3. Use the *Save As* dialog box to provide a filename and path for the graph.
4. Click the **OK** button. The results will then be saved.

Note, that when you save the output results to file, all of the results in the output window are saved, not just the highlighted text. if you are only interested in a small section of the output results you just need to cut the unwanted sections (either before saving, or later in a word processing program - like Word).

12 Common problems encountered

12.0.3 Error message - 'Package not found'

There are two common reasons for this:

1. When installing R from the CDROM, you ran the file called rw2000. This purely installs R. Solution: to install everything (including all the packages), install by running the **install.bat** provided.
2. When installing R, you asked for R to be installed in a location other than the default location. As a result, when the packages were installed, they were not installed in the same location as R - solution: uninstall R and install it again, this time allow it to be installed in the default location

12.0.4 I requested a graph, but I haven't been given one

- Usually this is because graphs appear as windows of **RGui**. You need to switch to **RGui** to see the graph.

12.0.5 I asked to create a new data set or clicked on Edit data set and nothing happened

- Usually this is because the spreadsheet for creating/editing data sets appears as a window of **RGui**. You need to switch to **RGui** to see the spreadsheet.

12.0.6 The window or dialog box disappeared

- Under some circumstances (such as moving a dialog box) under windows, a dialog box or window losses focus (that is it gets pushed behind another window). You just need to switch to this window or dialog box using either the **Alt-c** key combination or using the windows navigation bar.

12.0.7 Error message - 'There is no active data set'

- Import (see section 3.2) or manually create (see section) a data set
- Click on **Data set** panel (which probably says No active data set) in the main R Commander window

12.0.8 Working on the incorrect data set

- R has the capacity to have multiple data sets open simultaneously - this is one of its great strengths. However, with multiple data sets open at a time, it is necessary to be organized to prevent confusion. Each data set has a unique name and this helps to manage the different data sets, however it is still easy to lose track of which data set contains which data. It is therefore highly recommended that the names you give to each data set are highly descriptive.
  In Rcmdr, only one data set is considered to be the active data set. It is from this data set that it retrieves variables. To ensure that Rcmdr is operating on the correct data set, check that the **Data set** panel is displaying the name of the correct data set.

R is free software distributed by the R core development team under a GNU-style copyleft

Murray Logan, School of Biological Sciences, Monash University.
January 20, 2008