

SPSS Notes

(SPSS version 15.0)

Annie Herbert
Salford Royal Hospitals NHS Trust

July 2008

Contents

	Page
Getting Started	1
1 Opening SPSS	1
2 Layout of SPSS	2
2.1 Windows	2
2.2 Saving Files	3
3 Creating a Data File	4
3.1 Data View	4
3.2 Variable View (Defining Variables)	5
3.3 Applying attributes to multiple variables	9
3.4 Entering data	10
4 Data Manipulation	10
4.1 Editing	10
4.2 Sorting	11
4.3 Obtaining a File Summary	11
4.4 Searching through the contents of a variable	12
4.5 Go to Case	12
4.6 Go to Variable	12
5 Computing and Recoding	13
5.1 Compute	13
5.2 Recode	15
5.3 Automatic Recode	19
6 Opening an Existing File	20
6.1 Opening a SPSS file	20
6.2 Opening an Excel file	20
7 Split File and Select Cases	22
7.1 Split File	22
7.2 Select Cases	22
8 Displaying Data	23
General notation	23
8.1 Pie charts	24
8.2 Bar charts	25
8.3 Histograms	27
8.4 Box plots	28
9 Editing Graphs	30
9.1 Changing axis scales	30
9.2 Changing text	30
9.3 Changing the colour	30
9.4 Changing the size or style of the font	30

10	Summarising Data	31
10.1	Percentages	32
10.2	Descriptive statistics	33
11	Statistical Tests	35
11.1	T-test	35
11.2	Chi-squared test, Chi-squared test for trend and Fisher's Exact test	37
11.3	Mann-Whitney U test	39
11.4	Kaplan-Meier graph and Log-Rank test	41
11.5	ANOVA	45
11.6	Kruskal-Wallis test	46
12	Linear Regression	48
13	Logistic Regression	51
14	Survival Regression	54

Getting Started

- In the following sections the notes may sometimes just indicate which options to select.

Example If the notes indicate:

☞ *Graphs > Legacy Dialogs > Bar... > Clustered > Summaries for groups of cases > Define*

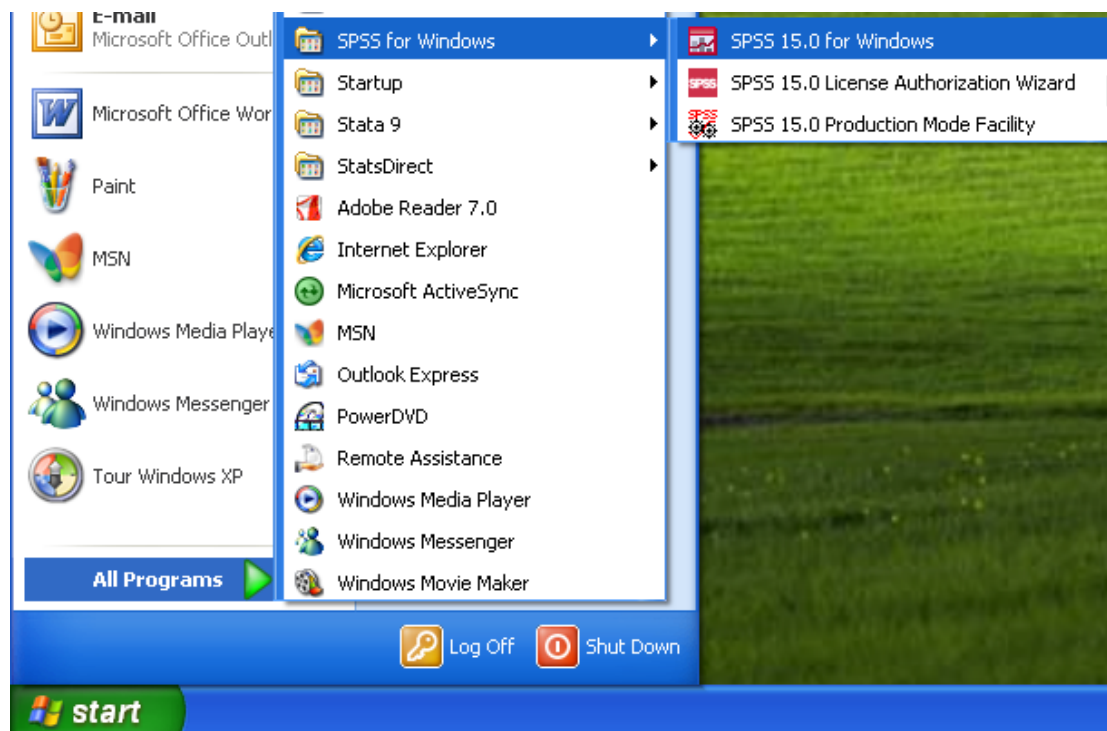
This would mean: Click on the *Graphs* menu, click on the *Legacy Dialogs* menu, select the *Bar...* option and when the above window appears, click on the *Clustered* graph and the *Summaries for groups of cases* option. Then press the *Define* button.

- *Legacy Dialogs* (which can be found in the *Graphs* menu) is an option available on the latest version of SPSS, which presents options as was commonplace in the older versions. These options are considered more user friendly, especially if the user has experienced previous versions of SPSS.
- *Interactive* (also found in the *Graphs* menu) presents options in the latest way.

1 Opening SPSS

To get into SPSS:

Either click:



Or:



Double click on the SPSS icon on your desktop

NB. An error message may appear the first time you open the package, just click OK.

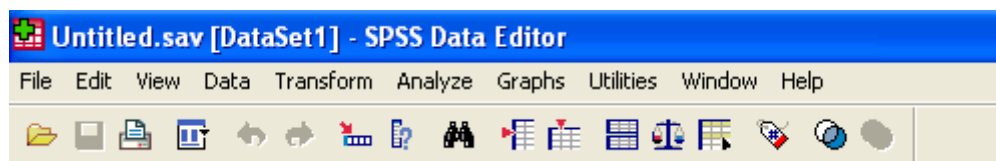
2 Layout of SPSS

2.1 Windows

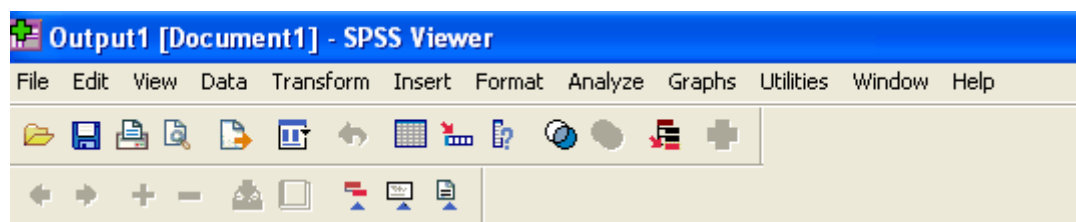
SPSS works through windows, when you first enter SPSS you'll enter the *Data Editor* window. When you produce graphs, summary statistics or statistical tests, they will be displayed in the output *Viewer* (or on older versions *Output Navigator*) window.

The menu headings in the two windows differ slightly, reflecting the different roles they take.

Data Editor Menu:



Output Viewer Menu:



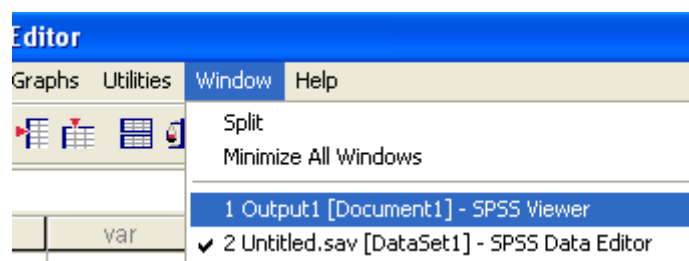
You can do the some of the same things from each window, but manipulation of data is done from the *Data Editor* window (using the menu headings *Data* and *Transform*) and manipulation of how the output looks from the *Viewer* window (using the menu headings *Insert* and *Format*).

The following table is a summary of which window to use for different procedures:

Procedure	Data Editor	Output Navigator
Data Entry	✓	
Data Manipulation	✓	
Computation or recoding variables	✓	
Production of graphs	✓	✓
Statistical Analyses	✓	✓
Editing graphs (or other output)		✓

The windows can be moved between as usual. Two ways are:

1) Using the menu heading *Window* and highlighting the window that you wish to move to. For example, if you're in the *Data Editor* window and want to move to the *Viewer* window:



2) Using the buttons on the task bar:



2.2 Saving files

Because the two windows contain two different files, they need to be saved separately:

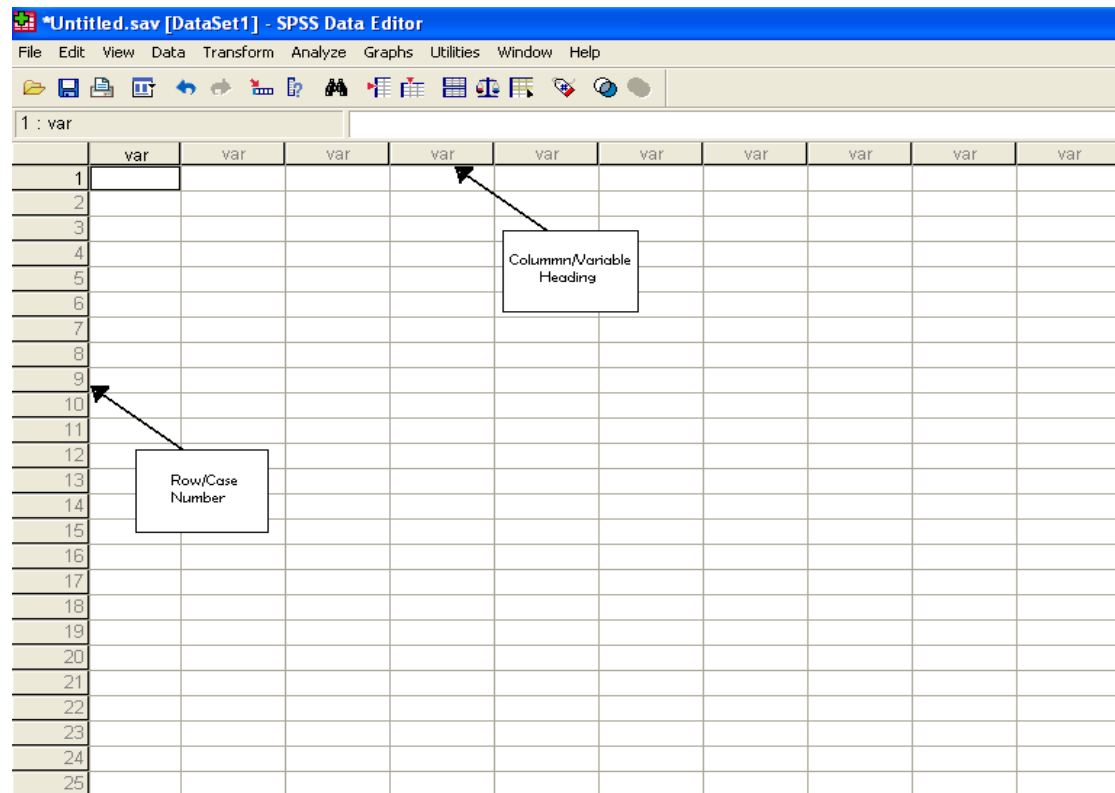
- *Data Editor* files are saved with a ".sav" file extension/ending
- *Viewer* files are saved with a ".spo" file extension/ending.

3 Creating a Data File

There are two parts to *Data Editor* – these are *Data View* and *Variable View*.

3.1 *Data View*

A new data sheet will look like this:



This is the *Data View* part to *Data Editor*. A typical way to enter data into *Data View* is to have one row per subject and one column per variable.

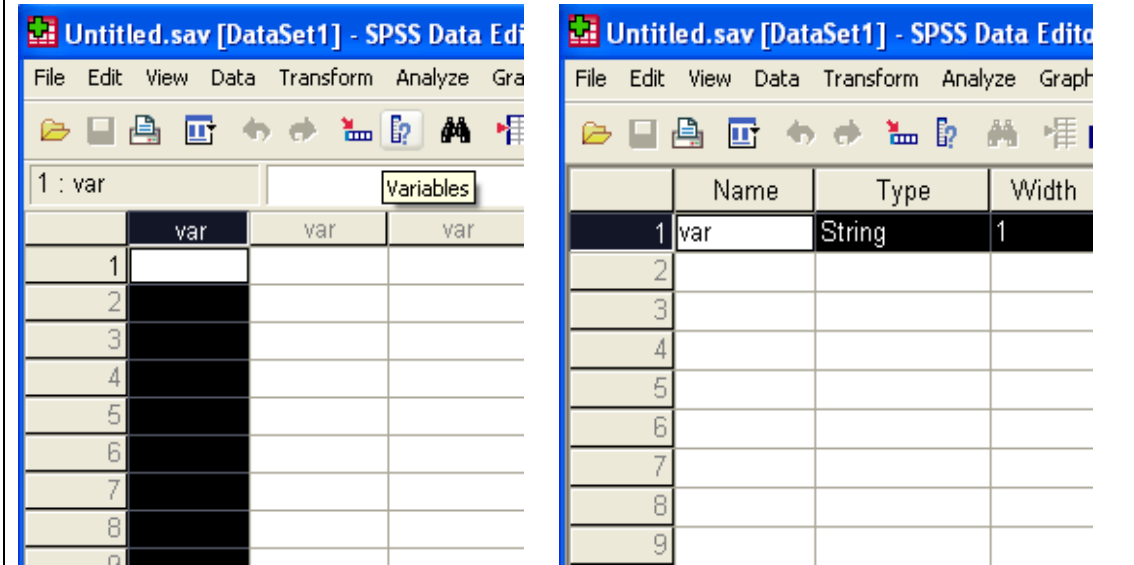
If you click on a grey column heading the whole column is selected/highlighted, and similarly if you click on a grey row number the whole row is selected/highlighted.

3.2 Variable View (defining variables)

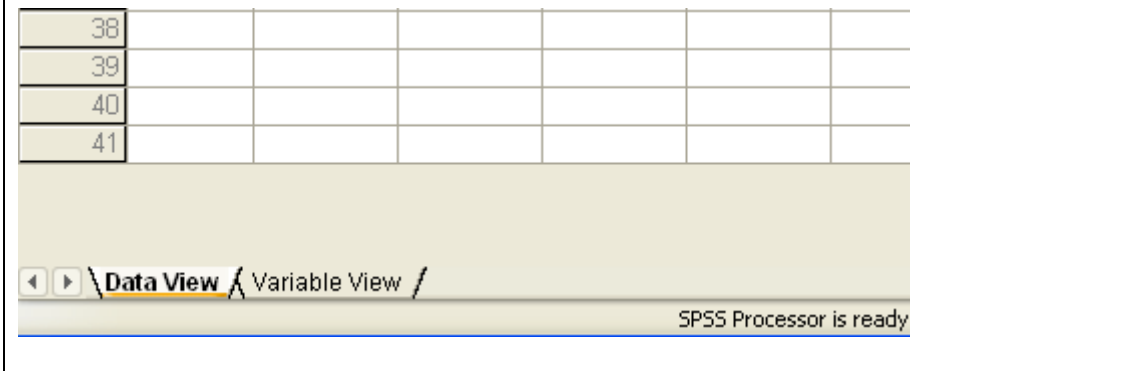
The characteristics of a variable may be assigned through *Variable View*.

You may navigate between *Data View* and *Variable View* in either of two ways:

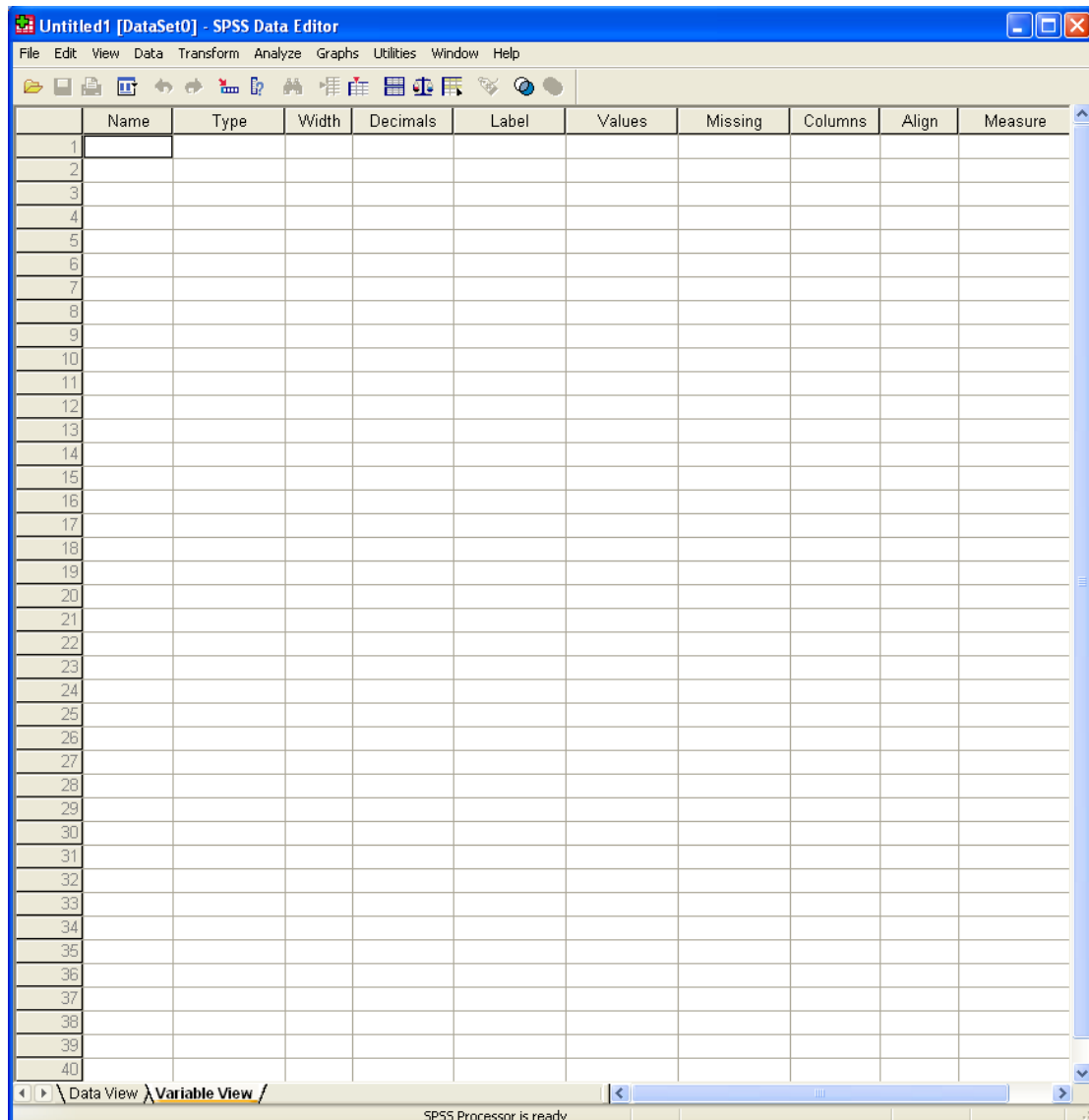
- 1) Double-clicking on a column heading to get from *Data View* to *Variable View*, and a row number to get from *Variable View* to *Data View*.



- 2) Using the tabs at the bottom of the *Data Editor* window:



When you navigate from *Data View* to *Variable View* the following screen will appear:



Each row contains information on a different variable and the variable attributes may be described under the different column headings.


3.2.1 Name

Here you may type in a suitable name for the variable and press ENTER to have it assigned. Names must start with a letter, no spaces are allowed, and names should not finish with a full stop or underscore.

As soon as a name is assigned to a variable, the default options are also allocated, such as *Type*. To change a default characteristic click on the appropriate cell for the corresponding variable and characteristic, and scroll through the options.

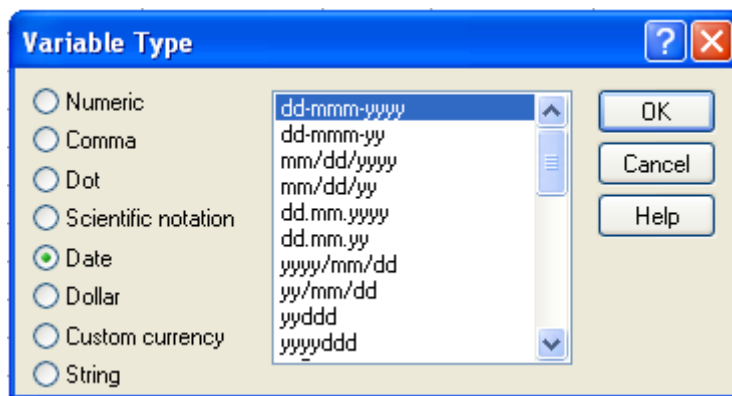
3.2.2 *Type*

The *Type* setting depends on the data to be entered, the default being *Number* with two decimal places being displayed – although you may want to change the number of decimal places if for example the variable is a count, in which case you would change the it to zero.

By selecting the cell and clicking on the  button on the right hand side, other options will be displayed. Some of the most useful are *Date* and *String*.

➤ *Date* (this also includes options for timings)

The date formats that are displayed on the right hand side just indicate how the date will be displayed, dd standing for day, mmm standing for three letter abbreviation of month, mm standing for two digit month, etc. You should always use yyyy for year. Time formats may be chosen in the same way.



➤ *String*

If the column is going to contain letters, click the string option – if this variable will contain a comment, you may want to increase the length from 8.

Note: It is advisable to avoid using *String* - it is better to enter a numeric code and then add a value label (see section 3.2.5).

3.2.3 *Width* and *Decimals*

Both can be set when you are defining the type. But you may also change them under these headings.



3.2.4 Label

Because the variable name is usually short and cannot contain spaces you will often have to abbreviate it somehow. This means that it may not be clear to you, or someone else using the file later what the variable actually is. However, you can give the variable a label which may be up to 256 characters long and contain spaces. This will be displayed in any output that is produced later.

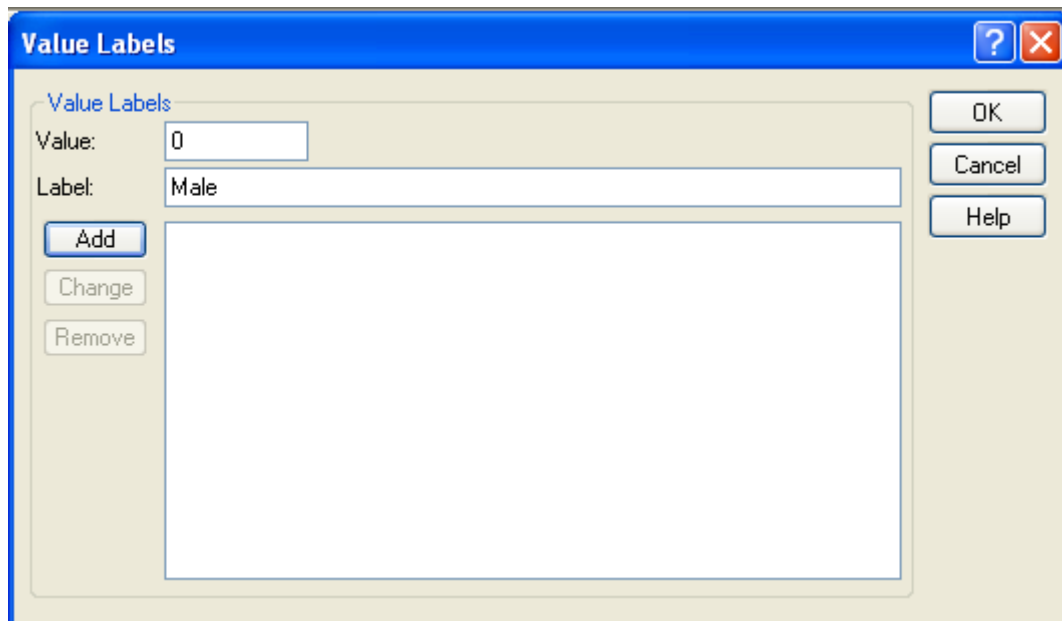
3.2.5 Values

It is good practice to enter all data as numbers even if some are code for something else. For example, for a variable indicating the sex of a subject, you can enter a '0' to indicate "Male" and a '1' for "Female". You could then attach a description for each level of the variable.

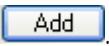
To do this:

- ✓ Click on the *Values* cell and 
- ✓ '0' in *Value*
- ✓  'Male' in *Value Label*

The screen should look as below:



- ✓ Click 

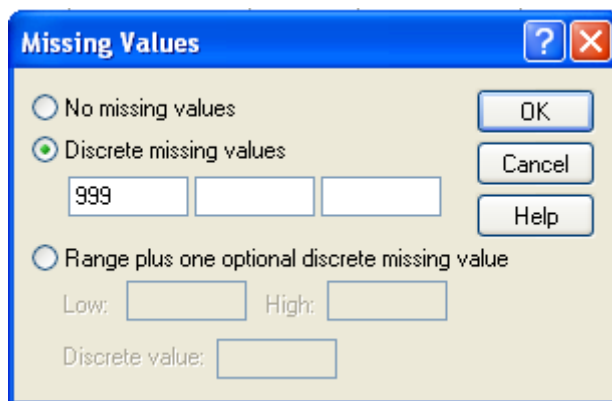
Repeat this with a '1' in *Value* and 'Female' in *Value Label*, and remember to click .

3.2.6 Missing Values

If a response is missing, rather than leave a cell blank it is good practice to enter another value such as '999', so that you know that the value is actually missing, and that a data entry error has not occurred. You must tell SPSS what your missing value code is.

☞ Missing Value cell and  button.

For example, to define '999' as a missing value, you type in the value as shown below. The default is no missing values. SPSS allows multiple missing value codes so that they are easy to distinguish - for example, using '888' and '999' to distinguish between "lost-to-follow-up" and "died".



3.2.7 Column, Align

These options allow you to specify how wide the column will be and the alignment of its contents.

3.2.8 Measure

Leave as default.

3.3 Applying attributes to multiple variables

Once you've defined *Type*, *Values*, *Label* or *Missing Values*, you can copy these attributes to other variables. This saves going through the same procedure with many similar variables, for example, with responses to a questionnaire.



To do this, make sure you're in the *Variable View* part of *Data Editor*:

- ☞ Click in the cell with the attribute you wish to copy
- ☞ *Edit>Copy* (or right-click *Copy*)

- ☞ Click in the cell(s) you wish to apply the attribute to
- ☞ *Edit>Paste*


3.4 Entering data

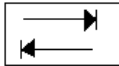
To enter data, make sure you are in the *Data View* part of *Data Editor*.

-  Click the cell you wish to start entering data into
-  Type the value

It's helpful to use the numbers on the right hand side of your keyboard; if the *Num Lock* light



is not lit on your keyboard, press the  key.

Press ENTER to move down to the cell below or the TAB key  to move across to the cell on the right. The arrow keys also work.



To correct a mis-entered value, click in the cell. The entry that is currently in the cell will be displayed in the white box above the grey column headings. You may either amend the value in this white box, or enter the new value directly into the cell itself; the old value will automatically be deleted.

4 Data Manipulation

4.1 Editing



4.1.1 *Cut*

Deletes a column, row or cell(s), making the contents available to paste

-  Select a column, row or cell(s)
-  *Edit > Cut* (or right-click '*Cut*')



4.1.2 *Copy*

Copies a column, row or cell(s)

-  Select a column, row or cell(s)
-  *Edit > Copy* (or right-click '*Copy*')

4.1.3 *Paste*

Pastes a selection that has been copied or cut (will start in top left-hand corner of selection)

-  Select (empty) column, row or cell(s). (If not empty, will replace current values).
-  *Edit > Paste* (or right-click '*Paste*')

4.1.4 Clear

Deletes a column, row or cell – no copy is made

- ☞ Select column, row or cell(s)
- ☞ *Edit > Clear*



4.1.5 Insert

Inserts a row or column

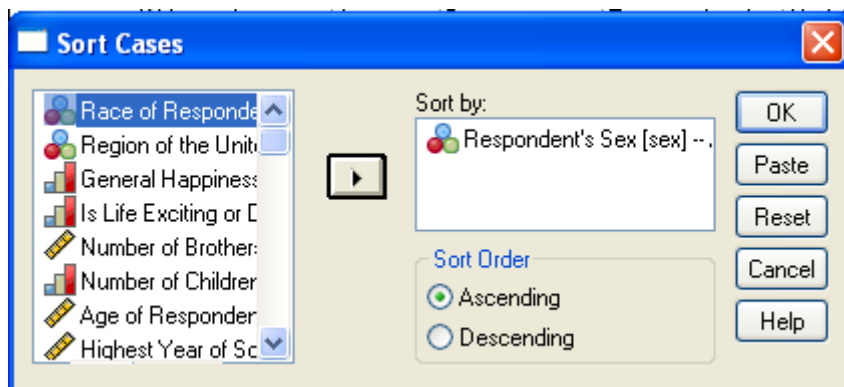
- ☞ Select a column or row
- ☞ *Data > Insert Variable or Insert Case*

4.2 Sorting

The dataset can be sorted according to the order of one or more variables. *Ascending* sorts from smallest to largest and *Descending* from largest to smallest.

- ☞ *Data > Sort Cases...*
- ☞ In the *Sort Cases* box, click the variable to sort by and use  to move it into the *Sort by:* section. You may make as many variables ascending or descending as you wish, simply select extra variables.
- ☞ Under *Sort Order* click *Descending* if you wish this variable to be so, otherwise simply leave at the default setting of *Ascending*.
- ☞ Click 

The *Sort Cases* box should look like this:







4.3 Obtaining a file summary

To display a summary of the variables within the file

- ☞ *Utilities > Variables...*




4.4 Searching through the contents of a variable

To search for a word, part of word, or number that has been entered into a variable

-  Select column
-  *Edit > Find...*
-  Type characters to be searched for in the *Find what* box
-  Click



4.5 Go to Case

To move to a specific row

-  *Edit > Go to Case...*
-  Type in the row number you wish to go to in the *Case Number:* box.
-  Click

4.6 Go to Variable

To move to a specific variable

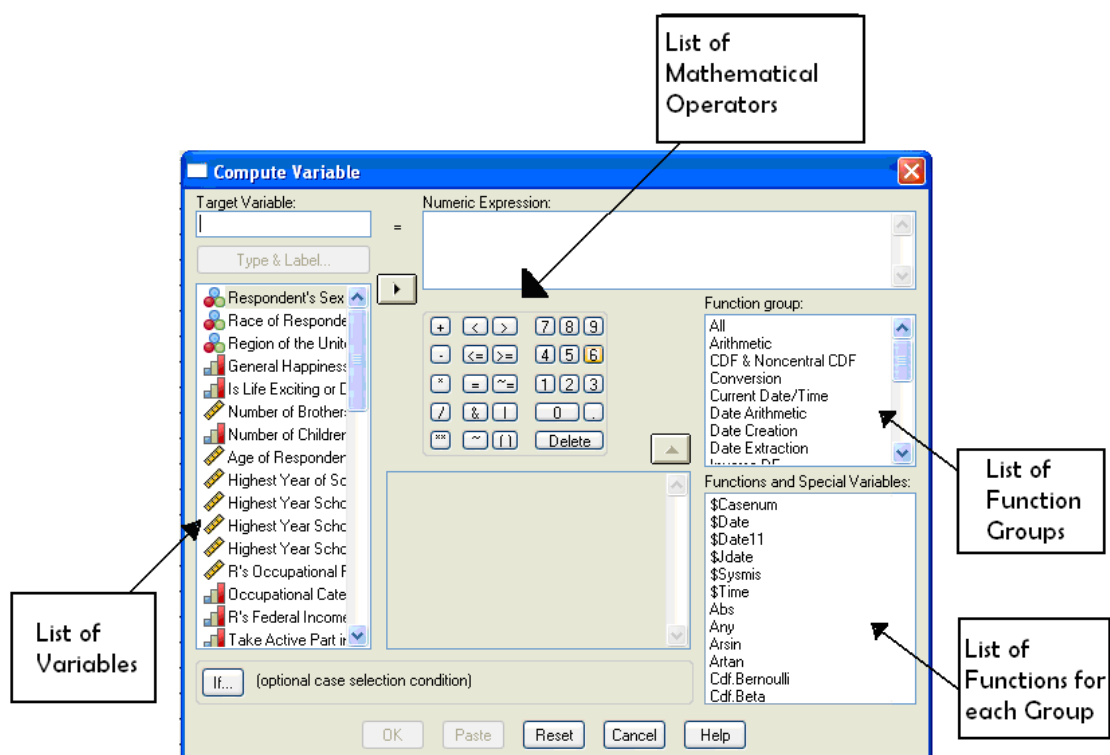
-  *Utilities > Variables*
-  Choose the variable from list and click

5 Computing and Recoding

5.1 Compute


New variables can be created from existing ones through *Compute Variable*.

☞ *Transform > Compute Variable...*




☞ Type the name for the new variable in the *Target Variable* box.

All existing variables are listed on the left hand side, a list of functions on the right hand side and the standard mathematical operators in the middle (see above). These can all be used to create a formula for the new variable.

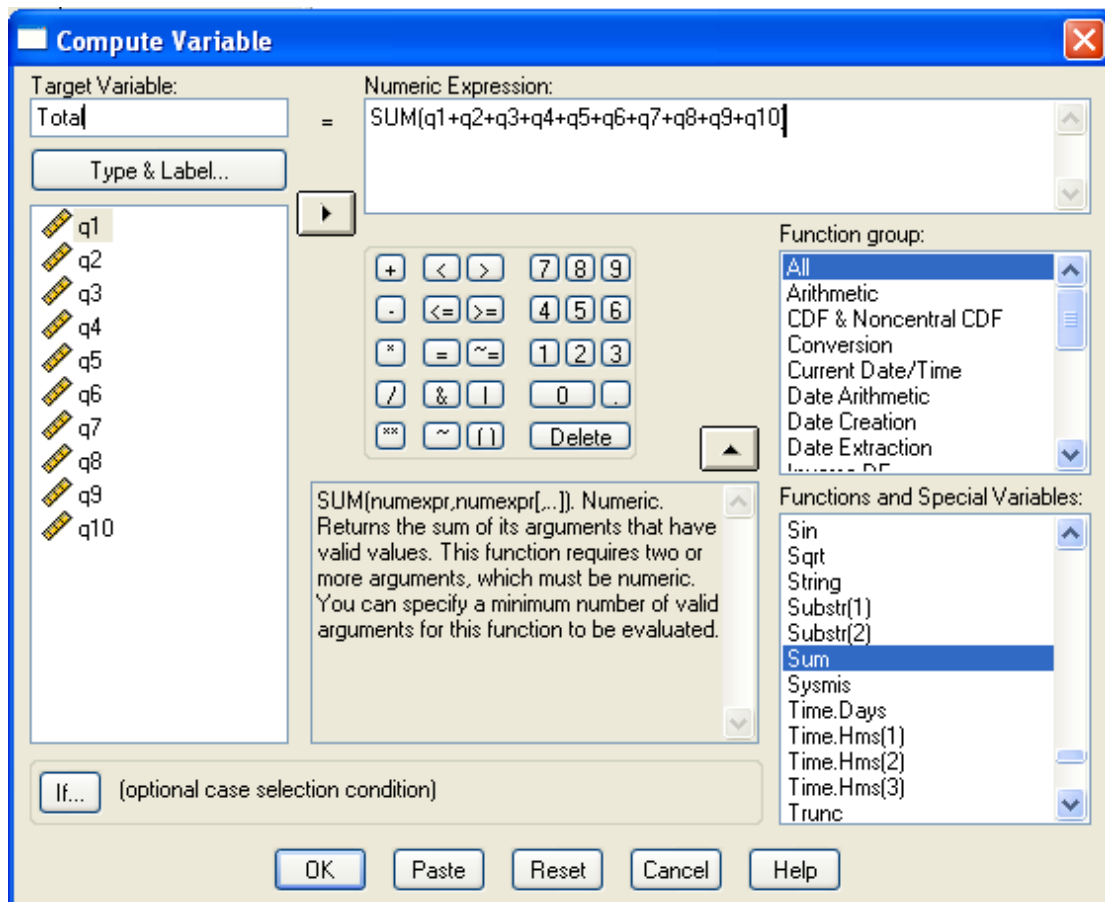
Scroll down the list until you find the function that you wish to use, and use  to place it in the *Numeric Expression* section. Similarly, you can create formulae using combinations of existing variables.

Example 1

Creating a variable that contains the total number of positive responses in a ten-question questionnaire, for each case. For each question a positive response was stored as '1' and a negative as '0'.

To create a sum of the variables 'q1', 'q2', 'q3', 'q4', 'q5', 'q6', 'q7', 'q8', 'q9' and 'q10' you can use , the list of variables and list of functions as below:

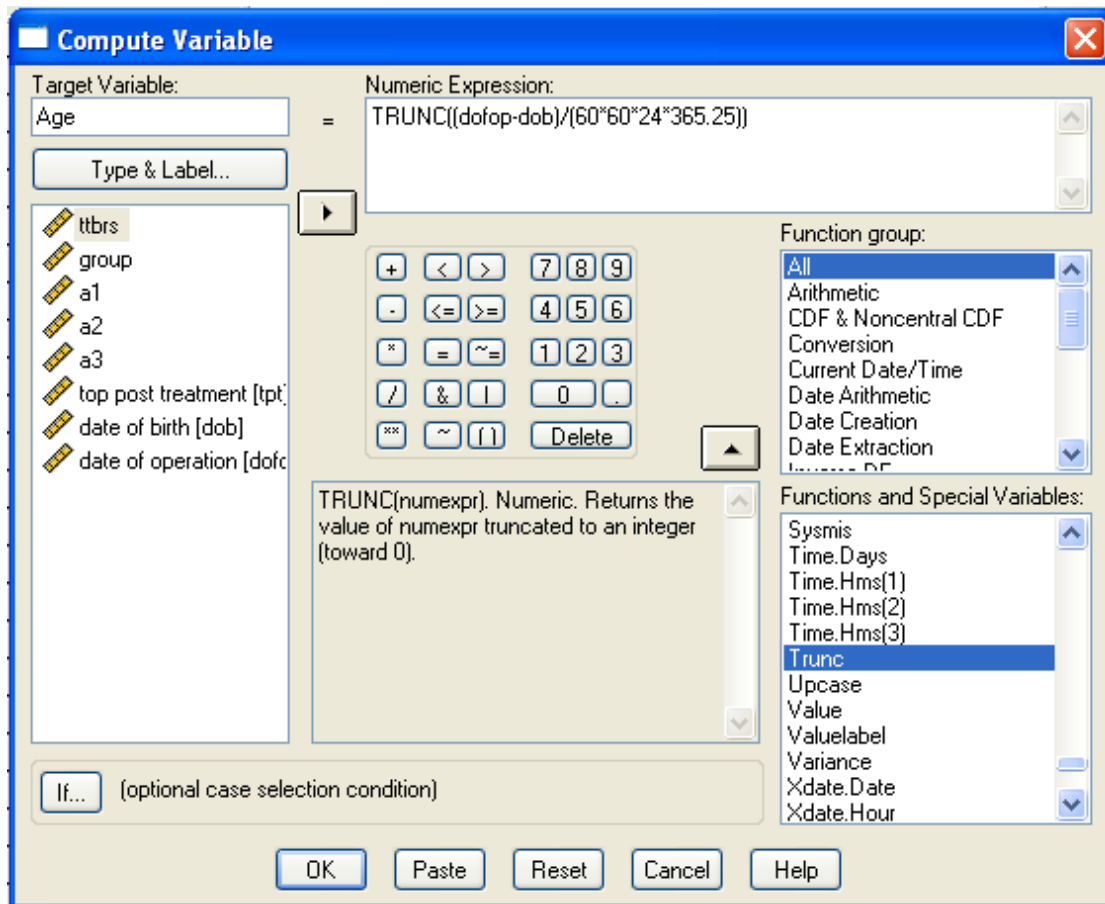
SPSS notes



Example 2

Creating an age variable from date of birth and date of operation. Here date of operation is called '*dofop*' and date of birth is called '*dob*'.

SPSS saves dates as the number of seconds since 14th October 1582 so when you do a computation on dates, the result is in seconds. To present age in years use the formula as shown below (the *trunc* function truncates (shortens) a number by ignoring decimal places).



Note that the expression on the previous page is equivalent to using the expression 'Datediff(dofop,dob,"years")'. The date-difference function must always be entered in the form: DATEDIFF(datetime2, datetime1, "unit"). This should only be used when you want to truncate the answer (that is in this case, when you want the number of complete year's difference) and not when the answer should be given to several decimal places or rounded to the nearest whole number.

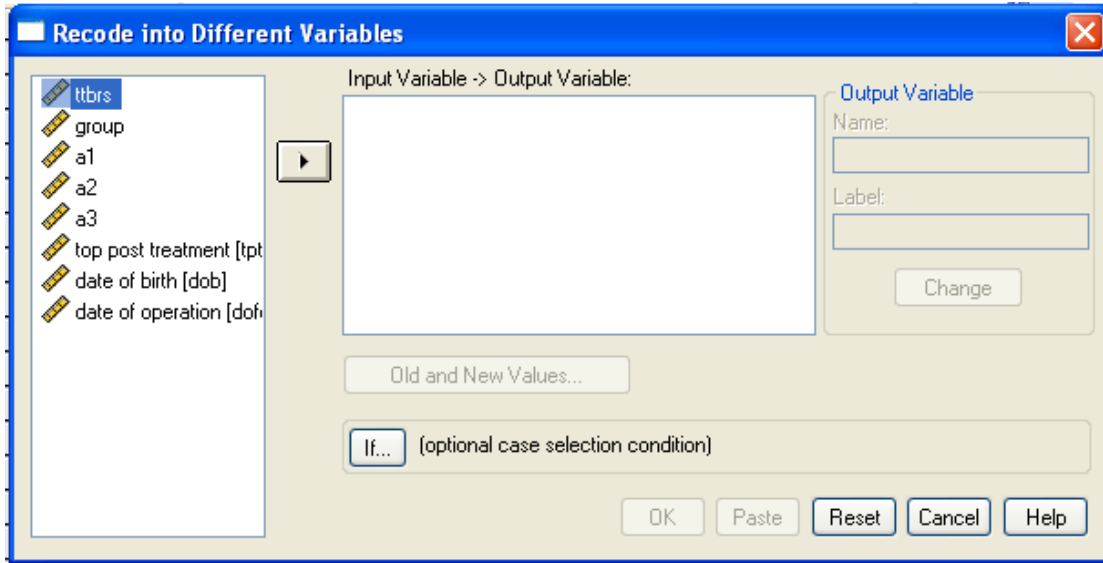
5.2 Recode




Existing variables can be recoded or re-classified by:

 *Transform > Recode Into Different Variables...*

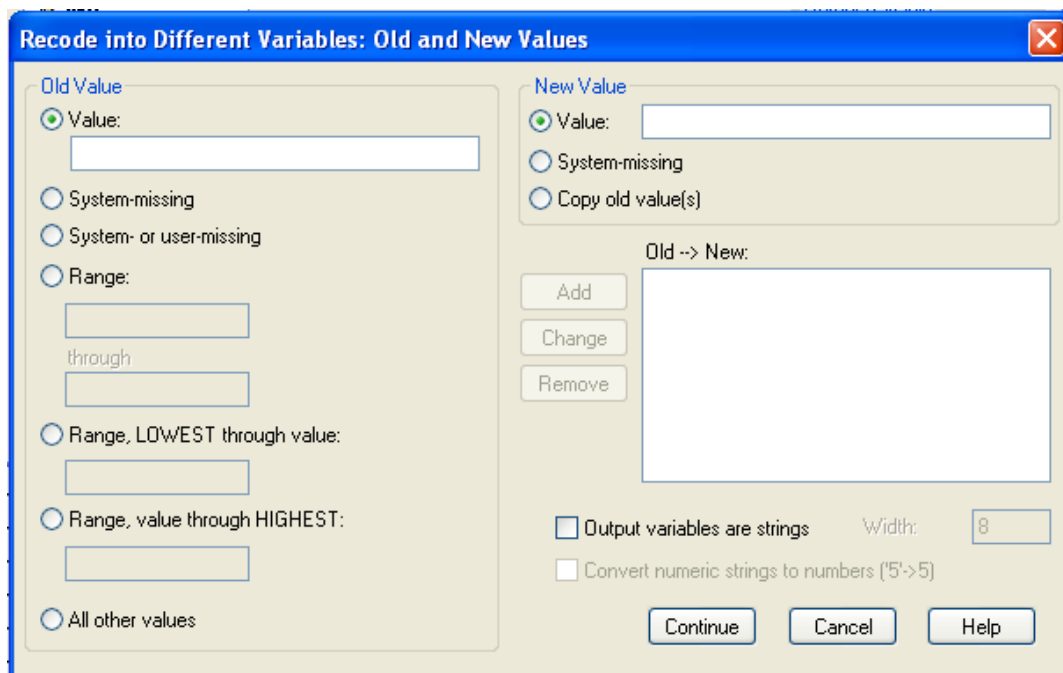
(Never use the *Transform > Recode Into Same Variables...* option because you will lose your data if you make a mistake)


A new window will appear:



- ☞ Put the name of the variable that you wish to recode into the *Input Variable -> Output Variable* box, using 
- ☞ Type the name for the new variable into the *Name* box
- ☞ Click 
- ☞ Click 

The following window will appear:



 Enter the how the old value is to be recoded, by defining what the old value is, and what you would like it to be recoded as (see below)

 Click 

Repeat this until all 'old values' have been defined.

 Click  and 

Old Value

The options on the left hand side are used to define which values in the existing variable are going to be recoded:

- *Value*: to recode a single value
- *System-missing* to recode an empty cell
- *System- or user-missing* to recode either an empty cell or a defined missing value
- *Range* through to recode a range of values between (value 1) and (value 2)
- *Range: LOWEST through value*: to recode anything less than (value)
- *Range: value through HIGHEST*: to recode anything greater than (value)
- *All other values* to recode anything else that has not already been specified

New Value

The options on the right hand side define what the old value is to be recoded to:

- *Value* the new recoded value that the new variable will take
- *System-missing* to recode as a system-missing value
- *Copy old value(s)* to recode to the same value as before

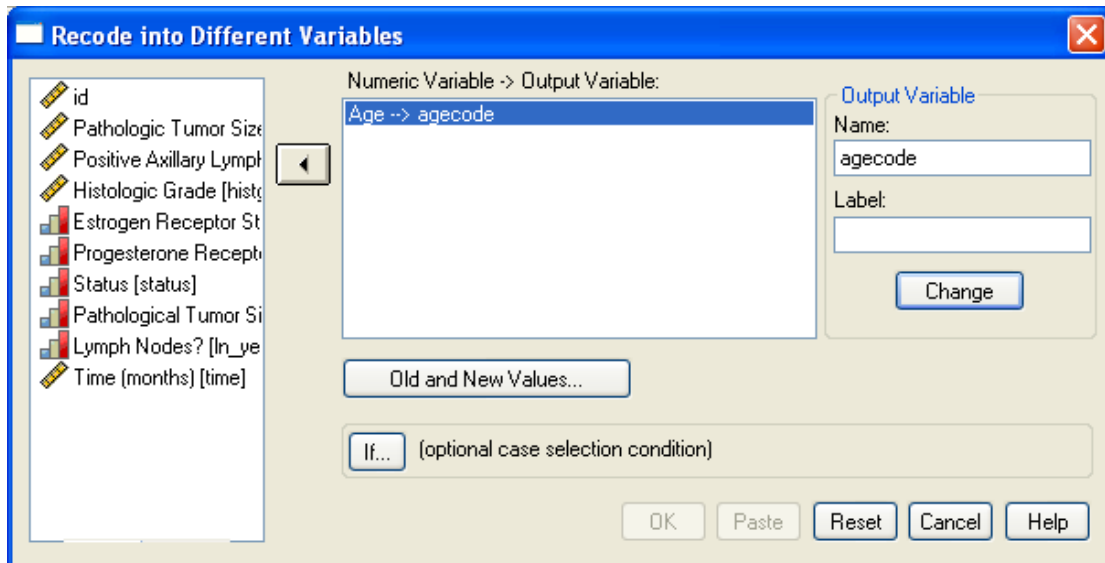
Example: Creating a new variable to distinguish under 50 years old from over 50 years old.

Age is the variable to be recoded.

 Using  put *Age* variable into *Input Variable -> Output Variable:* box

 Type 'agecode' into the *Name:* box

 Click  so that it appears in the middle box as shown below.



Click

In the *Old Value* side:

Click *Range: LOWEST through value:*

Type 49 into the box.

Under *New Value* make sure that *Value:* is checked.

Type '1' into the box.

Click . This coding will then be displayed in the *Old -> New:* box.

Now just want to code any other age as '2':

Under the *Old Value* side, click *All other values.*

Under *New Value* make sure that *Value:* is checked.

Type '2' into the box.

Click

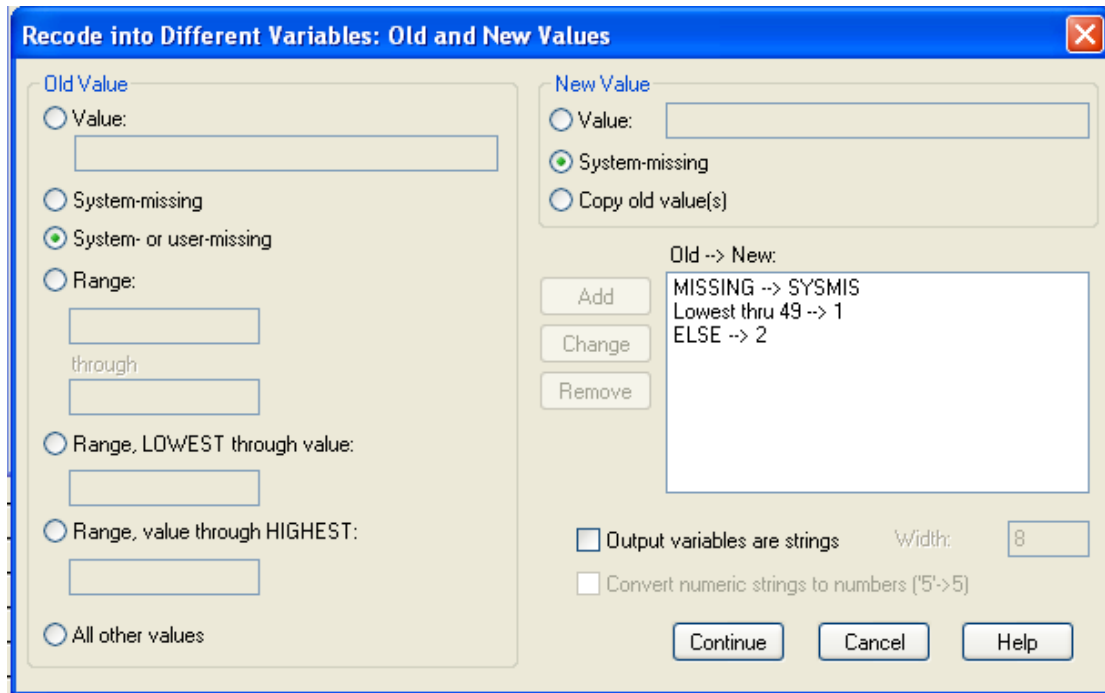
To make sure that missing values stay as missing values:

Under the *Old Value* side, click *System- or user-missing.*

Under the *New Value* side, click *System-missing.*

Click

The screen should look as below:



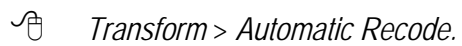

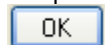
Click 

Beware – this works if *Age* is recorded as whole numbers and there are no invalid ages recorded. Just in case, could use:

Old Value	New Value
Range 0 through 49.99	1
Range 50 through 130	2
All other values	System-missing

5.3 Automatic Recode

You can recode an existing string variable into a numeric code using Automatic recode.

- Click 
- Put the variable to be recoded in to the *Variable -> New Name:* box.
- Type in the smaller box *New Name:*
- Click 
- Click an option under *Recoding Starting from:* (either *Lowest value* or *Highest value*).
- Click .

6 Opening an Existing File

6.1 Opening a SPSS file

To open a SPSS data file,

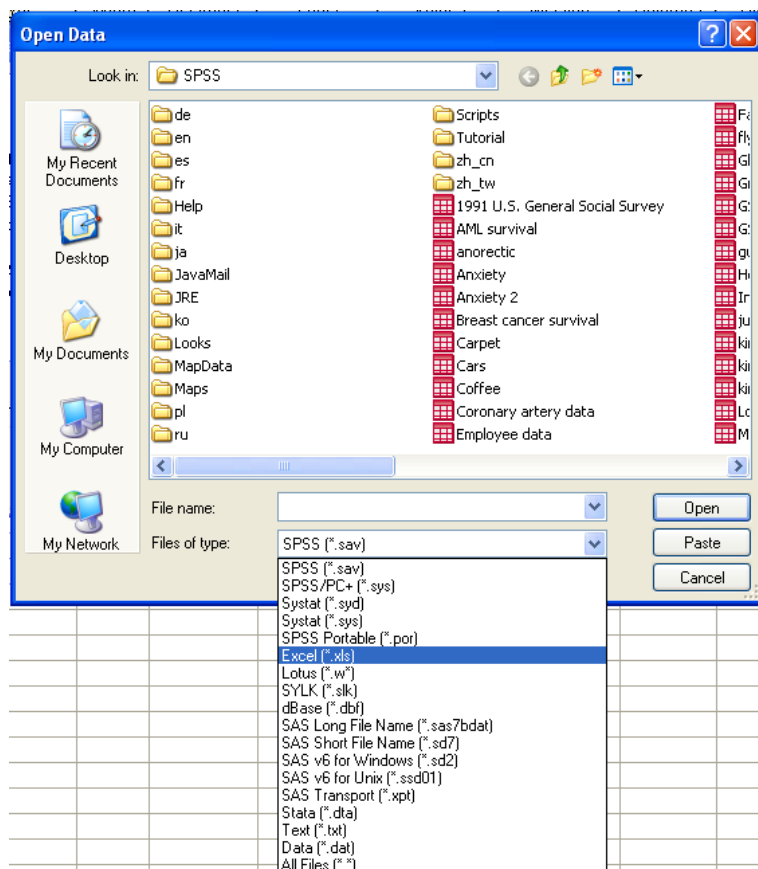
- ☞ *File > Open > Data...*
- ☞ Choose 'SPSS (*.sav)' under *Files of type:*
- ☞ Click on the desired file.
- ☞ Click

To open a SPSS output file,

- ☞ *File > Open > Output...*
- ☞ Choose 'Viewer document (*.spo)' under *Files of type:*
- ☞ Click on the desired file.
- ☞ Click

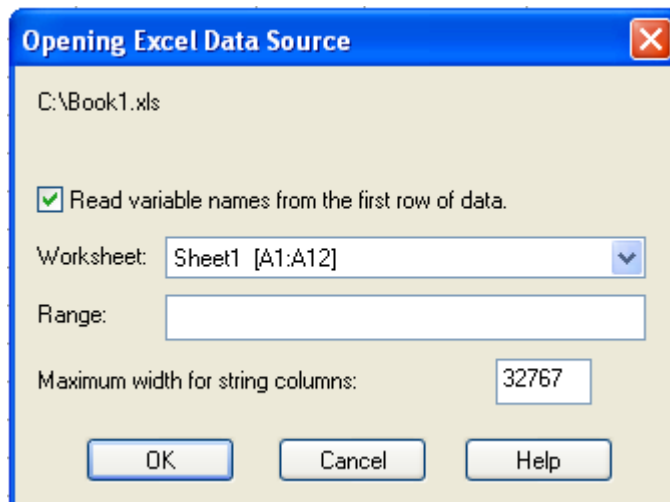
6.2 Opening an Excel file

- ☞ To open an Excel workbook it is the same as opening an SPSS data file in 6.1, but under to *Files of type:* click on 'Excel (*.xls)' as below:



- ☞ Select the Excel workbook you wish to open
- ☞ Click

A screen similar to below will open:



- ☞ If the first row of the excel workbook contains the variable names then make sure that this option is ticked.
- ☞ If the data are not on the first worksheet change the worksheet using .
- ☞ If you do not want all the data from the worksheet select the range that you do want.
- ☞ Click

NB: 1) You cannot open the file in SPSS if it is already open in Excel.
 2) If you get an error message that the file is not an Excel file, try saving your Excel file (using Excel) as an earlier version and then returning to SPSS.

7 Split File and Select Cases

7.1 Split File

To display statistical or graphical procedures separately for subgroups of data:

- ☞ *Data > Split file...*
- ☞ Click 'Compare groups'
- ☞ Put variable that defines the desired subgroups in the box under *Groups based on*
- ☞ Click

Notice that under the box, *Sort the file by grouping variables* is selected. When this is switched on, any output will be produced for each group separately.

To switch off this 'sort file' option:

- ☞ *Data > Split file...*
- ☞ Select *Analyze all cases, do not create groups.*
- ☞ Click

Note: It is unlikely that you will ever need to use the 'sort file' procedure.

7.2 Select Cases

This displays output only for specified cases:

- ☞ *Data > Select Cases...*
- ☞ Click *If condition is satisfied*
- ☞ Click
- ☞ Use the variable list, mathematical operators and *Functions*: to define criteria for selecting cases, e.g. Age > 50 to analyse only those subjects which are older than 50.
- ☞ Click
- ☞ Click

Notice that at the bottom of the window it says 'Current Status: Do not filter cases'.

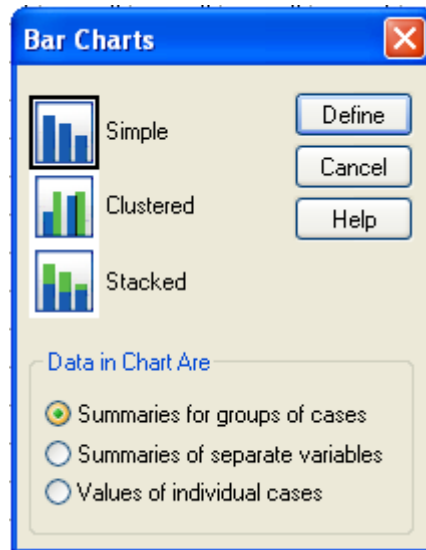
To switch off the select cases:

- ☞ *Data > Select Cases...*
- ☞ Click *All cases*
- ☞ Click

8 Displaying Data

General notation

There are many options available in SPSS when creating graphs. For example, after selecting *Graphs > Legacy Dialogs > Bar...* the following screen will appear:



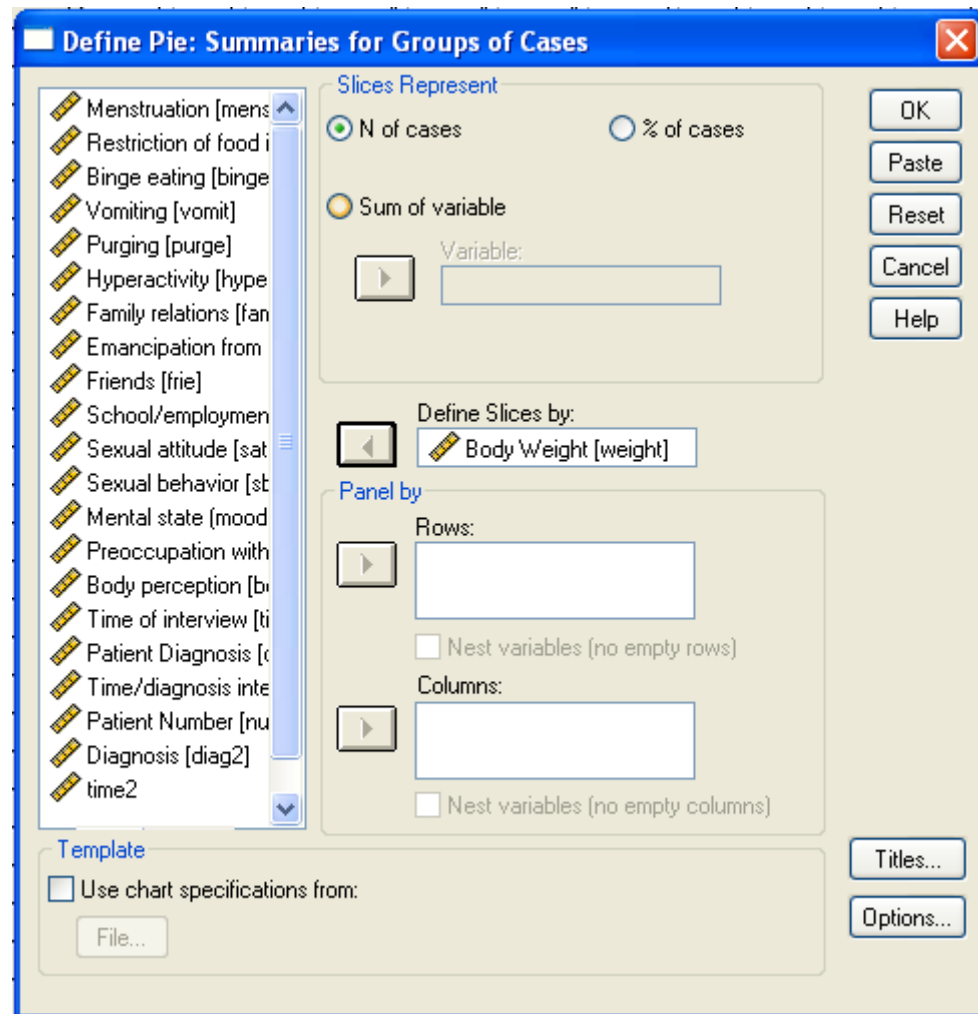
Simple, *Clustered* or *Stacked* simply define the style of the graphs and the *Data in Chart Are* options define the type of data collected and how they were entered. Instructions will be given for the *Legacy Dialogs* route rather than *Interactive* as they are considered more user-friendly.

8.1 Pie Charts

Graphs > Legacy Dialogs > Pie... > Summary for groups of cases > Define...

Select the variable that categorises the sample and put into *Define Slices by* box. Leave the *Slices Represent* option as *N of cases*. Click *OK*.

Example When creating a pie chart to show how many patients were in each 'Body Weight' classification, the window should look like this:

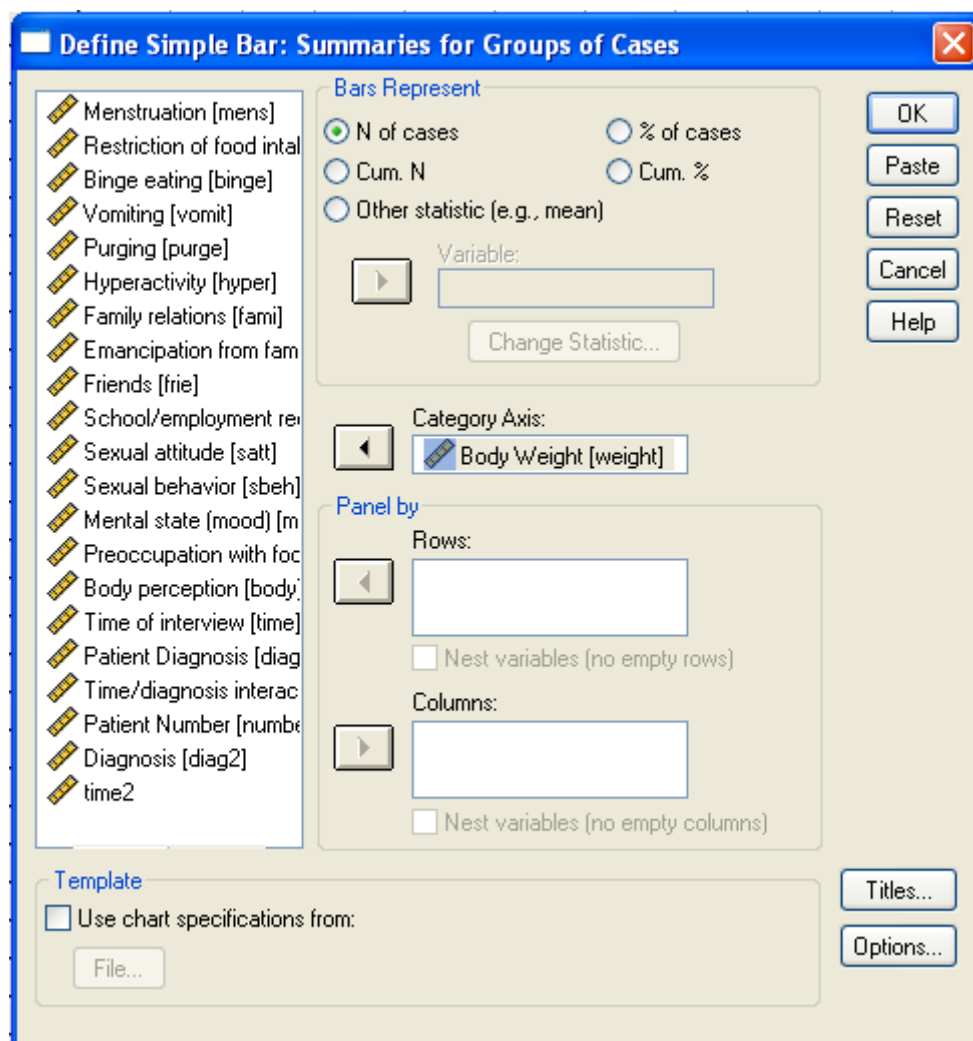


8.2 Bar Charts

8.2.1 To show the breakdown of the sample:

- ☞ *Graphs > Legacy Dialogs > Bar... > Simple > Summaries for groups of cases > Define.*
- ☞ Put variable into *Category Axis* box.
- ☞ Under *Bars Represent* leave as *N of cases*

Example When creating a bar chart to show many patients were in each 'Body Weight' classification, the window should look like this:



8.2.2 Comparing an outcome between groups using stacked bar charts:

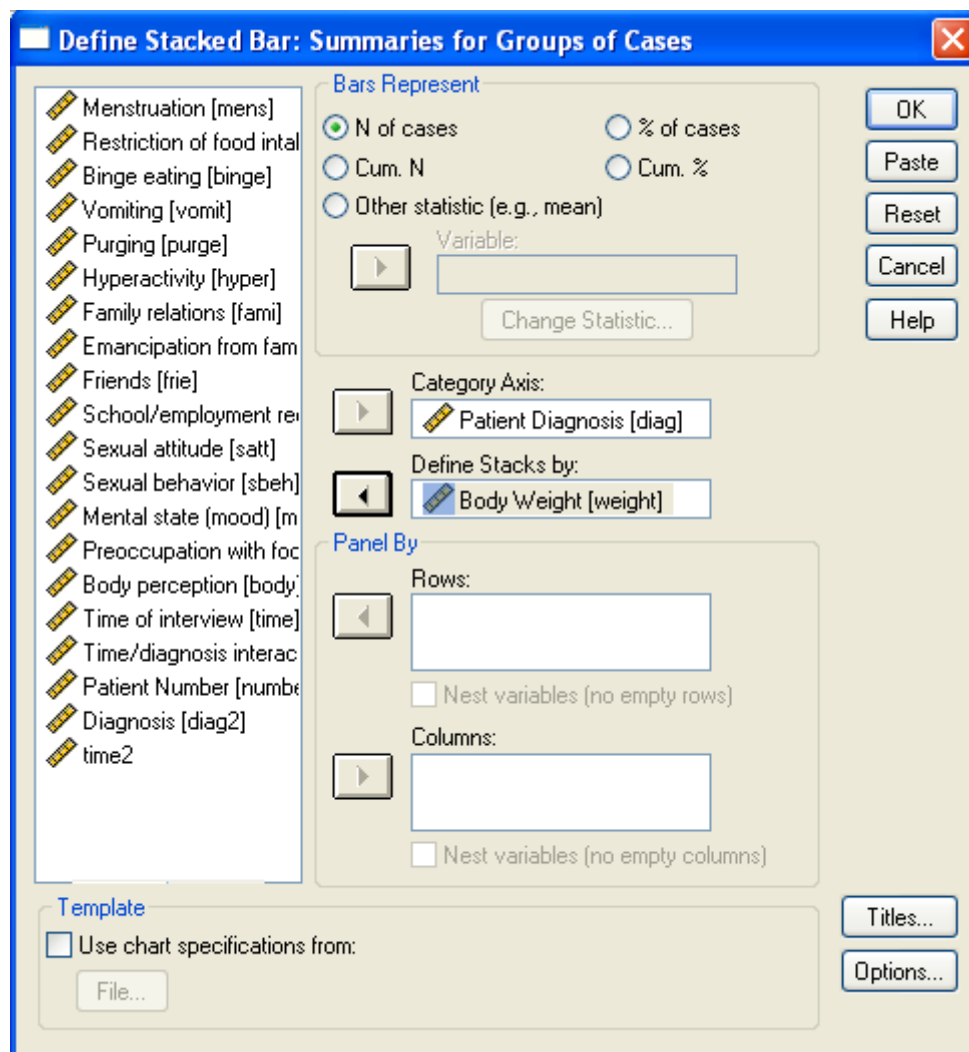
- ☞ *Graphs > Legacy Dialogs > Bar... > Stacked > Summaries for groups of cases > Define.*
- ☞ Put the grouping variable in the *Category Axis:* box
- ☞ Put the outcome in the *Define Stacks by:* box.
- ☞ Leave the *Bars Represent* option as *N of cases*.

☞ Click *OK*.

To show the bars as percentages:

- ☞ In the Output window, double click on the graph. Another window will appear the SPSS *Chart Editor*.
- ☞ Double click on one of the bars of the bar chart (the bars should be highlighted). A *Properties* window will appear.
- ☞ *Bar options > Scale to 100% > Apply*
- ☞ Close the *Chart Editor* window.

Example When creating a stacked bar chart to compare the proportion of patients in each 'Body Weight' classification between groups of patients with different diagnoses of eating disorder (categorised 1,2,3,4) the window should appear like the following:



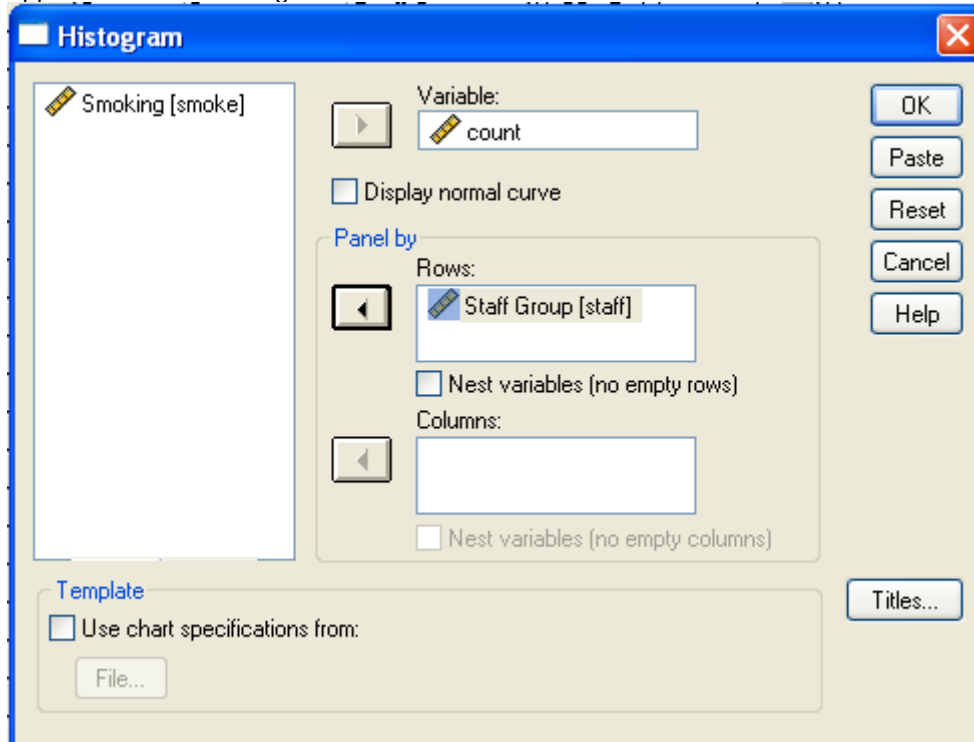
If you want to show a "100% bar", edit the graph in the output window as detailed previously.

8.3 Histograms

🔗 *Graphs > Legacy Dialogs > Histogram...*

Put the variable you wish to summarise into the *Variable:* box. To compare groups place the grouping variable in the *Panel by Rows* box.

Example: When creating a histogram to compare smoking counts between groups of staff at different responsibility levels (e.g. Senior Manager = 1, Junior Manager = 2), the window should appear like the following:



8.4 Box Plots (or “Box and Whisker” Plots)

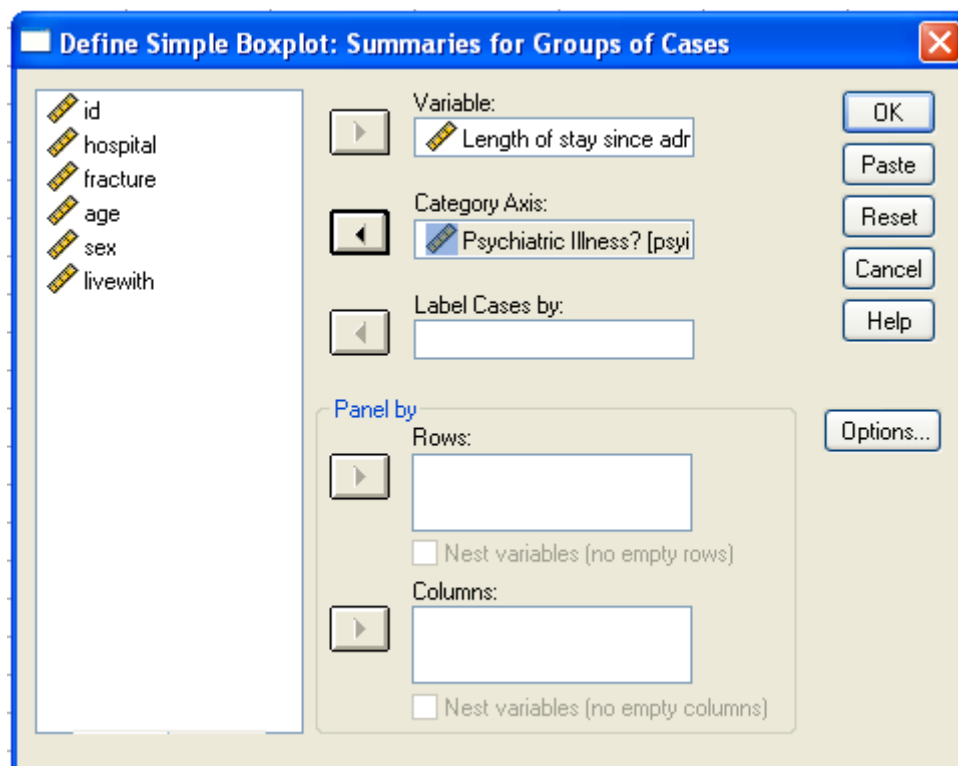
8.4.1 Single group:

- ☞ *Graphs > Legacy Dialogs > Boxplot... > Simple > Summaries of separate variables > Define*
- ☞ Put the desired variable into the *Boxes Represent:* box
- ☞ Click *OK*.

8.4.2 Comparing groups:

- ☞ *Graphs > Legacy Dialogs > Boxplot... > Simple > Summaries for groups of cases > Define*
- ☞ Put the grouping variable into the *Category Axis:* box
- ☞ Put variable to summarise into the *Variable:* Box
- ☞ Click *OK*.

Example When creating a box plot to compare length of hospital stay between groups of patients with and without psychiatric illness the window should look like this:



8.4.3 Comparing repeated measures:

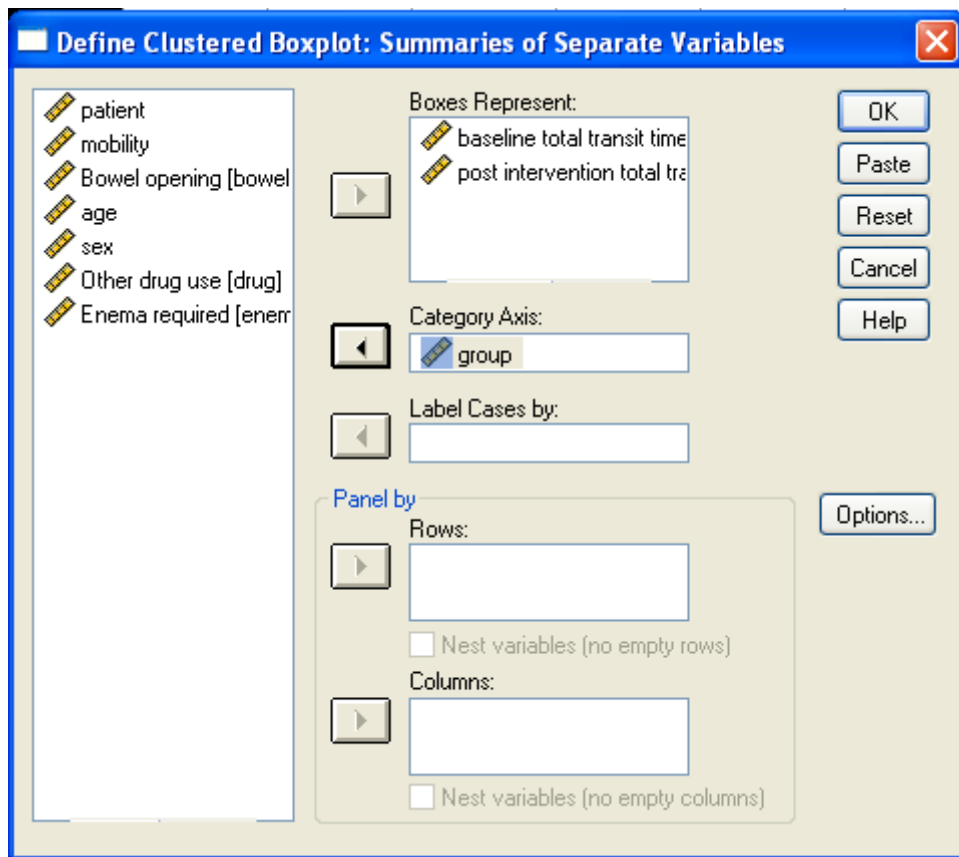
- ☞ *Graphs > Legacy Dialogs... > Boxplot > Simple > Summaries for separate variables > Define*
- ☞ Put the repeated measures into the *Boxes Represent* box

☞ Click *OK*

8.4.4 Comparing repeated measures between groups:

- ☞ *Graphs > Legacy Dialogs... > Boxplot > Clustered > Summaries for separate variables > Define*
- ☞ Put the repeated measures into the *Boxes Represent:* box
- ☞ Put the grouping variable into the *Category Axis:* box
- ☞ Click *OK*

Example When creating a box plot to compare the transit times (at baseline and post treatment) of patients between groups the window should look like this:



9 Editing Graphs

Graphs will not automatically be produced in the most appropriate way. Once the graph has been produced and displayed in the Output Window we may want to edit it in. For example, in 8.2.2 the graph was edited to show 100% bars.

Generally in order to edit graphs enter the chart editor by:

- ☞ Double-clicking on the graph you want to edit

The following sections assume you are in the chart editor.

9.1 Changing Axis Scales

In the Chart Editor window:

- ☞ Double-click on the axis you wish to edit (the properties window should appear)
- ☞ Choose *Scale* – from here you can edit the max value, min value and increment size.

9.2 Changing the text

- ☞ Click twice (slowly) on the text you wish to edit. The cursor should appear in the text to enable you to edit.

9.3 Changing the colour

- ☞ On the legend of the graph double click on the colour you wish to change
- ☞ Choose *Fill and border*
- ☞ Select the colour you want
- ☞ Click *Apply*

9.4 Changing the size or style of the font

- ☞ Double click on the text you wish to format
- ☞ Choose *Text Style*
- ☞ Alter format
- ☞ Click *Apply*

10 Summarising Data

10.1 Percentages

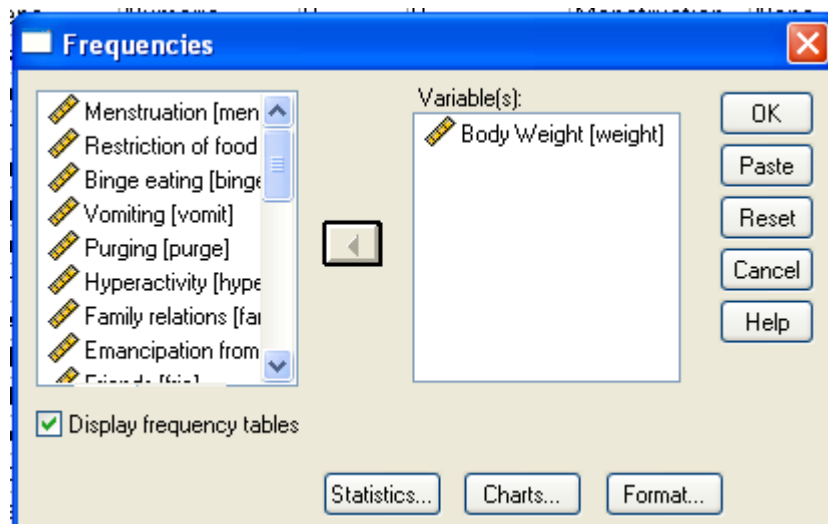
10.1.1 Overall Percentages

To obtain the overall frequency in each category:

- ☞ *Analyze > Descriptive Statistics > Frequencies...*
- ☞ Put measure in *Variable(s):* box
- ☞ Click *OK*

Make sure *Display frequency tables* is checked, and that in the *Statistics...* screen, none of the boxes are checked.

Example When obtaining the frequency of subjects in each group the window should look like this:

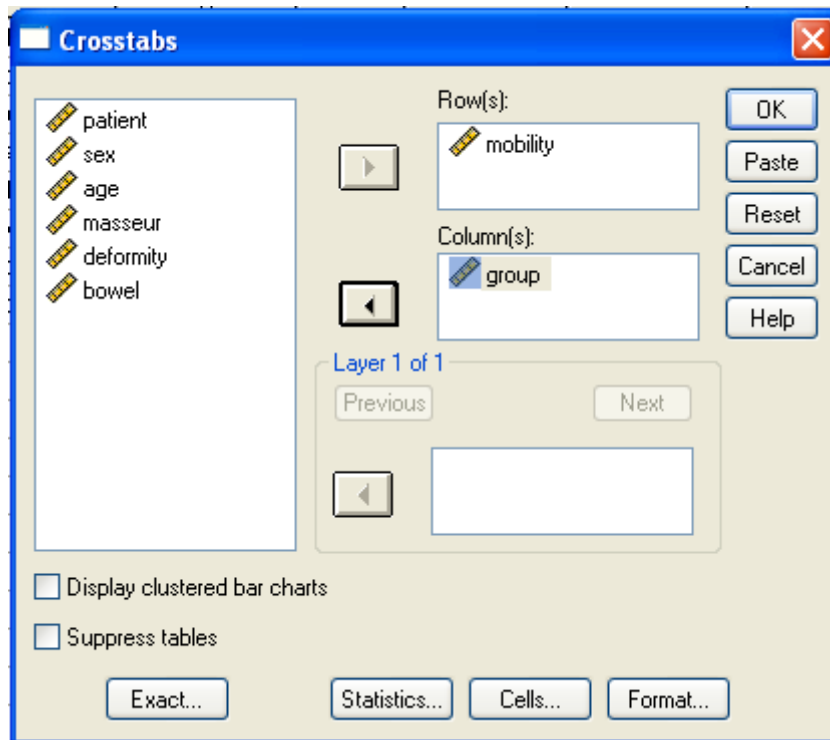


10.1.2 Percentages per group

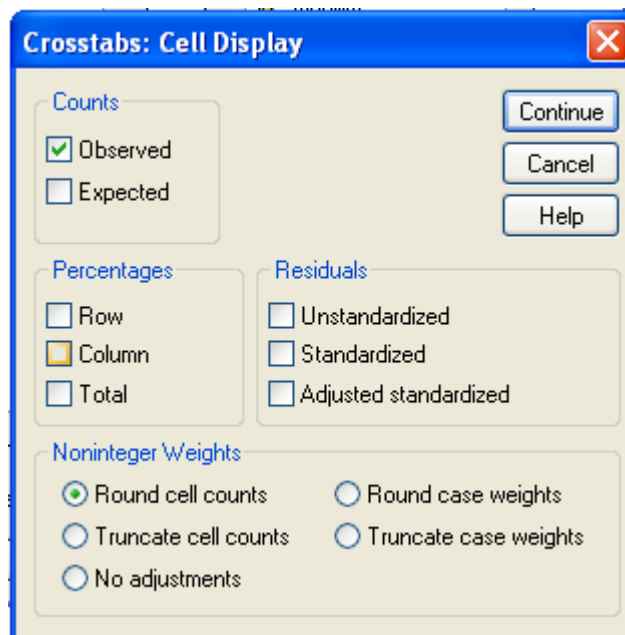
To examine how the frequencies differ between groups we generate **cross tabulations**, with percentages out of number in each group:

- ☞ *Analyze > Descriptive Statistics > Crosstabs...*
- ☞ Put the grouping variable into *Column(s):* and the other measure(s) in *Row(s):*
- ☞ *Cells... > under Percentages click Column > Continue > OK*

Example When obtaining percentages for mobility classification per group the main window should look like:



And the *Cells...* window should look like this:

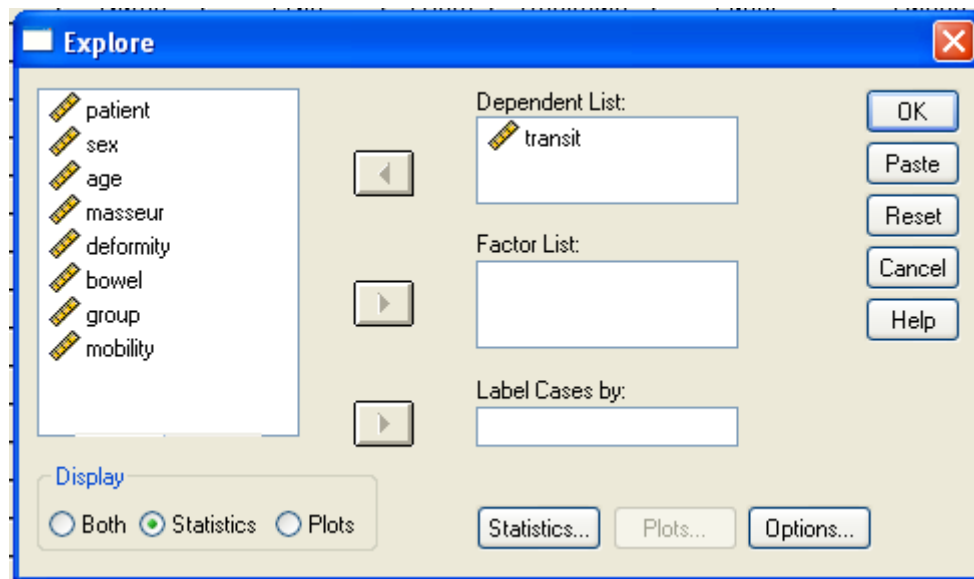


10.2 Descriptive Statistics

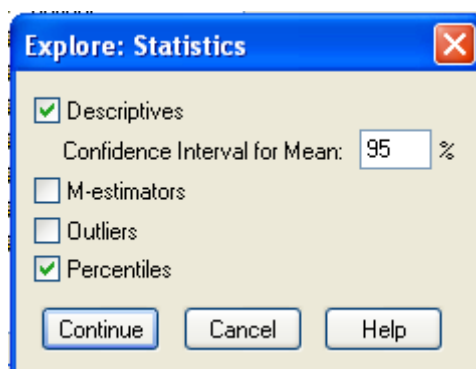
10.2.1 Overall Descriptive

- ☞ *Analyze > Descriptive statistics > Explore...*
- ☞ Put the measure to be summarised in the *Dependent List:* box
- ☞ Under *Display*, check *Statistics*
- ☞ *Statistics...* > make sure *Descriptives* and *Percentiles* are checked > *Continue*
- ☞ Click *OK*

Example When obtaining summary statistics for transit time the window should look like this:

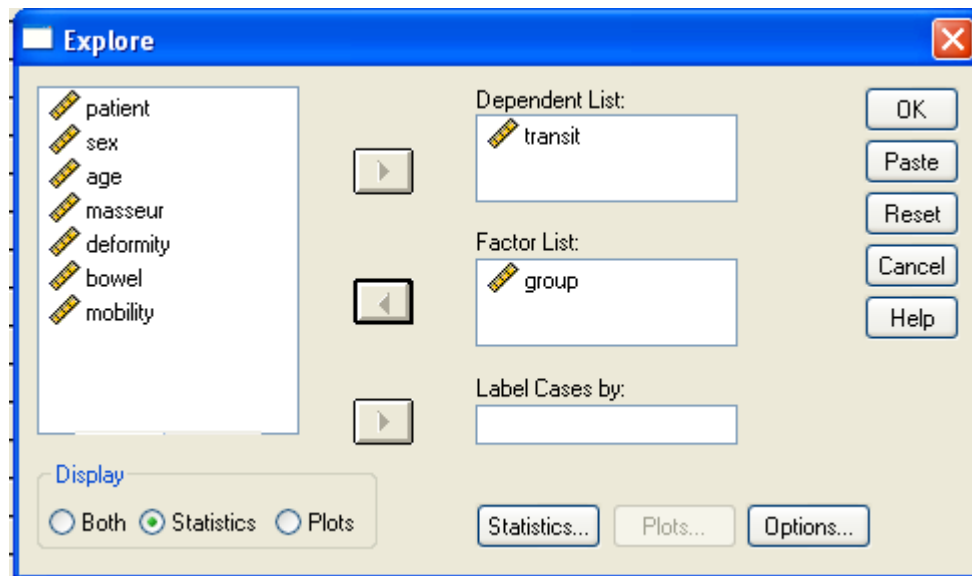


and the *Statistics..* window should look like this:



10.2.2 Descriptives per group

Follow instructions as for 10.2.1, but put the grouping variable into the *Factor List:* box in the *Explore* window as below:

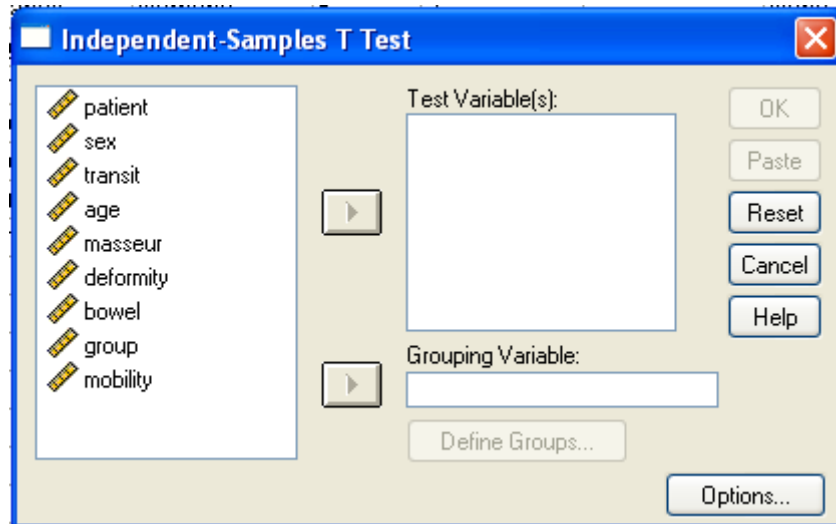


11 Statistical Tests

11.1 T-tests (parametric test comparing a numerical outcome across two groups)

☞ *Analyze > Compare Means > Independent-Samples T test...*

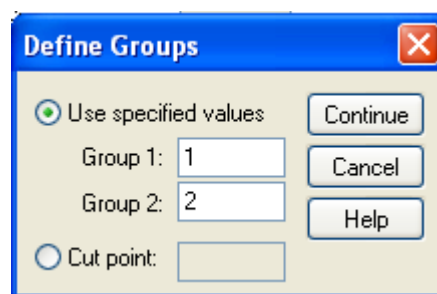
The following window will appear:



- ☞ Place the measure to be tested in the *Test variable(s):* box
- ☞ Place the group variable in the *Grouping Variable:* box
- ☞ Click *Define Groups...*

In the new window

- ☞ Put the value which the first group is coded as in the *Group 1:* box and the value of the remaining group in *Group 2:* box. For example, if your group is coded as 1 and 2 then it should look as below.



- ☞ *Continue > OK*

Below is an example of the output:

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
transit time, phase	Massage	16	159.19	59.44	14.86
a, total gi tract	Laxative	15	137.47	65.64	16.95

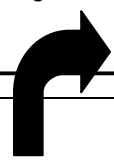
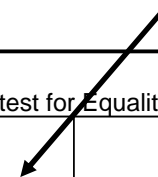
Summary statistics



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
transit time, phase a, total gi tract	Equal variances assumed	.645	.428	.967	29	.342	21.72	22.47	-24.23	67.67
	Equal variances not assumed			.964	28.229	.343	21.72	22.54	-24.44	67.88

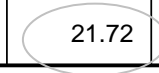
Otherwise known as p-value



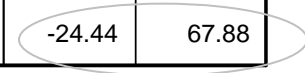
Use the results on this row



Ignore these columns



Estimated mean difference between the two groups

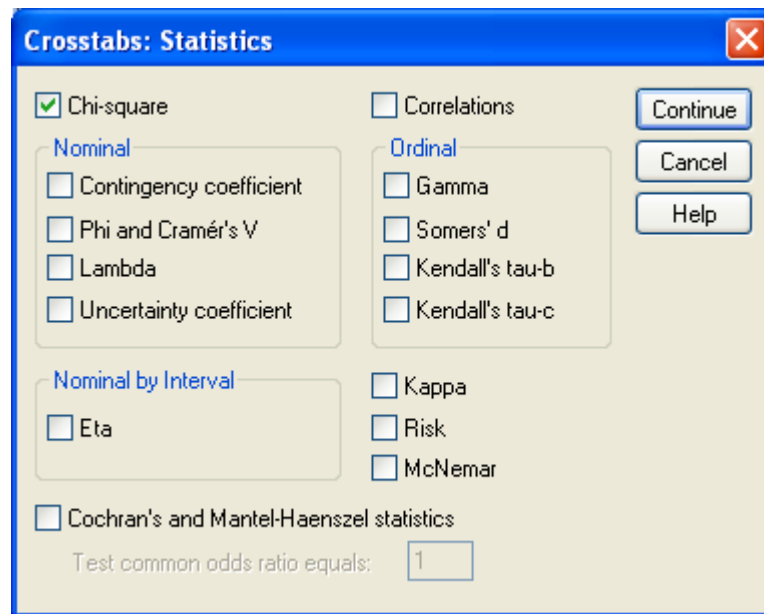


Confidence interval for estimated mean difference between the two groups

11.2 Chi-squared test, Chi-squared test for trend, and Fisher's Exact test (comparing a categorical outcome across two or more groups)

Create cross-tabulation as in 10.1.2:

- ☞ *Analyze > Descriptive Statistics > Crosstabs...*
- ☞ Put the grouping variable into *Column(s)*: and the other measure(s) in *Row(s)*:
- ☞ *Statistics...* > Check *Chi-square* as shown below:



Note: The result for chi-squared test for trend is automatically produced when *Chi-square* is selected. Fisher's Exact test is automatically calculated when you have a 2*2 table and select *Chi-square*.

- ☞ *Continue > OK*

An example of the output is shown below:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Clinical Improvement? * GROUP	31	100.0%	0	.0%	31	100.0%

Clinical Improvement? * GROUP Crosstabulation

			GROUP		Total
			Group 1	Group 2	
Clinical Improvement?	Yes	Count	4	10	14
		% within GROUP	25.0%	66.7%	45.2%
	No	Count	12	5	17
		% within GROUP	75.0%	33.3%	54.8%
Total		Count	16	15	31
		% within GROUP	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.427 ^b	1	.020		
Continuity Correction ^a	3.875	1	.049		
Likelihood Ratio	5.594	1	.018		
Fisher's Exact Test				.032	.024
Linear-by-Linear Association	5.252	1	.022		
N of Valid Cases	31				

p-value

Use top row, fourth column from left, for two sided p-value of Chi-squared Test

Use fourth row, second column from right, for two sided p-value of Fisher's Exact Test.

Use fifth row, fourth column from left, for two-sided p-value of Chi-squared test for trend.

Usual crosstab output

Note:
If more than 20% of cells have expected count less than 5, or minimum expected count is less than 1, then Chi-squared test is not appropriate. Use Fisher's Exact test.

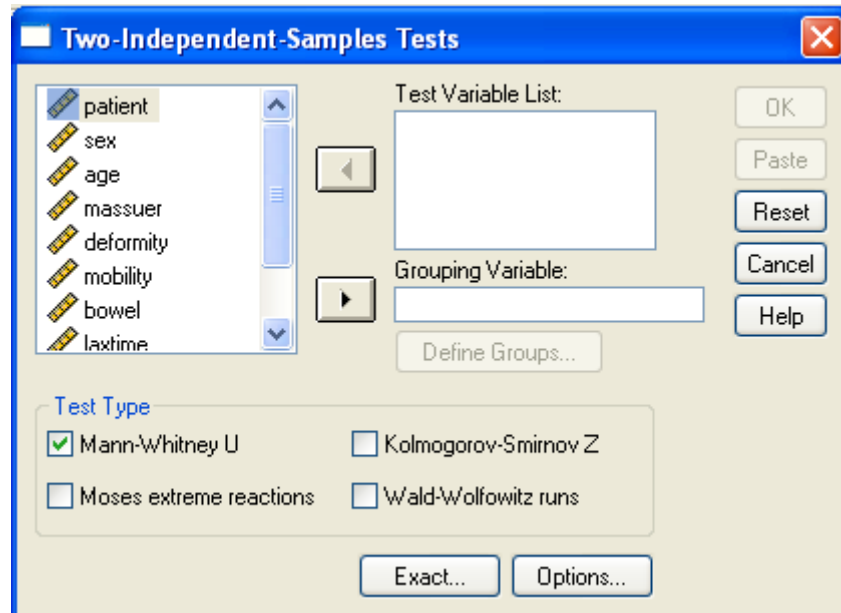
a. Computed only for a 2x2 table
b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.77.

11.3 Mann-Whitney U Test

(non-parametric test comparing a numerical outcome across two groups)

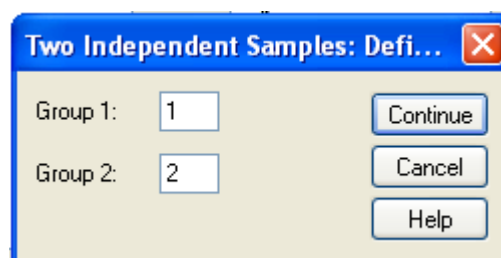
☞ *Analyze > Nonparametric Tests > 2 Independent Samples...*

The following window will appear:



- ☞ Put the measure to be tested in the *Test Variable List:* box
- ☞ Put the group variable in the *Grouping Variable:* box
- ☞ Make sure that under *Test Type*, Mann-Whitney U is checked.
- ☞ Click *Define Groups...*

☞ Assign a value to *Group 1:* and *Group 2:*
E.g. if your groups are coded as '1' and '2' then it will look as below:



☞ *Continue > OK.*

The output will be something similar to below:

NPar Tests

Mann-Whitney Test

Ranks

	GROUP	N	Mean Rank	Sum of Ranks
top transit time	Massage	16	18.19	291.00
post treatment	Laxative	15	13.67	205.00
	Total	31		

Test Statistics^b

	top transit time post treatment
Mann-Whitney U	85.000
Wilcoxon W	205.000
Z	-1.385
Asymp. Sig. (2-tailed)	.166
Exact Sig. [2*(1-tailed Sig.)]	.175 ^a

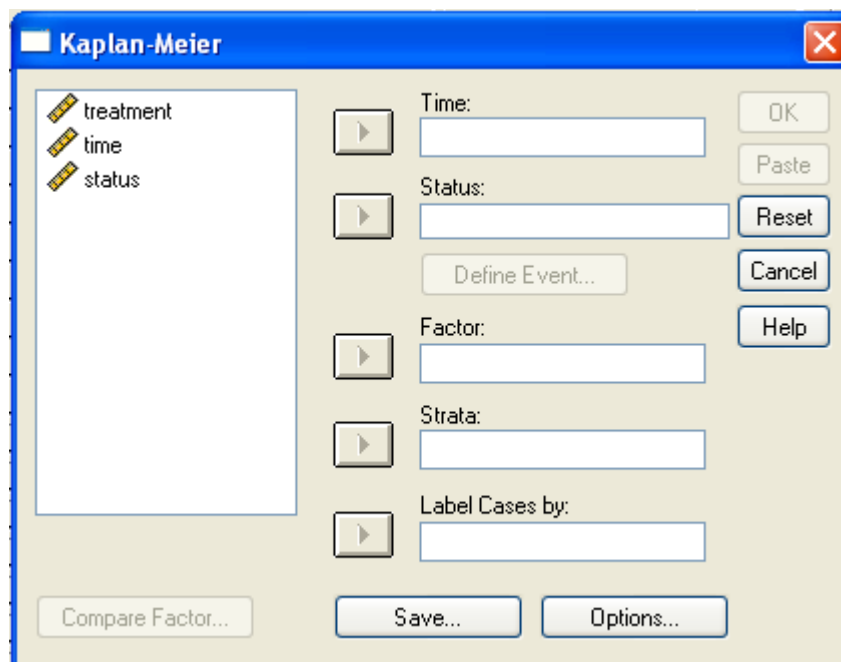
a. Not corrected for ties.

b. Grouping Variable: GROUP

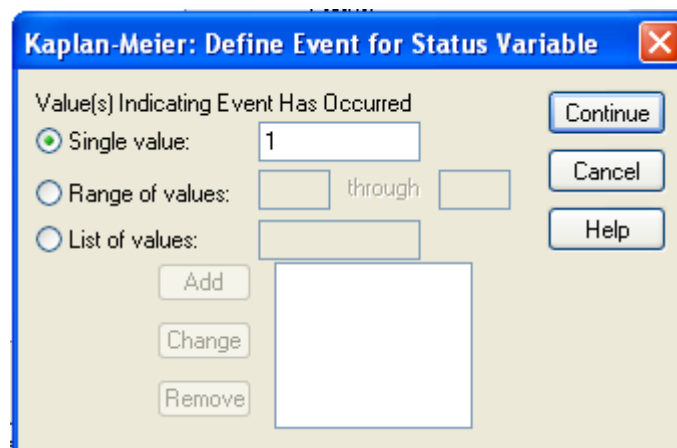
11.4 Kaplan-Meier graph and Log-Rank test (comparing a time-to-event outcome across two or more groups)

☞ *Analyze... > Survival > Kaplan-Meier...*

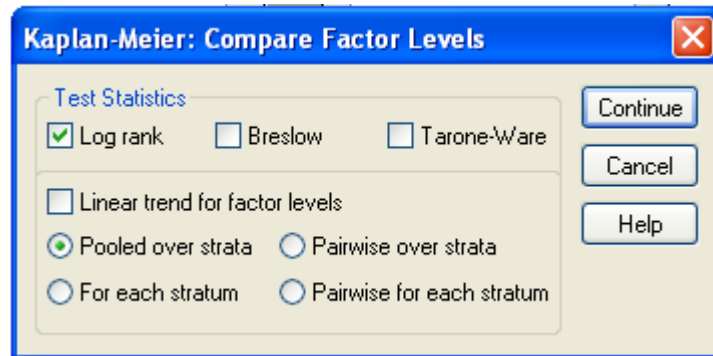
The following screen should appear:



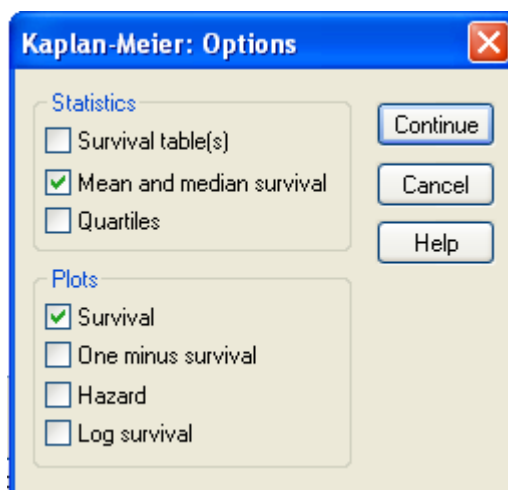
- ☞ Place the time to event measure in the *Time:* box
- ☞ Place the explanatory factor or group in the *Factor:* box
- ☞ Put the measures which indicates whether an event has occurred (e.g. death) in the *Status:* box
- ☞ Click *Define Event...*
- ☞ In the window that appears enter the value that defines whether an event has occurred into the "Single value:" box, as below:



- ☞ Click *Continue* > *Compare Factor...*
- ☞ In the window that appears check the *Log rank* box as below:



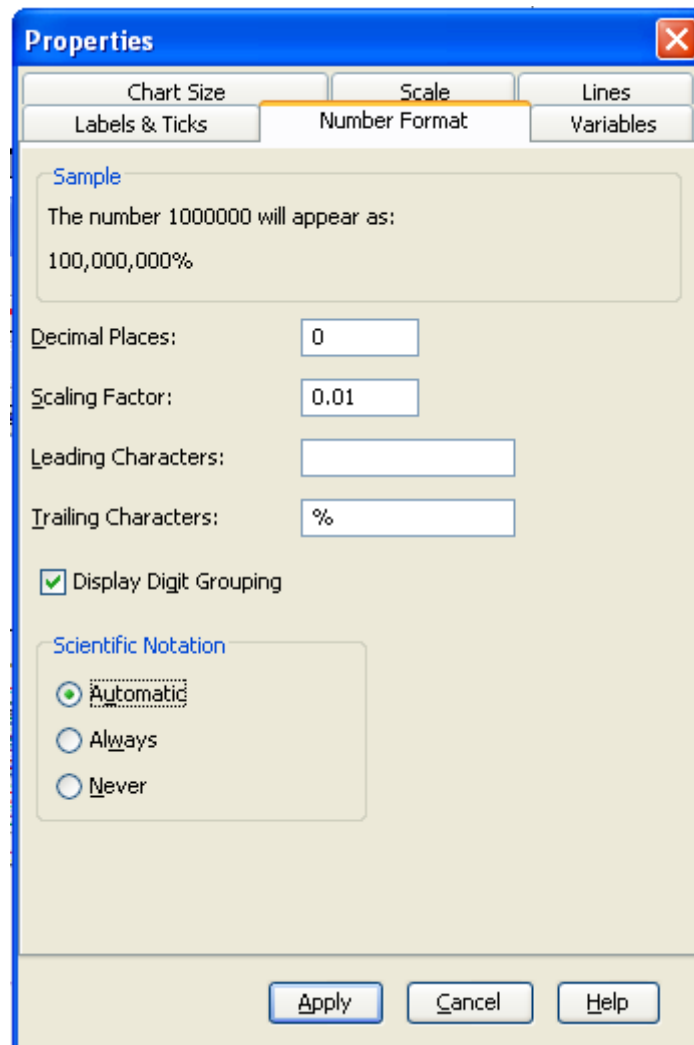
- ☞ Click *Continue* > *Options...*
- ☞ Check (or un-check) the boxes so the window looks as the one below:



- ☞ *Continue* > *OK*

In the output window:

- ☞ Double click on the Kaplan-Meier graph so that you can edit it
- ☞ In the graph editor window, double click on the vertical (Y scale) axis
- ☞ Choose *Number Format*
- ☞ Change the entries so that the *Properties* window looks as on the following page.



- ☞ *Apply* > Close *Properties* window
- ☞ Double click on the horizontal (X scale) axis
- ☞ Choose *Scale*
- ☞ Change the *Minimum* and *Maximum* entries and *Major Increment* to an appropriate scale i.e. if in time is days, change *Major Increment* to 7, 30 or 365 depending on what would be appropriate for the data.

For example, if the time is in weeks, the *Major Increment* could to change to 52, the *Minimum* to 0 and the *Maximum* to 204. This means the time can easily be related to years.

The output is shown below:

Kaplan-Meier

Case Processing Summary

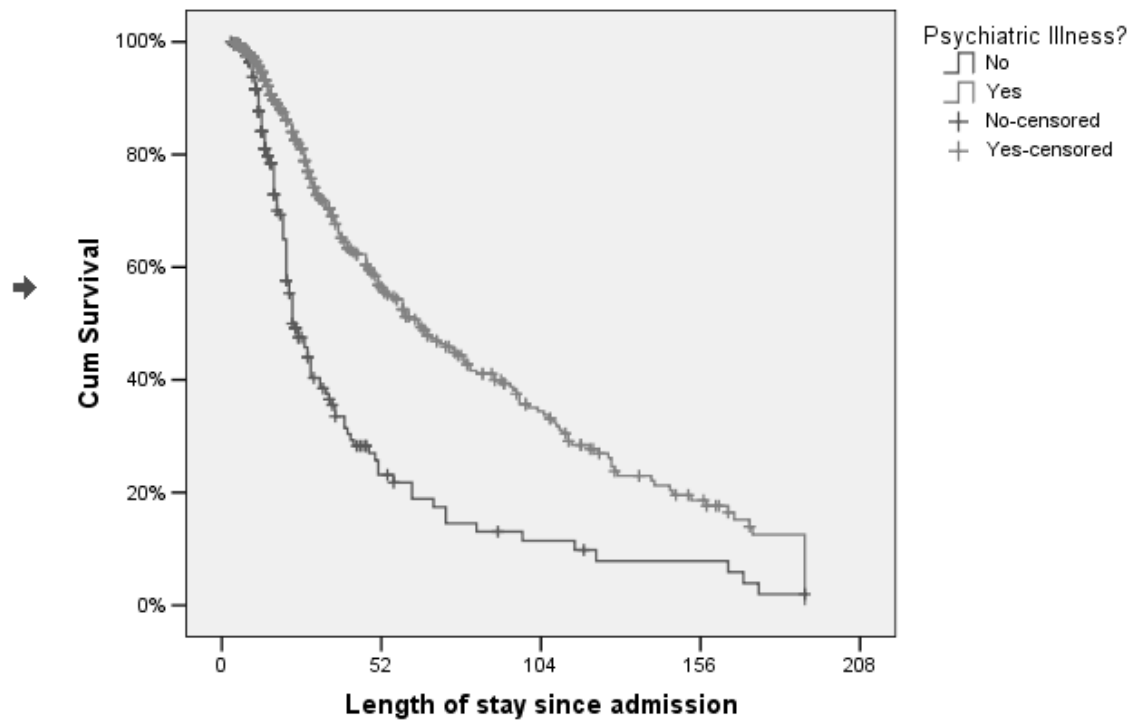
psyillyn	Total N	N of Events	Censored	
			N	Percent
No	208	117	91	43.8%
Yes	523	222	301	57.6%
Overall	731	339	392	53.6%

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	58.841	1	.000

Test of equality of survival distributions for the different levels of Psychiatric Illness?.

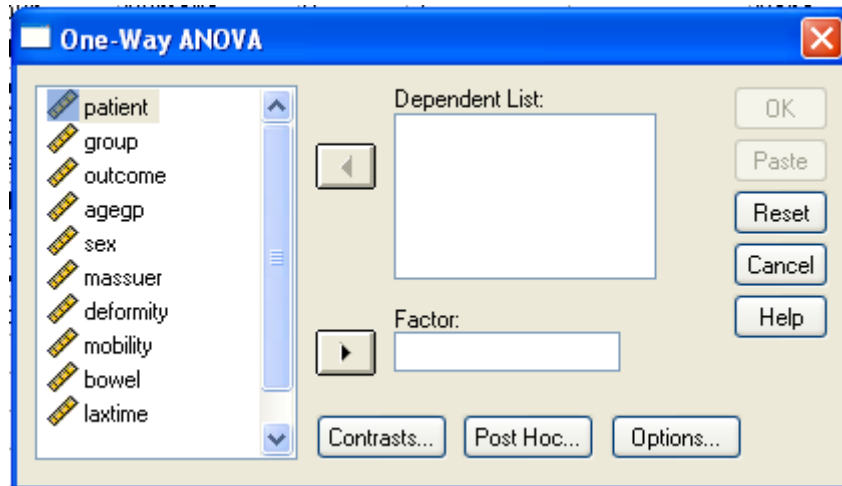
Survival Functions



11.5 ANOVA

(parametric procedure comparing a numerical outcome across more than two groups)

☞ *Analyze > Compare Means > One way ANOVA...*



- ☞ Place the outcome into the *Dependent List*: box
- ☞ Place the measure to compare between (e.g. group) in the *Factor*: box
- ☞ Click *OK*

An example of the output is shown below:

Oneway

ANOVA

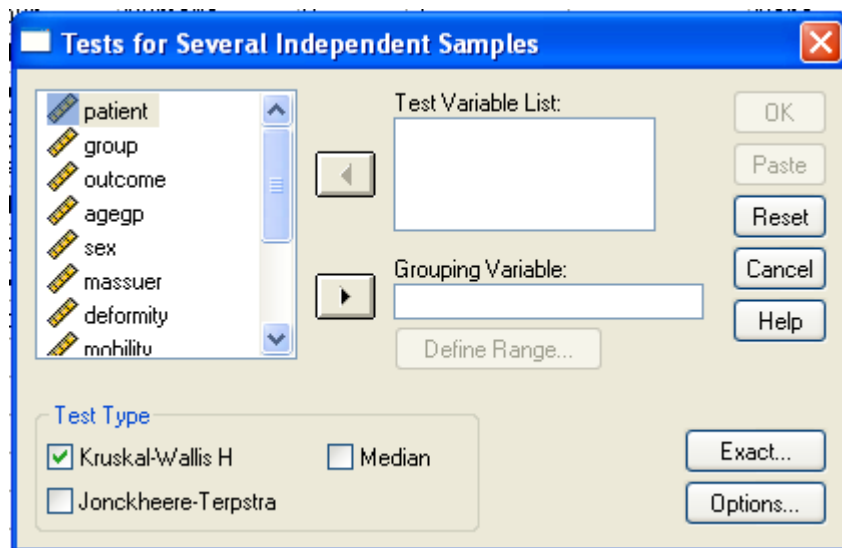
Outcome Measure

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9646.333	2	4823.167	4.895	.023
Within Groups	14779.667	15	985.311		
Total	24426.000	17			

P-value

11.6 Kruskal Wallis test (non-parametric alternative to ANOVA)

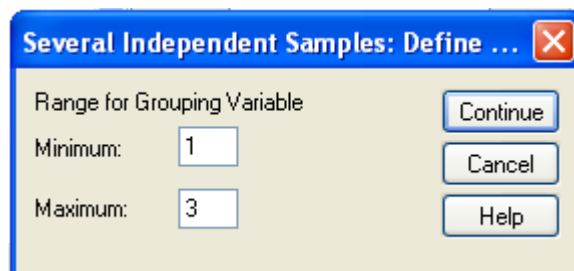
☞ Analyze > Nonparametric Tests > K independent Samples...



- ☞ Put the outcome into the *Test Variable List*: box
- ☞ Put the measure to compare between (e.g. group) in the *Grouping Variable*: box
- ☞ Click *Define Range...*

In the new window:

- ☞ Put the range of numbers for the group coding into the boxes
E.g. if groups have been coded as '1', '2' or '3' then the window should look as below:



The output will be something similar to below:

NPar Tests

Kruskal-Wallis Test

Ranks

	Group	N	Mean Rank
Outcome measure	Group 1	6	13.00
	Group 2	6	8.25
	Group 3	6	7.25
	Total	18	

Test Statistics^{a,b}

	Outcome measure
Chi-Square	3.982
df	2
Asymp. Sig.	.137

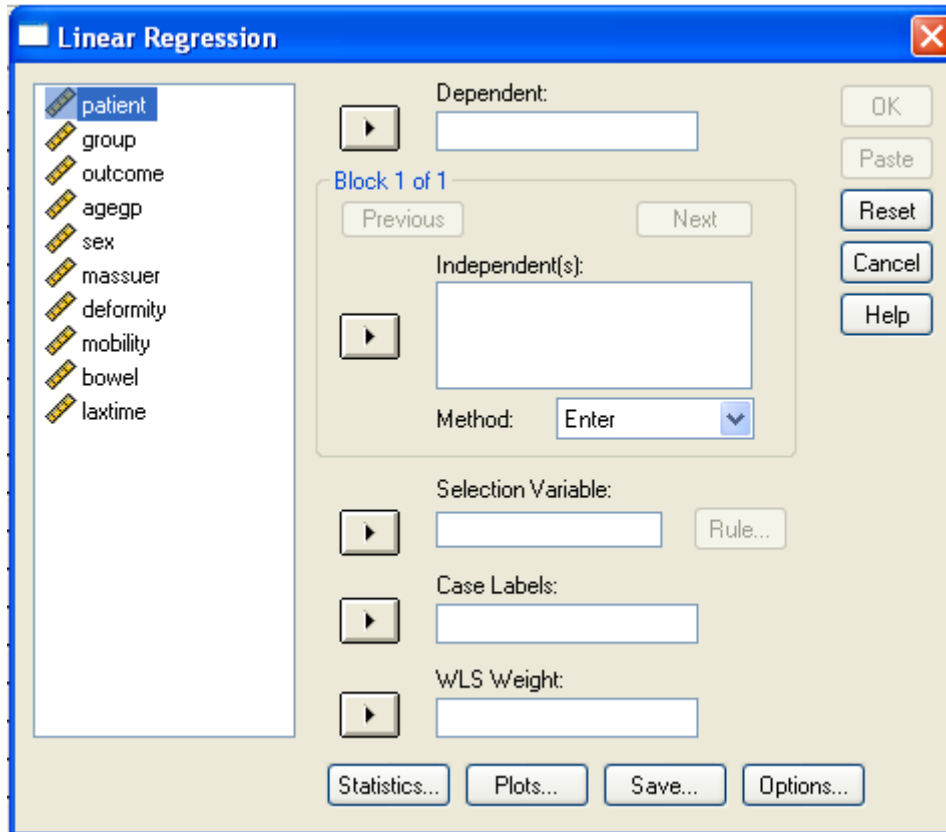
a. Kruskal Wallis Test

b. Grouping Variable: Group

p-value

12 Linear Regression (numerical, normally distributed outcome)

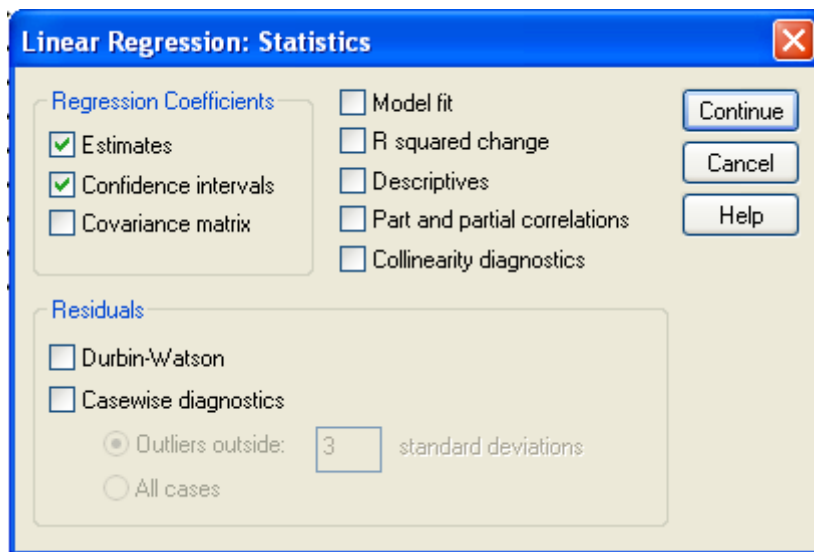
☞ *Analyze > Regression > Linear...*



- ☞ Place outcome variable in *Dependent:* box
- ☞ Place explanatory factors in *Independent(s):* box

Make sure *Method:* status is *Enter*.

- ☞ Click *Statistics...*
- ☞ Check *Confidence intervals* and un-check *Model fit* as on the following page.

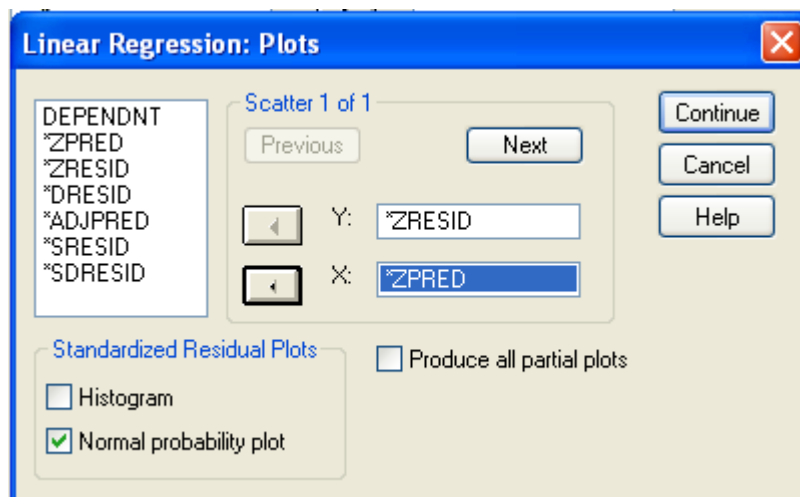


☞ Click *Continue*

To check linear regression assumptions:

☞ Click *Plots...*

☞ Place **ZRESID* in Y: box, place **ZPRED* in X: box and check *Normal probability plot* as below:



☞ *Continue > OK*

An example of the output is given below:

Regression

		Coefficients ^a				95% Confidence Interval for B		
		Unstandardized Coefficients		Standardized Coefficients				
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	89.650	57.224		1.567	.129	-27.763	207.06
	GROUP	-9.278	21.645	-.075	-.429	.672	-53.691	35.135
	transit time, baseline, total gi tract	.489	.198	.448	2.469	.020	.083	.895
	AGE	-.226	.887	-.045	-.255	.801	-2.046	1.594

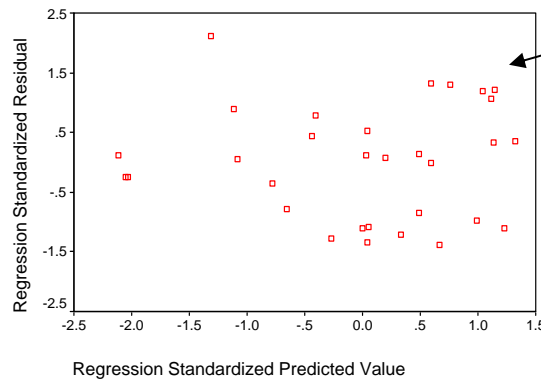
a. Dependent Variable: transit time, post treatment, total gi tract

Charts

Scatterplot

Transit time

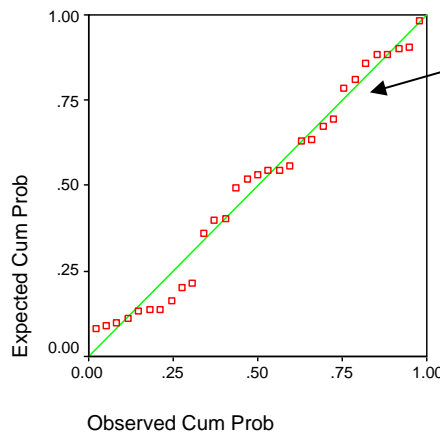
scale should be no more than +3 to -3 unless hundreds of cases



should look like there isn't a pattern

Normal P-P Plot of zres

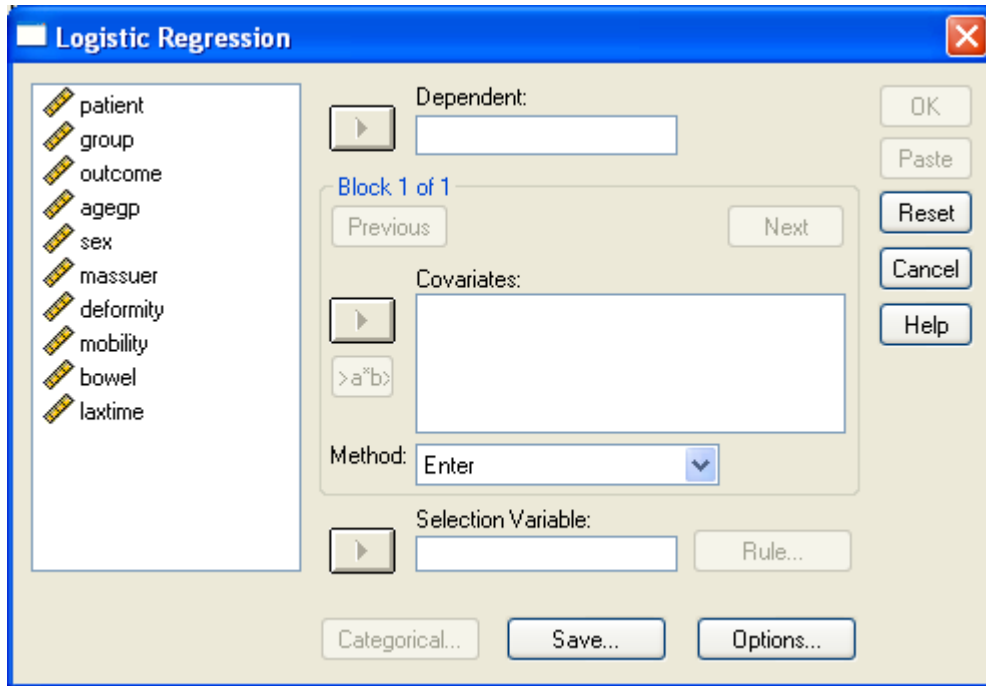
Transit time



should be straight line at 45° with any misfit having no pattern

13 Logistic Regression (binary outcome)

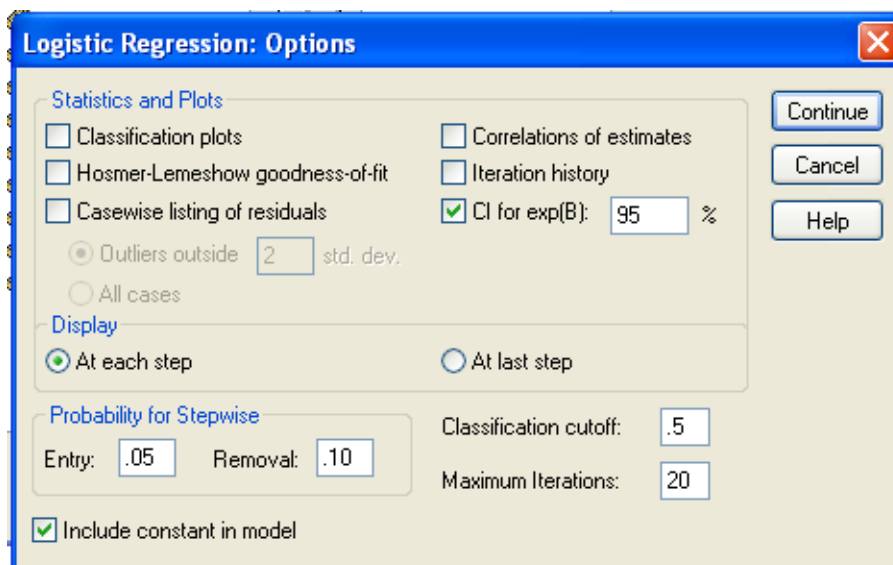
☞ Analyze > Regression > Binary Logistic...



- ☞ Place outcome variable in *Dependent:* box
- ☞ Place explanatory factors in *Covariates:* box

Make sure *Method:* status is *Enter*.

- ☞ Click *Options...*
- ☞ Check *CI for exp(B):* and leave status as *95%* as below:



☞ Click *Continue*

You only need to do the following if you have included any explanatory factors that are categorical but are not coded as '0's and '1's.

☞ Click *Categorical...*

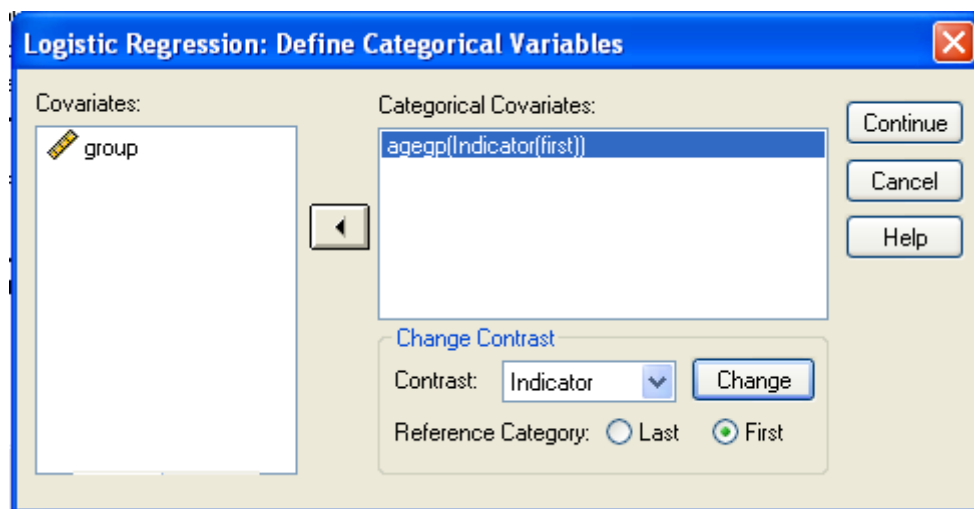
☞ Out of the *Covariates*: list, place explanatory factors which are not coded as '0's and '1's in the *Categorical Covariates* box.

☞ Leave *Contrast*: status as *Indicator*.

☞ Choose the *Reference Category*: to be either clinically sensible or one with many cases in it.

☞ Click *Change*

An example of how the screen could look is given below:



☞ Click *Continue* > *OK*

An example of the some of the output is shown below:

Logistic Regression

Dependent Variable Encoding

Original Value	Internal Value
2 weeks or less	0
more than 2 weeks	1

An odds ratio greater than one means that the outcome is more likely

Table showing how the dummies were created for categorical case-mix factors

Categorical Variables Codings

		Frequency	Parameter coding		
			(1)	(2)	(3)
Lives Alone	339	.000	.000	.000	
with Family	210	1.000	.000	.000	
Home	180	.000	1.000	.000	
Other	2	.000	.000	1.000	

Block 1: Method = Enter

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	813.236	.147	.205

p-value

Odds ratio for group effect

Variables in the Equation

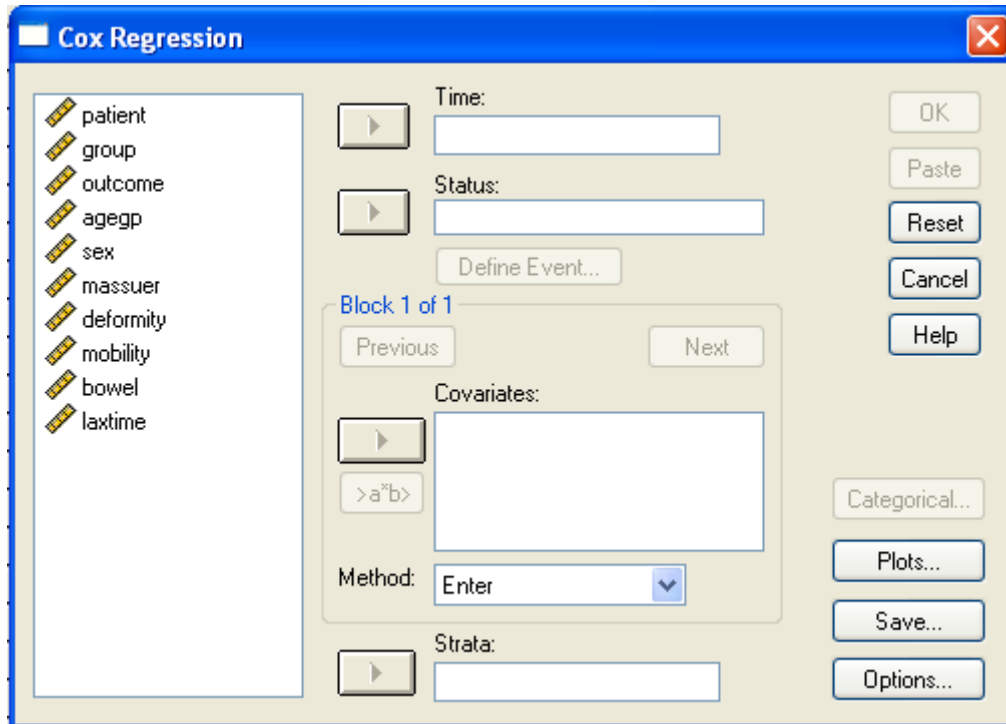
Step	Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
								Lower	Upper
1 ^a	GROUP	.719	.205	12.348	1	.000	2.052	1.374	3.064
	AGE	.054	.012	18.702	1	.000	1.055	1.030	1.081
	LIVewith			89.743	3	.000			
	LIVewith(1)	-.719	.210	11.737	1	.001	.487	.323	.735
	LIVewith(2)	-2.197	.232	89.610	1	.000	.111	.071	.175
	LIVewith(3)	-.964	1.429	.455	1	.500	.381	.023	6.280
	Constant	-3.328	.991	11.271	1	.001	.036		

a. Variable(s) entered on step 1: PSYILLYN, AGE, LIVewith.

14 Survival Regression (time-to-event outcome)

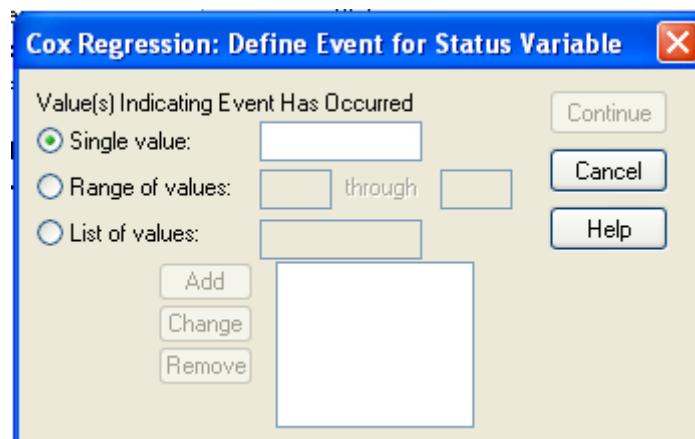
☞ *Analyze > Survival > Cox Regression...*

A screen similar to below should appear:



- ☞ Place time-to-event measure in *Time:* box
- ☞ Place measure which indicates whether an event occurred or not in *Status:* box
- ☞ Click *Define Event...*

A screen similar to below will appear:



- ☞ Enter the value that indicates that the event occurred in the *Single value:* box.
- ☞ Click *Continue*

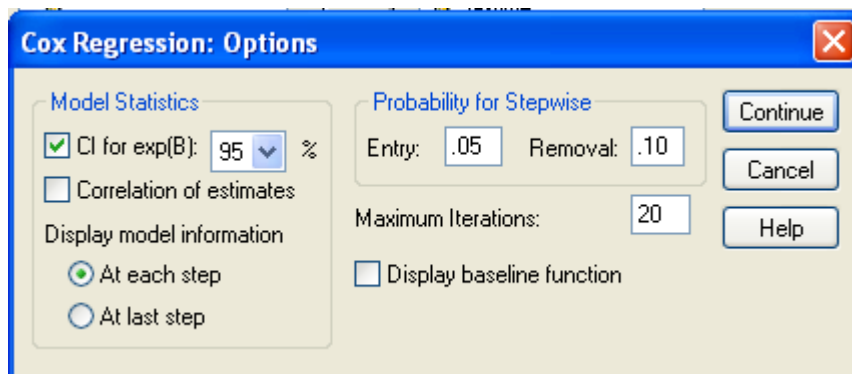
For example if you were interested in time to death, and you had a measure which was coded '1' if the patient died and '0' otherwise, you would enter '1' in the *Single value:* box.

☞ Place explanatory factors in *Covariates:* box

Make sure *Method:* status is *Enter*.

☞ Click *Options...*

☞ Check *CI for exp(B):* and leave status as *95%:* as below:



☞ Click *Continue*

If you have included any explanatory factors that are categorical but are not coded as '0's and '1's:

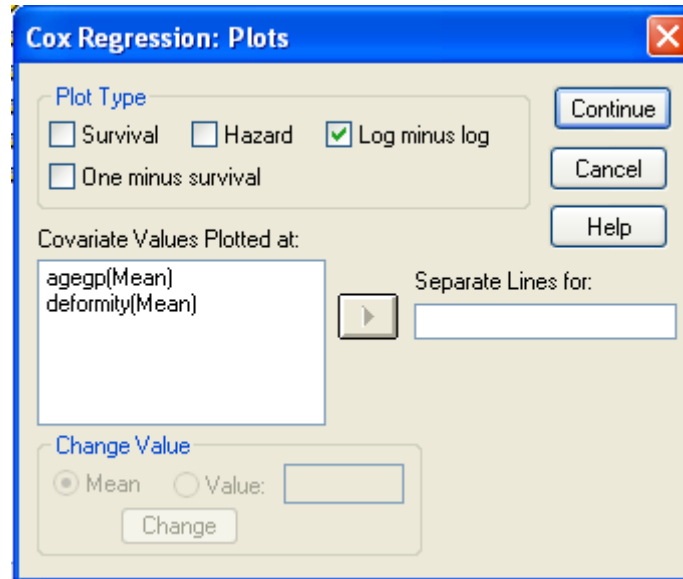
☞ Click *Categorical...* and follow the same method as in **Section 13** for selecting *Categorical Covariates* and appropriate *Contrast*.

To check constant hazard ratio assumption:

Perform the analysis as before but in the *Cox Regression* box place one of the categorical explanatory factors into the *Strata:* box instead of the *Covariates:* box.

☞ Click *Plots...*

☞ Check *Log minus log* as below:



☞ Click *Continue > OK*

Repeat for all explanatory factors. For any factors that are not categorical, create a new variable which categorises the factor and put this in the *Strata:* box (having removed it from *Covariates:*).

An example of some of the output is shown below:

Cox Regression
Block 1: Method = Enter

p-value

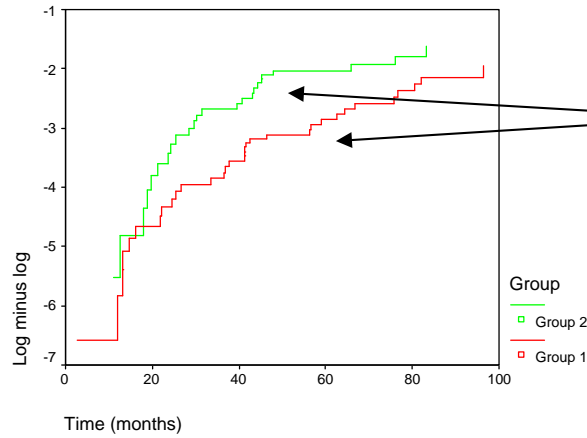
hazard ratio for group effect

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Group	.692	.277	6.266	1	.012	1.998	1.162	3.436
AGE	-.013	.011	1.601	1	.206	.987	.966	1.007
HISTGRAD			4.978	2	.083			
HISTGRAD(1)	-1.1	.737	2.426	1	.119	.317	.075	1.345
HISTGRAD(2)	-.521	.278	3.514	1	.061	.594	.344	1.024

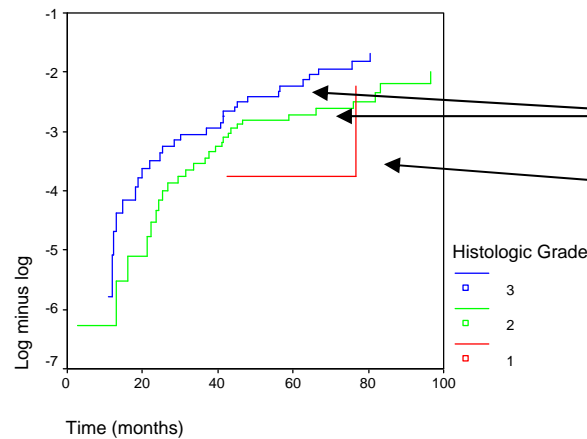
Complementary log-log plots

LML Function at mean of covariates



*this is fine,
roughly
parallel*

LML Function at mean of covariates

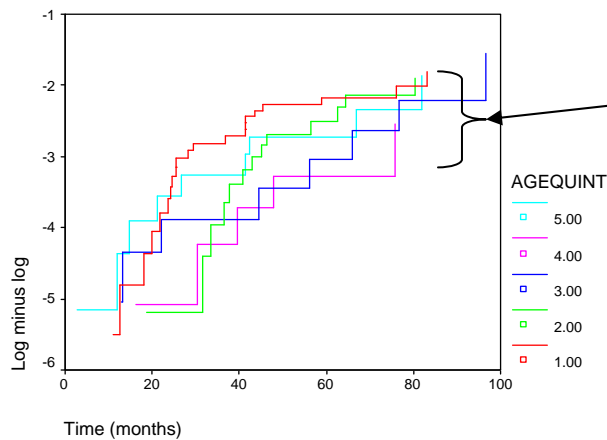


*these two
are fine*

*maybe few cases
in this category so
could combine with
another category*

*for each
explanatory
factor, the
category lines
should be
parallel and not
cross*

LML Function at mean of covariates



*this is not
good, most
lines cross
over each
other - leave
age as strata*