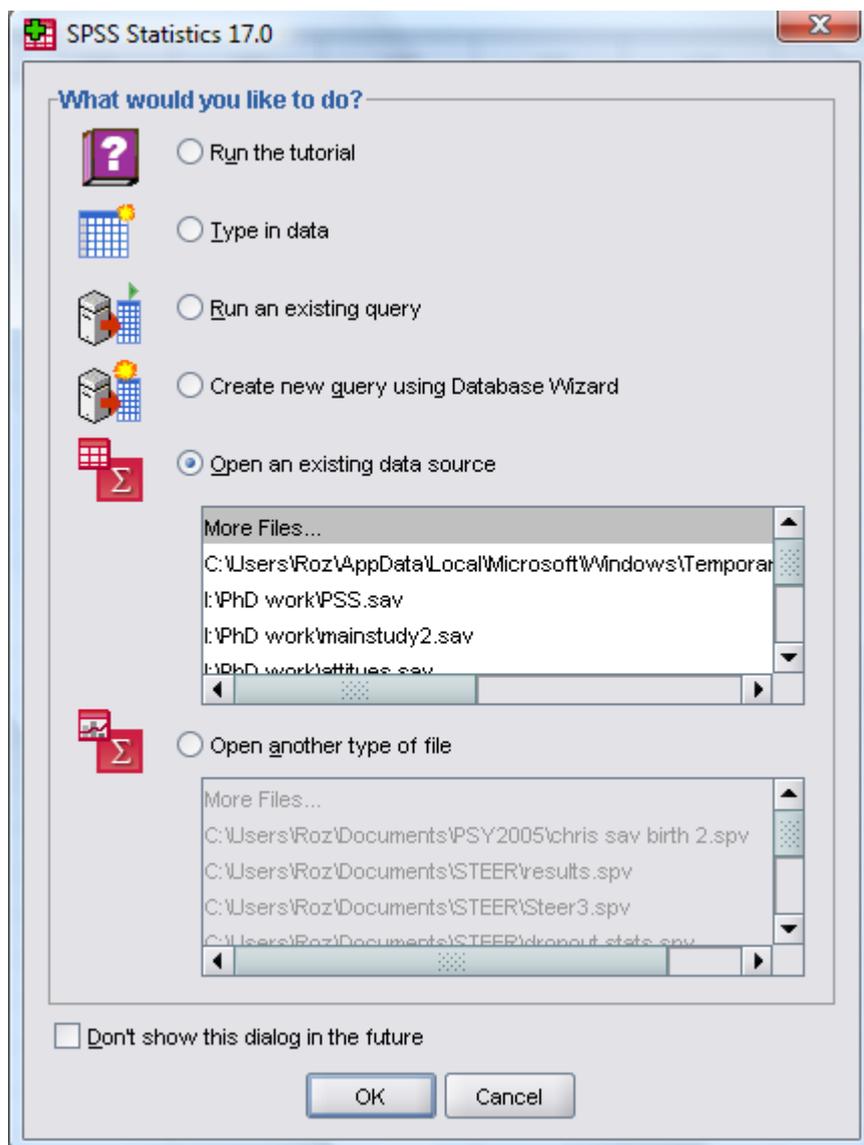# Getting to grips with SPSS

When you open SPSS it will ask what you want to do first (see Fig 1). If you have been working on an existing file you can open it directly from here by highlighting and clicking OK. To start a new data base simply close this dialog.

Fig 1: SPSS dialog

When you first open SPSS you will notice 2 different forms of the data base Variable View where you name the variables within the data base and Data View where you type in all the numbers.

Fig 2 shows a blank data base in variable view.
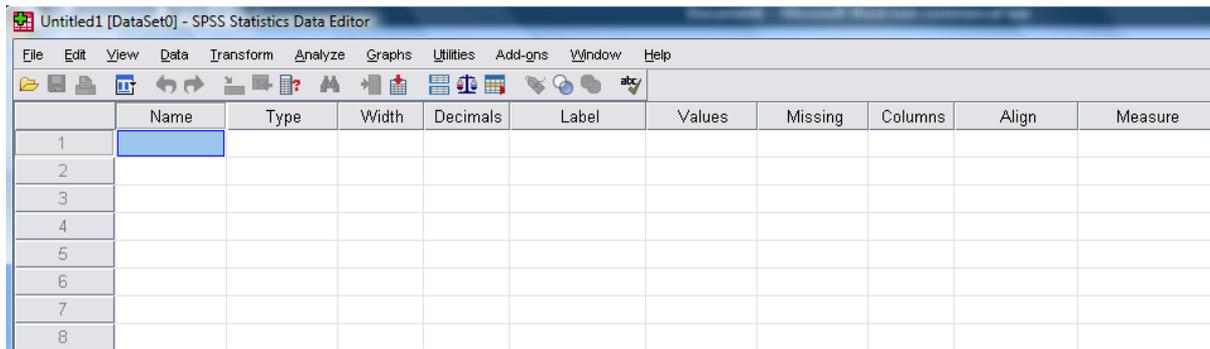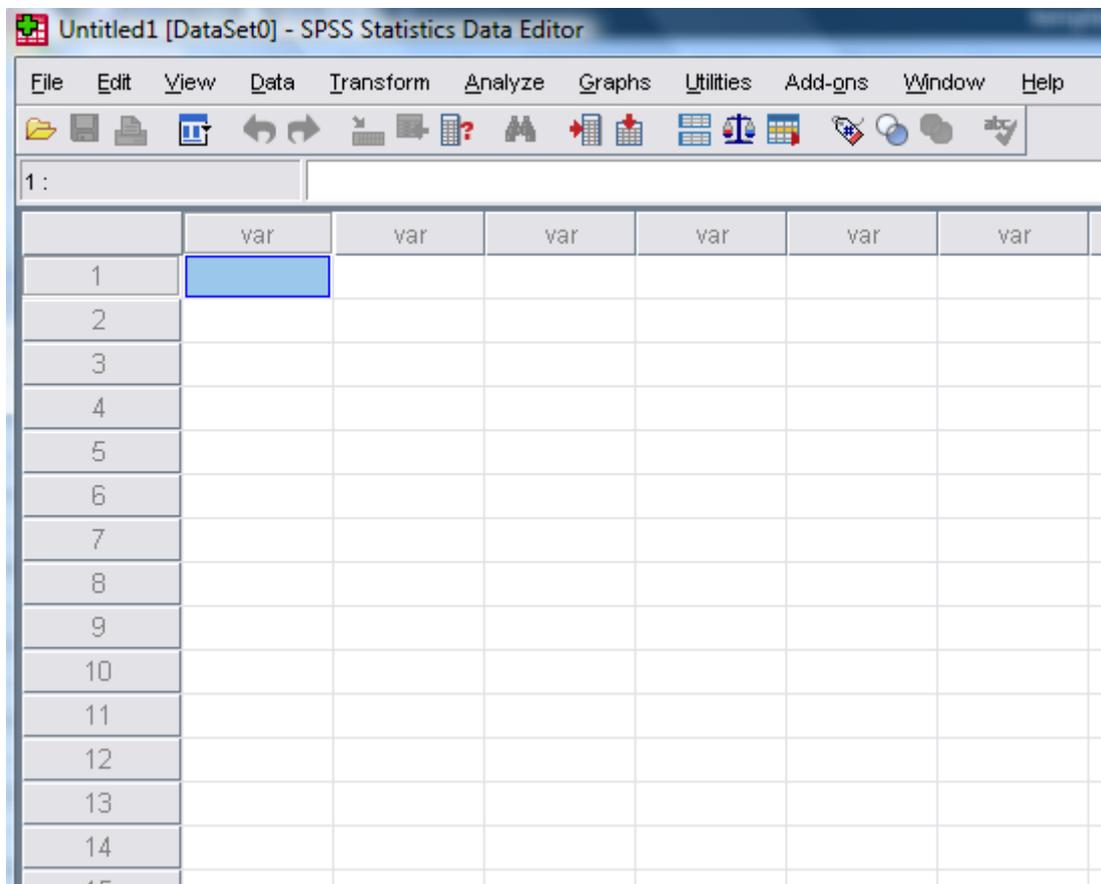
Fig 2: Variable view

Fig 3: Data view blank

You need to be in variable view to type up what variables you have in your data set.

You will need to know whether you have Within Subjects data or Between Subjects. Your data will be entered very differently for the different designs.

Additionally you will need to know whether your variables are categorical or continuous. Categorical variables require an additional step.

Fig 4: Variable view with two variables inserted. "Levels" is categorical (note the 'nominal' under 'Measure)'. "Recall" is continuous (note the 'scale' under 'Measure'.

You can label your variable if you would like. For example if the name is too long you may want to have an abbreviated word in 'Name' but have the longer version in 'Label'

For categorical data you will need to value the categories.

Click on the box in the Values column.
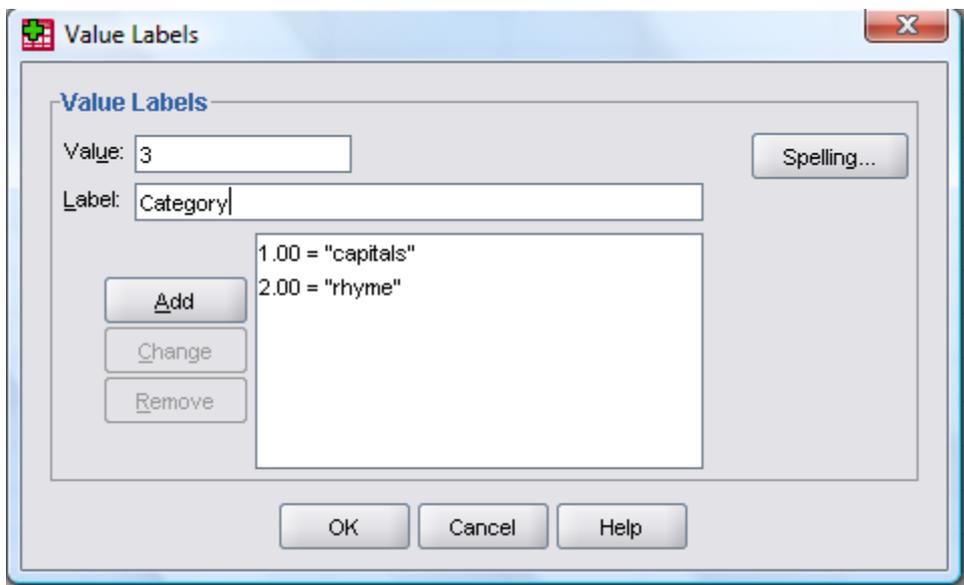
Within this box you need to state the value you would like to give for each level of the variable.

Don't forget to press add after each value.

Once this is complete you can return to data view and start inserting your raw data (see Fig ?)

Once this is complete you can return to data view and start inserting your raw data (see Fig ?)

This is how a data set looks for Between subjects data

Each row equates to 1 participant.

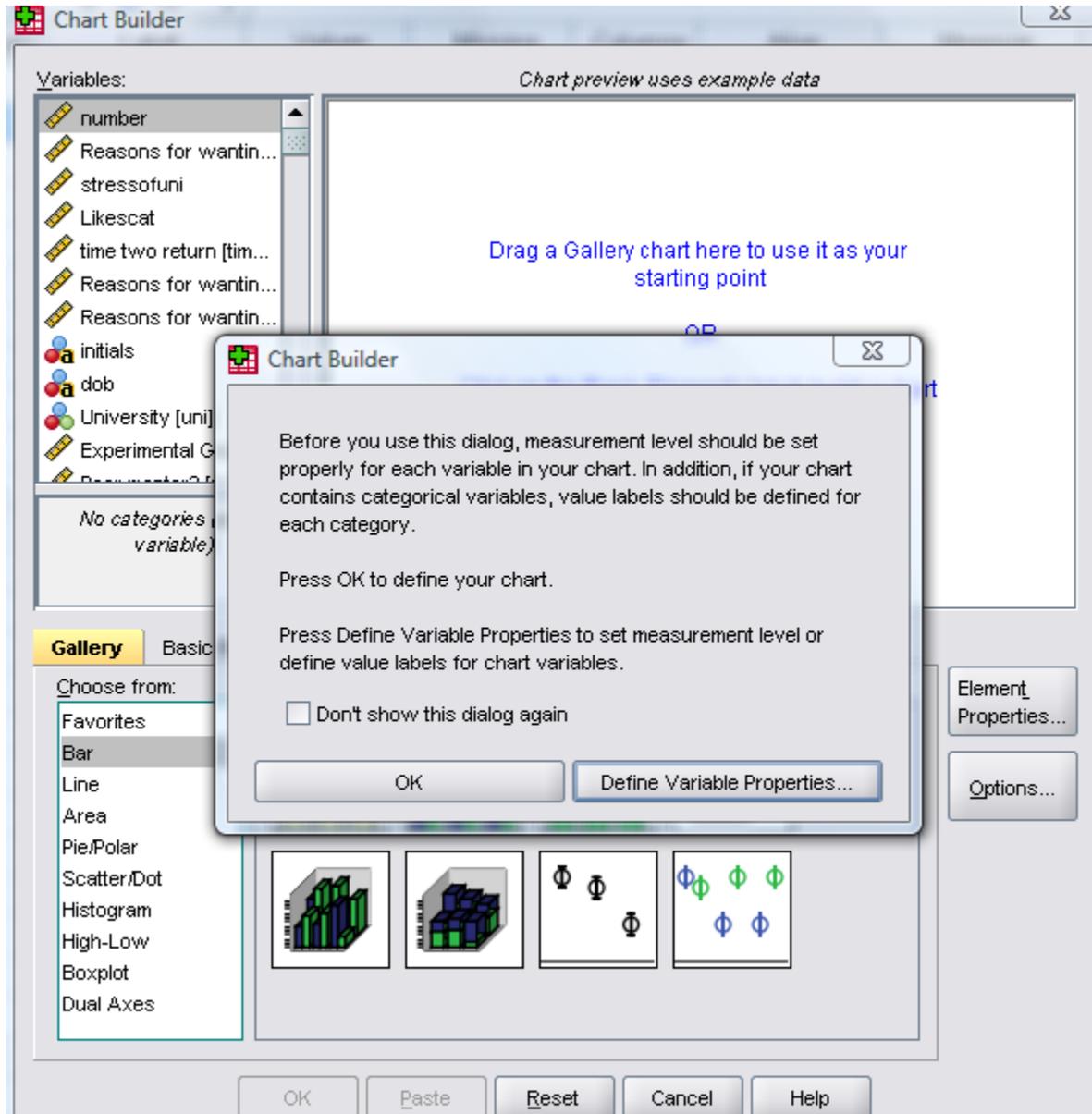In this case one row is for the IV

The second row is the DV

Fig ?:Data view for within participants effects. In this case each level of the variable is a column.
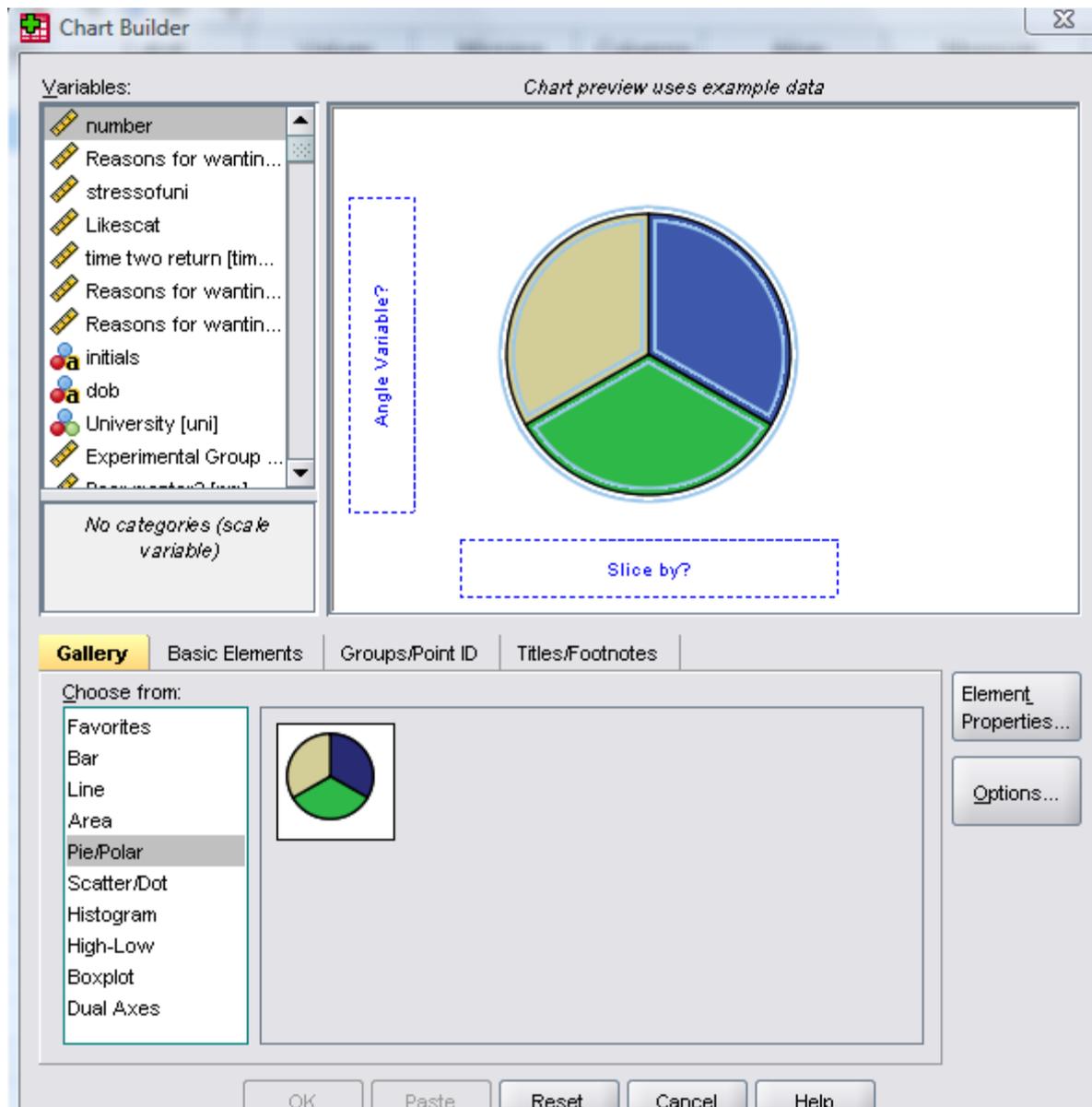
Graphical displays of Data.

Categorical data: Pie Charts

Graphs> Chart Builder> Ignore the dialog box (press OK)

Highlight Pie and then drag the pie diagram into the Gallery Chart (Fig ?)

Select the variable you are interested in looking at and drag to 'slice by'. You can change the measure from count to percentages using the element properties option. Once you have finished select OK.

After you have completed this you can change aspects of your graph such as colours used, shading, as well as adding titles etc. Simply double click on the graph in output to enter Chart Editor and play around. You can do this with all graph types.

Categorical Data : Bar Chart

Graphs: Chart Builder> Select Bar Chart > drag across the variable of interest> decide if you would like percentages or counts and press OK.

Using the Bar chart for more than just counts: (categorical IV and Continuous DV)

Between Subjects design example: assessing the difference between peer mentored and non peer mentored individuals on adjustment to university.

Graphs: Chart Builder > select bar > drag across your categorical IV onto the x axis box (in this case mentored/ not mentored). Finally drag your continuous DV (in this case adjustment) into the y axis.

You will need to go into element properties and select Mean

Additionally you can see the distribution of scores by selecting 'Display Error Bars' which I would recommend within reports.

Within subjects design example: assessing the change over time for social support amongst university students.

Graphs: Chart builder > select bar > drag both time points across. Leave the x axis blank. Remember to go to element properties and change this to mean.

Continuous variables:

Histogram. This type of graph is most often used to check the assumptions of parametric testing. IN parametric tests we argue that the data must be normally distributed. A normally distributed data has 3 properties …

To complete a histogram> Graphs > Chart Builder > Highlight histogram > drag simple histogram into the box > select the continuous variable of interest across. Because you are interested in testing your parametric assumptions you will need to check @display normal curve' in the 'element properties'

 If you want to assess the distribution of your data according to a categorical data (for example wellbeing according to gender) you can select the split histogram.

Chart editor will ask for a split variable (Gender) and a distribution variable (wellbeing)

Two continuous variables: assessing the relationship in graphical format

Graphs > Chart Builder > Scatter/ dot. Select the simple scatter > drag up > select the two variables of interest placing one in the x axis and one in the y axis. I would suggest that if you argue that one variable predicts the other have the predictor in the x and the predicted in the y. This is essential when graphing regression analysis although isn't important in correlation (because we cannot imply causality in correlation). Thus I assume that well being will predict how well adjusted you are to university. When you have created a graph I would suggest 'fitting a line'.

Line of best fit in correlation graphs:

Double click on your graph in output to bring up the chart editor. Within chart editor click the diagram corresponding to line of best fit to fit the line.
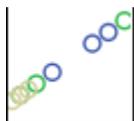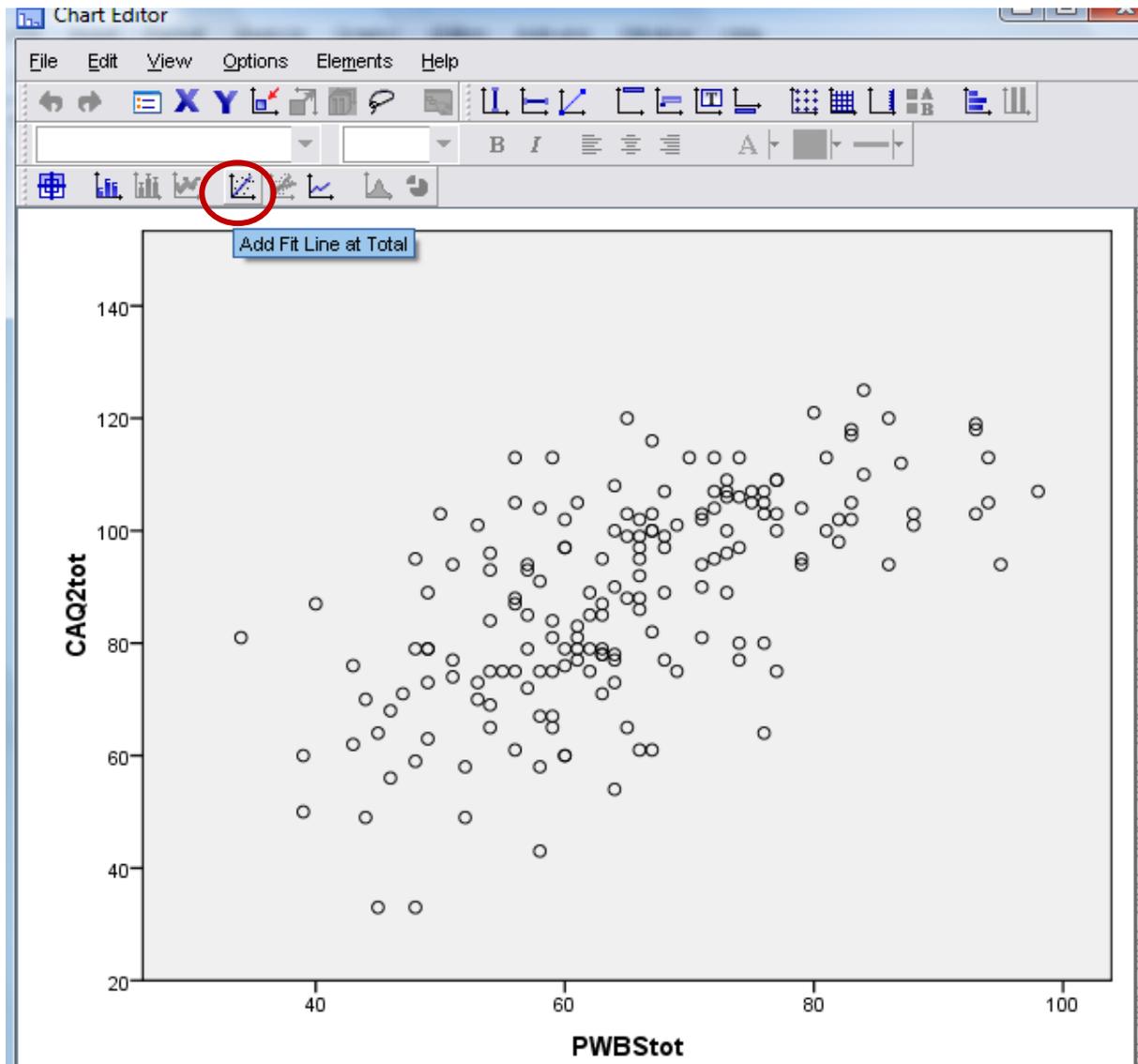
You can also produce scatter plots split by a categorical variable. For example I could assess the relationship between wellbeing and adjustment depending upon gender. In order to do this select the grouped scatter/ dot and enter your categorical variable into the 'set colour by' box that appears in the top right hand corner.

I have set color by whether or not the participants had a peer mentor.

Descriptive Statistics

To get all the descriptive statistics you need: Analyze > Descriptive Statistics>Explore.
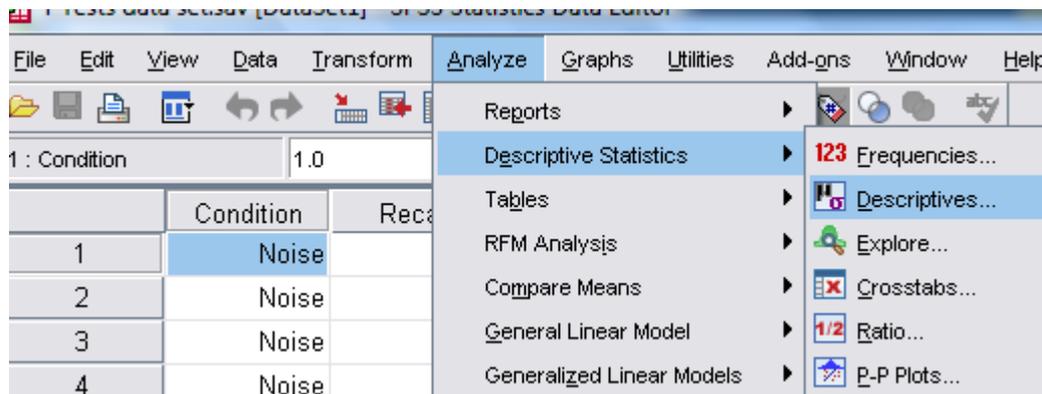
Enter the IV into the Factor list and the DV into the Dependent list.

Select plots and tick Histogram and Normality plots

Descriptives

Condition				Statistic	Std. Error
Recall	Noise	Mean		7.25	.719
		95% Confidence Interval for Mean	Lower Bound	5.67	
			Upper Bound	8.83	
		5% Trimmed Mean		7.28	
		Median		6.50	
		Variance		6.205	
		Std. Deviation		2.491	
		Minimum		3	
		Maximum		11	
		Range		8	
		Interquartile Range		5	
		Skewness		-.003	.637
		Kurtosis		-1.104	1.232
	No Noise	Mean		13.83	.796
		95% Confidence Interval for Mean	Lower Bound	12.08	
			Upper Bound	15.59	
		5% Trimmed Mean		13.87	
		Median		14.00	
		Variance		7.606	
		Std. Deviation		2.758	
		Minimum		9	
		Maximum		18	
		Range		9	
		Interquartile Range		5	
		Skewness		-.198	.637
		Kurtosis		-.937	1.232

This table can be used to find the mean, 95% CI around the mean, and SD. Additionally you can find information on the distribution of data: skewness and kurtosis.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition	Statistic	df	Sig.	Statistic	df	Sig.
Recall	Noise	.192	12	.200[*]	.938	12	.467
	No Noise	.164	12	.200[*]	.963	12	.825

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

These are statistical tests for normality of data. I read the Shapiro -Wilk and if this is significant then we have a problem with our distribution. However, in large sample sizes these become overly sensitive when small deviations produce significant results, therefore you should also visually inspect the histograms.

Because the above test is unreliable in larger sample sizes I often assess for normality using a z score calculation instead. This is based on figures found in the explore table – the skewness and kurtosis and their respective standard error terms. To test for normality using this you calculate: $z = s/se$ (i.e. skew / standard error of skew). In the above example this would be: $- 0.003 / 0.343 = 0.0175$; thus $z = 0.0175$. The general rules are: for small sample size (which is what you will be dealing with) a score of $z = 1.96$ or above indicates a problem.

—— Normal

Histogram

for Condition= No Noise

Mean =13.83
Std. Dev. =2.758
N =12

Recall
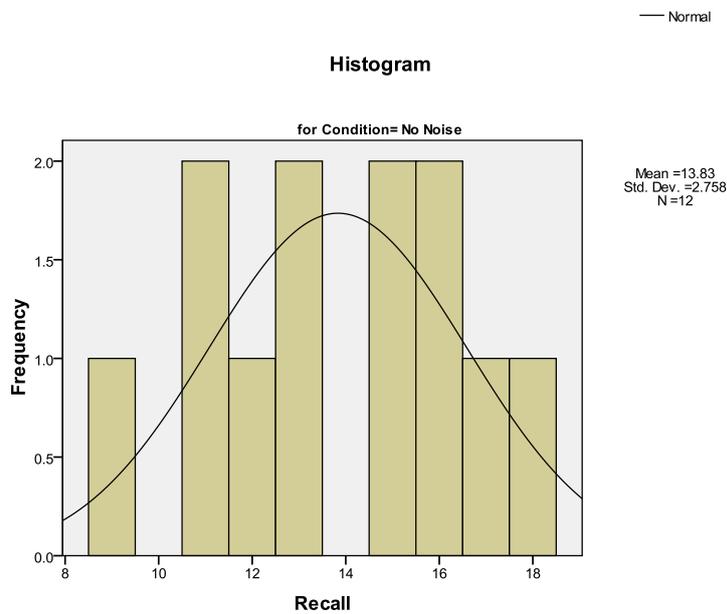
In Psychology you rarely get a perfectly normal distribution. I would be happy with the above histogram. Further investigation can be made from the normality plots: see Figure 2.

Normal Q-Q Plot of Recall

for Condition= No Noise
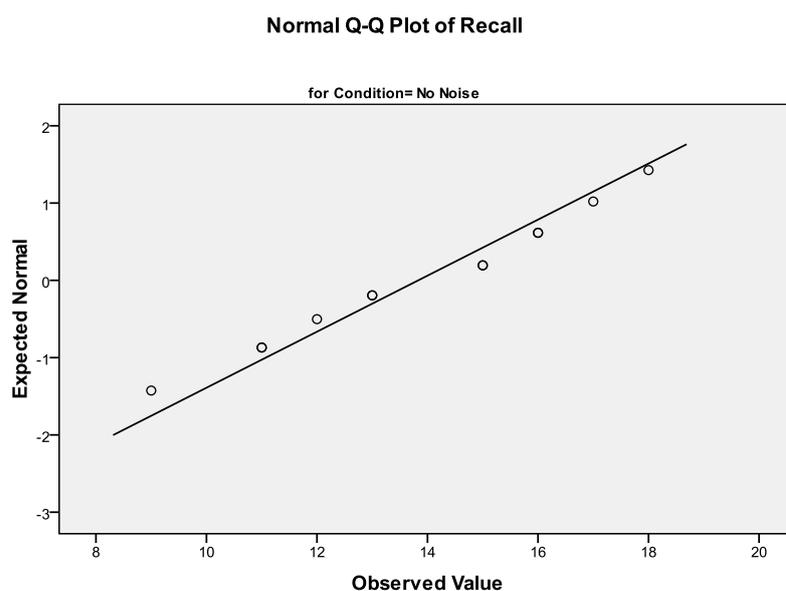
Observed Value

The Q-Q plots the observed values against those that are predicted if your data were normal (a bit like a scatterplot). Therefore your line of best fit indicates a perfect fit with normality. Any deviations from that line indicate some form of abnormality. Your aim is to have a no deviation (rarely happens) or a gentle s shape around the line without too greater deviation.