

Guide to using SPSS

[Using SPSS](#)

[Getting started](#)

[Starting the analysis](#)

[Graphs](#)

[Descriptive statistics](#)

[Tests for differences between means](#)

[t tests](#)

[Paired – t –test](#)

[One-way ANOVA](#)

[Correlation and Regression Analysis](#)

[Non parametric tests](#)

Using SPSS

IBM® SPSS Statistics software (SPSS®) is a statistical package for social science.¹ The package is designed and sold by despite its name it's applicable to most circumstances where the generation of statistics is needed. SPSS is a widely used software package that has been specifically designed to calculate and produce statistics. You can find SPSS being used by numerous government agencies, researchers and independent organisations. SPSS is versatile and can be used to compute a very wide range of statistics. SPSS is produced by IBM <http://www-01.ibm.com/software/analytics/spss/>

You will need to invest time in learning how to use SPSS, but this time will be well-rewarded. SPSS is an extremely versatile package with a large number of data analysis facilities. It can be used for basic as well as complex statistical tests. As with many modern computer packages SPSS is menu

¹ *All screenshots of SPSS in this document are reproduced with permission. Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation. SPSS Inc. was acquired by IBM in October, 2009. IBM, the IBM logo, ibm.com, and SPSS are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "IBM Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

driven and operates within a “Windows” type environment. This guide is based on SPSS release 11.5.

Getting Started

To get started click on the SPSS programme's icon, which is often found on a computer's desktop or sometimes within a folder called Programmes. Having started the programme, the first window you will see is the data entry window (you initially see a dialogue box giving you the option to open existing files – if you are a new user just close this box).

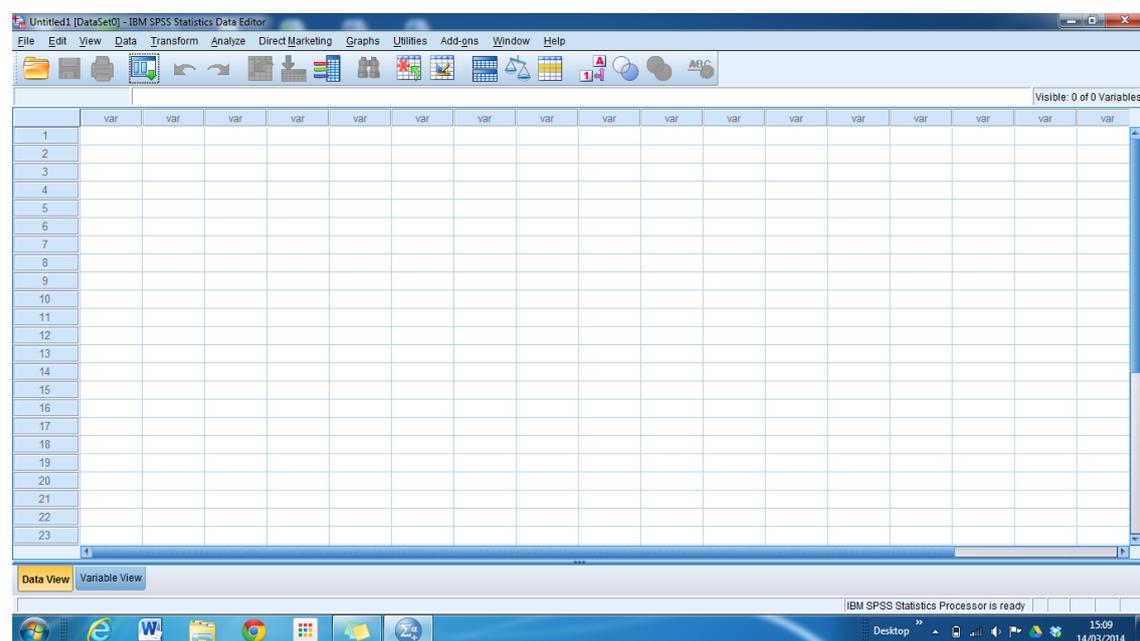


Figure 1: Data Editor

The data entry page, called the Data Editor, has a similar look to that of a spreadsheet. It has a series of rows and columns, where the data will be entered (Figure 1). In the first cell of each column is a heading. This heading will form the name for each variable that you wish to use in your analysis. When a new data file is being created SPSS will call the variables v1, v2, v3 etc. You can give each of the columns a specific name to represent the variables that will be entered. To do this double click on the variable heading, a dialogue box will appear. This dialogue box allows you to enter a specific name for the variable of each column and asks for details of the type of variable to be analysed. You can indicate for example to how many significant places after the decimal a variable should be entered and whether or not the variable will be entered as a number or as an alpha numeric. An alpha

numeric is the term given to variables that are signified in the data set as letters.

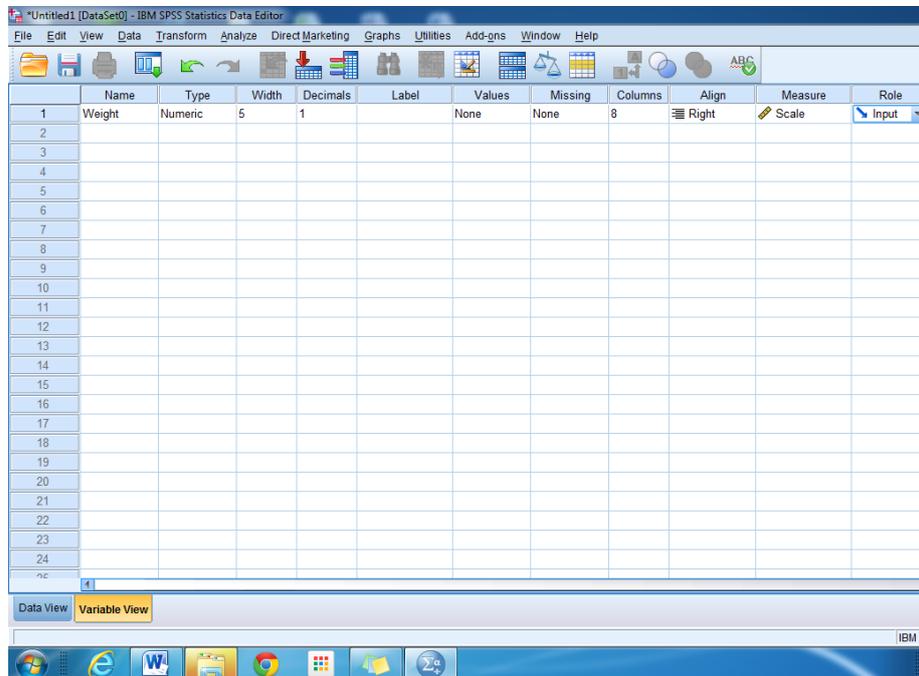


Figure 2: Data Editor, Variable View

You can also indicate the width (number of columns a variable should occupy – this is a useful feature that helps to reduce data entry errors. The last column allows you to indicate what measurement scale the variable was recorded on; there are three options, scale (ratio /interval-see Chapter 3), ordinal and nominal. The column with heading Label allows you to give your variable a tag or name, which will then be used in any analysis output produced by SPSS. The column missing should be used if you want to use a specific value (-9) to signify that there is no available data to enter for that particular cell. Some researchers use this facility if they wish to signify that a particular item of data was not collected, i.e. that its absence has been noted. SPSS will normally not use these values in analysis. If some of the information for a particular case is missing and you have not specified a particular “missing value number”, leave the cell blank, do not enter zero, as the computer will read this as a zero, not a missing value.

Take care with how you define your variables because SPSS will interpret them very specifically. Once you have defined your variables you can move

on to entering your data. To move back to the data entry screen click on the first column of the page that is highlighted in grey.

When you enter data each case has a row – do not use the same row for more than one case. It is possible to enter data directly without specifying its type, SPSS will then interpret the data you have entered and define the nature of the variable for you. The variable name will be given by SPSS as VAR, and a number will also be added in sequence, so the first variable will be called VAR0001 and the second VAR0002 etc.

It is possible to enter files directly from Excel. You may want to try this with some of the data that we have supplied. The two files of data supplied are that generated by the studies described in Chapter 5 of the book. To enter data from an external file from the data entry window select File then Open then Data; a dialogue box will appear, you must tell SPSS both the file type (in this case Excel) and its location on your computer system.

Starting the analysis

Once you have finished entering, you are ready to begin the analysis. As suggested in the book, the first thing that you need to do is to produce graphs and descriptive statistics for your data. This is so that you can get a better understand and a “feel” for the data.

At the top of the Data Editor, you will see a menu bar with a series of commands (Figure 1). You will probably make most use of the commands **Analysis** and **Graphs**. The other commands are concerned with manipulating the data set.

Graphs

SPSS is not a graphics package and if you are interested in producing graphs for presentation or publication we would not recommend using SPSS. However, the graph facility of SPSS does allow you to produce graphs quickly such that you check over and review your data.

From the Data Editor if you click on Graphs you will see that there is a large range of options. There are options that cover all the graph types described in the book (Chapter7). If you want to produce a frequency histogram (Chapter

9), click on the option **Histogram**. A dialogue box will appear containing a list of your variables down the left-hand side. Click on the variable that you wish to use and then the arrow button in the middle. Your selection will now appear in the right hand box. There is also a check box to indicate if you want SPSS to superimpose on the graph a normal distribution curve using the mean and standard deviation for your sample. This is a useful option as it allows you to gauge relatively quickly whether or not your data are likely to be normally distributed (Chapter 10). Having made your choice you can now select **OK**. Note that an alternative to selecting **OK** does exist and this is to choose **Paste** instead. If you select **Paste**, SPSS will put the commands to produce the histogram into a separate file. By using the paste facility, you can build up a series of commands to run all at once. This facility is very useful if you have several sets of data that wish to analyse in the same way. When you are getting started with SPSS however it is a good idea to just select **OK** and let the command to be executed immediately.

Once you click on **OK** a new window will be produced, with a record of files where the output from your analysis is stored down the left hand side (looking like a file tree) and the actual output of the last command in the right hand window. If you have requested a histogram, this is what will be displayed. If you wish to edit the graph (if it is going to be used for publication or presentation) double click on the graph and the graph editor will appear. The graph editor allows you to manipulate the appearance of the graph; we will not go into details of how to use this facility and suggest that you spend some time exploring it for yourself. The graphs can be printed, by clicking on the graph and the selecting print from the file menu.

Having finished working with your graph if you wish to close the output window (so that you can get back to the Data Editor, do so (as with most windows packages) by clicking on the cross (**x**) at the top right hand corner of the window. You will be asked if you wish to save the file. If you do, complete the dialogue box, entering an appropriate file name. Note that one feature of SPSS is that it saves the data file (from the Data Editor) any output files (graph and results) to separate files. Thus take care; saving the output file will not ensure that the data file is also saved.

The other graphing options operate in a similar way, you select the graph type you wish from the graph menu and then complete the dialogue box. The dialogue boxes will vary depend on the specific nature of the graph type. For

example to produce a scatter graph SPSS needs to know which variables are being investigated and which you wish to use for the x-axis and which you wish to use for the y-axis. Some of the graph options also have more complex variations, for instance, in the scattergraph option you can produce figures that use more than two variables. Take time to explore this option as you become more comfortable with using statistics and the terminology associated with them, so using SPSS graphics capability will become easier.

Descriptive statistics

SPSS can produce a wide range of descriptive statistics. The descriptive statistic functions of SPSS can be obtained by selecting **Analysis** from the Data Editor and then choosing **Descriptive Statistics**. You will see a number of options we will not discuss them all here, but will focus on those that are useful, easy and straightforward to use.

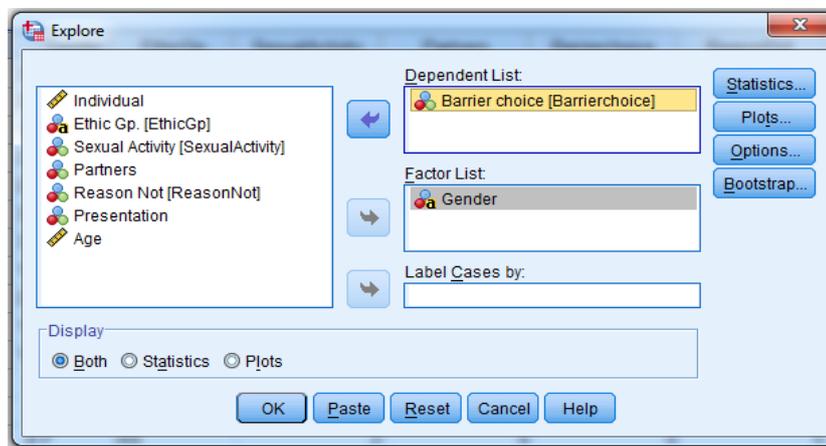
Selecting the option **Descriptives** under the **Descriptive Statistics** option will bring up a dialogue box. Once you have selected the variables you wish to analyse, you can then click options. A second dialogue box will appear which allows you to select certain descriptive statistics – you will recognise all of the options if you have read chapters 6 and 7 of the book. Select the options that you want by checking the appropriate boxes. Note that SE of the mean is also included even though it is not strictly speaking a pure descriptive statistic. Once you have made your selections select Continue, this will return you to the original dialogue box, if you now click on OK, the output will be produced. Note that this system of menus and a hierarchy of dialogue boxes is the main way in which, you interact with the SPSS computer programme.

The output is produced as a table, one issue with using SPSS is that if you request a wide range of descriptive statistics and/or a large number of variables the output can become very crowded and therefore difficult to interpret. If you are looking at large sets of data with large numbers of variables, consider performing your requests in more than one stage; for example you could request **Descriptives** for the first half of your set of variables, and then repeat for the second half.

Another useful option under the **Descriptive Statistics** section is the **Explore** option. If you click on **Explore**, you will again see a dialogue box with a

number of options. A feature of this option is that it allows you to produce descriptive statistics for your variables on a category by category basis. For example, for the data collected for the study on sexual health (Chapter 5) we may want to look at summary statistics for each ethnic group separately. We may want to produce summary statistics for choice of contraceptive by gender. In the **Explore** dialogue box (Figure 3) you thus have an option of selecting not just the variables that you wish to produce statistics for but also to select variables to categorise the data set by; SPSS refers to these variables as factor.

The explore option, by default, will produce a wide range of summary statistics most of which were discussed in Chapter 6. The default mode will also produce some plots of the data. These plots are known as box and whisker plots and show, median, quartiles, and extreme values within a category. The box displays the middle 50% of cases, the line in the middle of the box is the median, the lines outside the box (whiskers) represent values up to 2 standard deviations away from the mean, all other values are indicated as extremes (outliers).



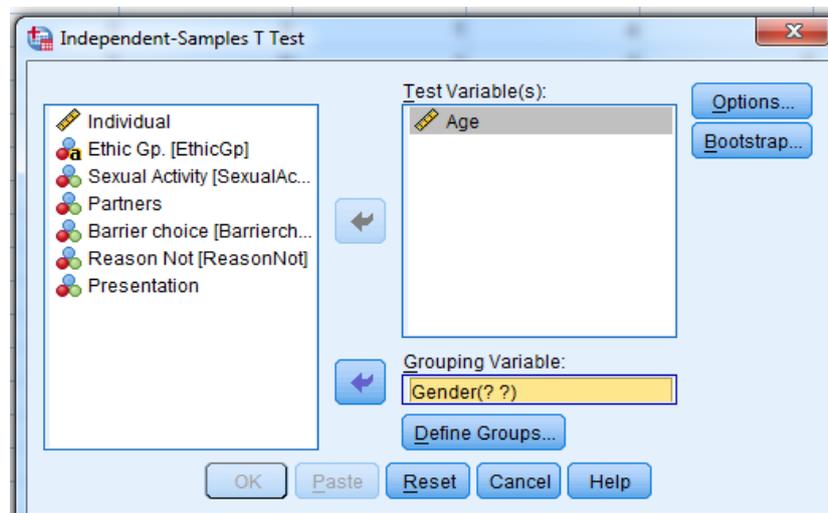
Tests for differences between means

This section looks at how you can use SPSS to test for differences in the means of samples. It will first look at test for two samples and then test for more than two samples.

t tests

Figure 3: The Explore dialogue box

To perform these statistical test (Chapter 12) from the **Analysis** menu from the Data Editor select **Compare Means**. Here you will see a number of



options, and within those options **T-tests**. The **Independent-Samples T test** command will give a student-t test and the **Paired-Samples T test** will give the paired t tests for a repeat measures design. Clicking on these options will produce a dialogue box (Figure 4).

For the t-test, you will need to select a **Test variable**, i.e. the variables that you want to perform a t-test for and you also need to define the experimental or test groups, SPSS calls this the grouping variable. In the example shown in Figure 4, we are testing to see if there is a significant difference in the age of the Males and Females in our sample. Gender (sex) is thus the **Grouping variable**. The grouping variable distinguishes the two samples being

Figure 4: dialogue box for student's t-test

compared. Once you have selected the grouping variable, you must also define the range of values that will be used to distinguish between groups. In the example above, this is obvious because it will be males and female. In the symphadiol experiment however there are, for example, four treatment groups and we may want to compare any two pairs of these groups; SPSS needs to know which pair to compare.

The output produced will be a table with summary statistics and a table with test related to the t test. Below is an example related to the symphadiol experiment, comparing the control group and treatment group two (Note this is not an appropriate test to use if all the treatment groups are to be compared against each other (see Chapter 13). The output table will include two values of t; one value assumes the variances of the samples is not different the other takes into account this variation.

Independent Samples Test

Tests for equality of variance.

		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	
Weight loss	Equal variances assumed	.511	.481	.430	28	.670	1.27	2.943	-4.762	7.295
	Equal variances not assumed			.430	26.87	.670	1.27	2.943	-4.773	7.307

Test statistic

Significance level
(i.e. the P value)

Figure 5 Output table for students t-test.

Paired – t –test

If you select the Paired-Samples T test option you will see that the dialogue box is slightly different. You are asked to supplied paired variables, thus SPSS treats the before and after measures of the test as being different variables not simply different samples. You must therefore have at least two entries for each case that you have. If for instance you have measured the effect of a therapy on a group of participants' respiration rates, you will probably have measurements before the intervention and after. The measurements before the treatment form one of the variable and the measurements after the other variable. The output from the paired-t-test is essentially the same as for the student's-t-test. When selecting the variables (paired samples) from the variable list you must choose both variables before clicking on the arrow to select the variables. The paired t-test is sometimes used in experiments where measurements are not repeated on the same participants; in such case, participants are matched, so they are very similar.

One-way ANOVA

Under the compare means menu you will also see the command one-way ANOVA as the name implies this command allows you to perform a one-way analysis of variance (Chapter 13). The dialogue box for this command is very similar to that for the independent sample t-test. The main difference with this command is that you do not need to specify the individual samples that are to be compared as SPSS will assume that all the groups within the variable specifying the groups will be used. In addition, there are a wider range of options (Options command), including summary statistics and test for equality of variance. There is also an opportunity to specify which post-hoc test will be used. The other difference with command is that the output is very different from that to the t-test as an ANOVA table is produced (Chapter 13).

Table 1: Output from One-way ANOVA command

Weight loss

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	428.133	3	142.711	2.388	.079
Within Groups	3346.800	56	59.764		
Total	3774.933	59			

Correlation and Regression Analysis

Correlation

To produce correlation select **Correlate** from the **Analysis** menu. You will have three options and for simple correlation's select bivariate (i.e. two variables). The menu produces is straightforward and all you need to do is select the variables to be correlated and which correlation methods you wish to use (Peasons, Kendal's or Spearman's). You can select more than one method if you wish. Correlation statistics will be produced for each potential combination of the variables that you selected. You can also select if you wish the significance level to be expressed for a one tailed or a two-tailed test.

Regression

For simple regression from the **Analysis** menu select **Regression** and then **linear**. Notice the wide range of regression techniques that are actually available. The regression command has a wide range of options as it also the menu used to produce multiple regressions (Figure 6).

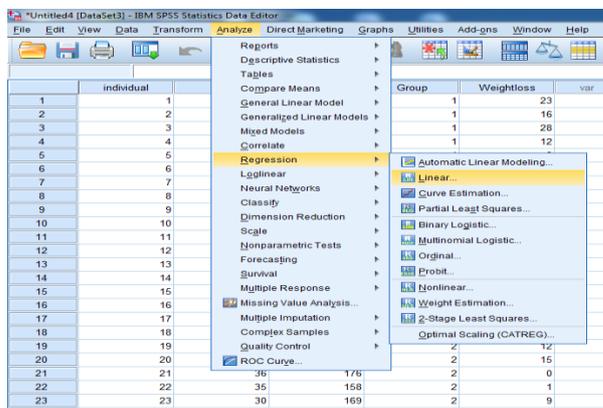


Figure 6, the Regression dialogue box

Once you have selected, **Linear** a dialogue box will appear. To produce a regression analysis, select from the list of variables the dependent and then the independent variables (see Chapter 17), then click on **OK**. You can explore the additional options, as you become more familiar with using statistics (Figure 7).

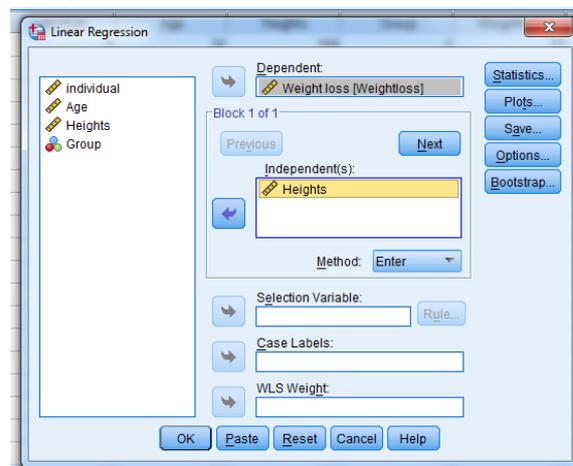


Figure 7: the Linear Regression dialogue box

The output from this command is in the form of four tables; the first is a table reminding you of the variables that were used. The second table gives values for the regression coefficient the coefficient of determination and their significance. The second table is an ANOVA table. The ANOVA uses the slope of the regression as a variable and has little relevance for regressions where just two variables are involved. The significance of the F value, of the ANOVA, should match with that for the slope given in the third. The third table gives the results of t-tests of the slope and the intercept (test to determine if the slope and intercept are significantly different from zero).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.273(a)	.075	.059	7.761

a Predictors: (Constant), HEIGHTS

Correlation Statistics

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-22.476	16.685		-1.347	.183
	HEIGHTS	.202	.093	.273	2.162	.035

a Dependent Variable: Weight loss

Value of the slope

Value of the intercept

Figure 9: the second and fourth output tables from a regression analysis

Non parametric tests

From the analysis menu select **Nonparametric Test**, a new menu will pop-up with a list of further options (Figure 10).

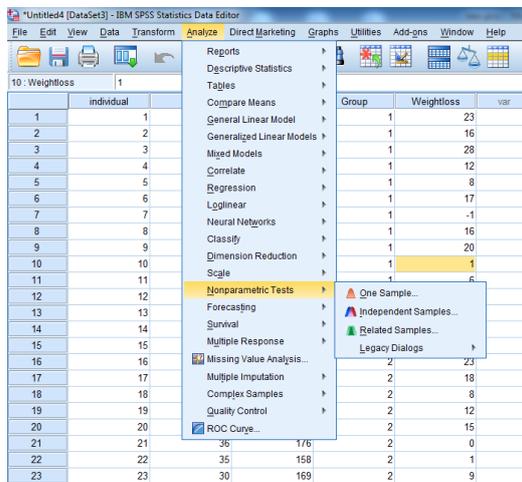


Figure 10: Selecting a non-parametric

Clicking on the **Independent Sample** command will produce a dialogue box, like the one below. The tests accessed from this command are designed to test for difference between two samples that have been sampled independently of each other (Figure 11).

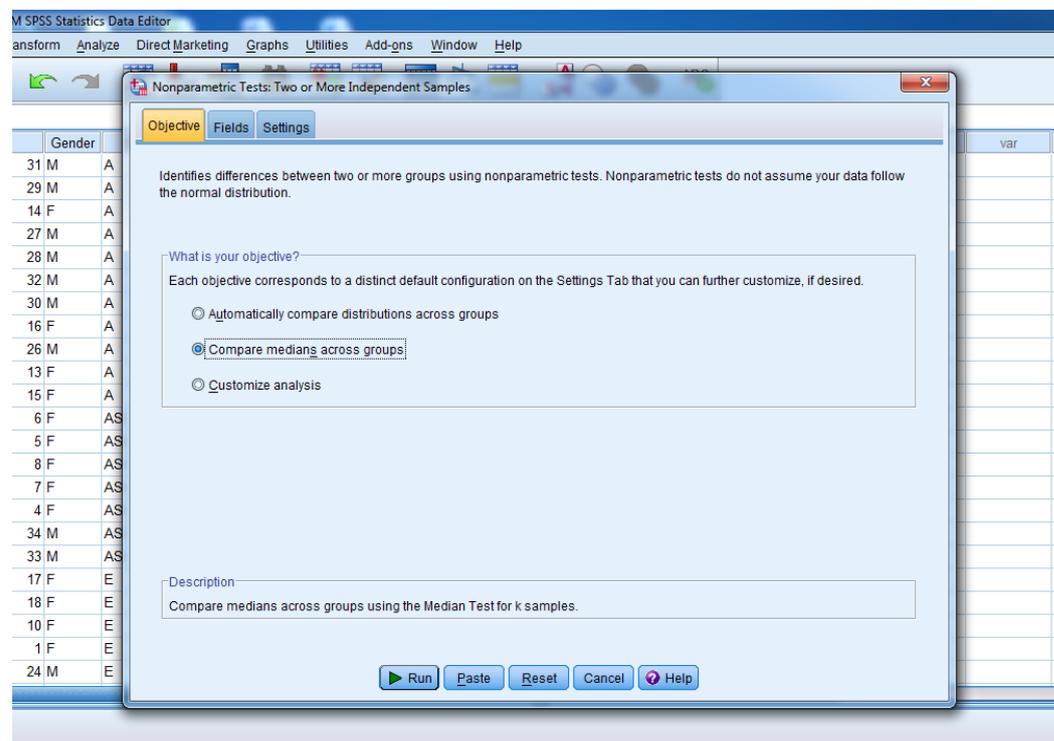


Figure 11: non-parametric test, two or more independent samples

Within this box, there are three radial buttons and 2 tabs. In the first tab you must select the objective, this lets you either compare the distribution of your samples or compare the medians, SPSS will provide you with the appropriate choice of statistical test, which helps avoid selecting the wrong test. In order for SPSS to do this for you, when you first entered your data, you will have needed to enter the type of variable scale correctly, on the variable view, data entry window. In Figure 11 we have selected the 'Compare medians' radial button and we now need to select the **Fields** tab, which will allow us to determine which data sets we wish to use.

In this example (Figure 12) we have selected barrier choice and age, that is we are looking to see if there is a difference in the median age between different barrier methods of protection.

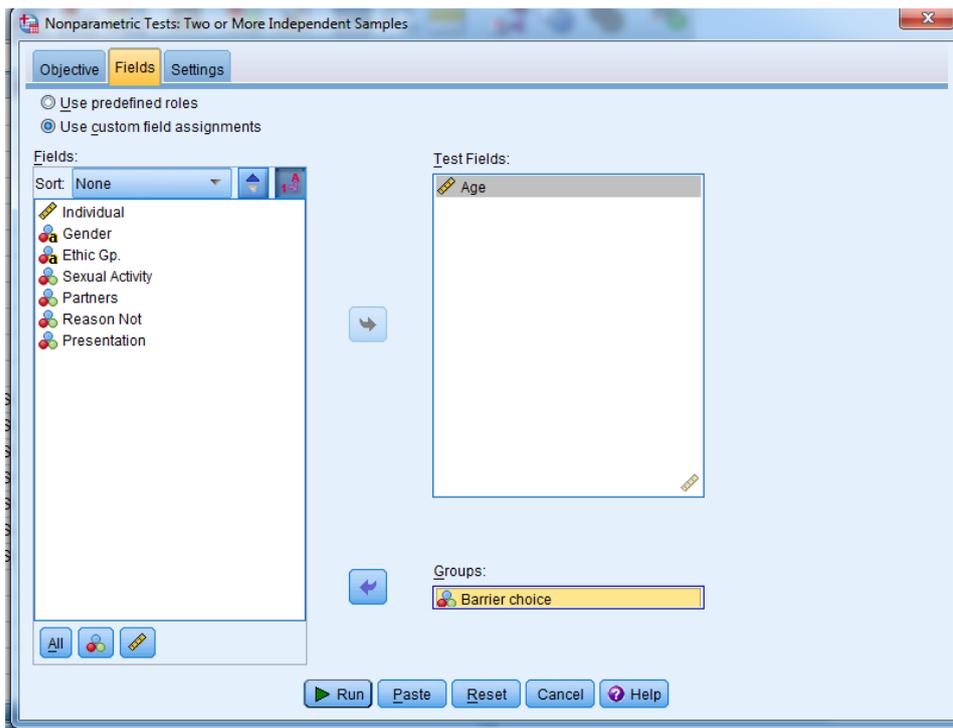


Figure 12: non-parametric test, selecting the variables to use.

The settings test gives you a range of appropriate options – in this case the Kruskal-Wallis test has been selected. (Figure 13) When you select Run and output a window will be generated (Figure 14).

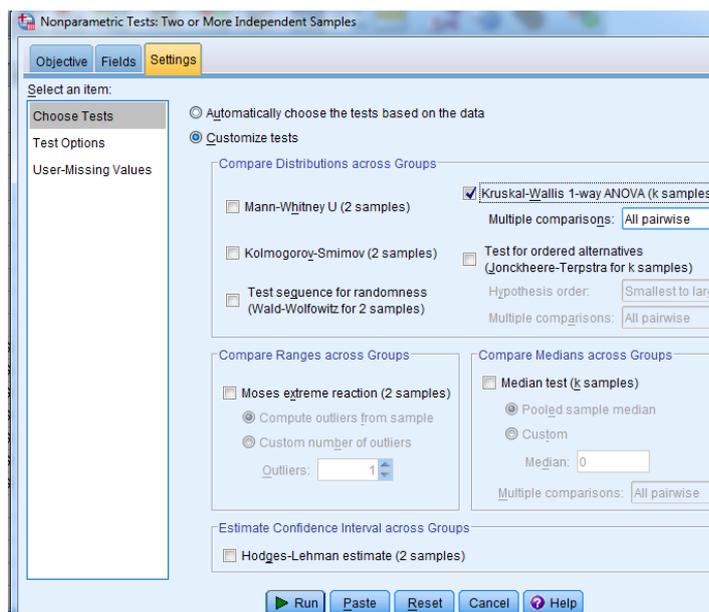


Figure 13: non-parametric test, selecting the test to use.

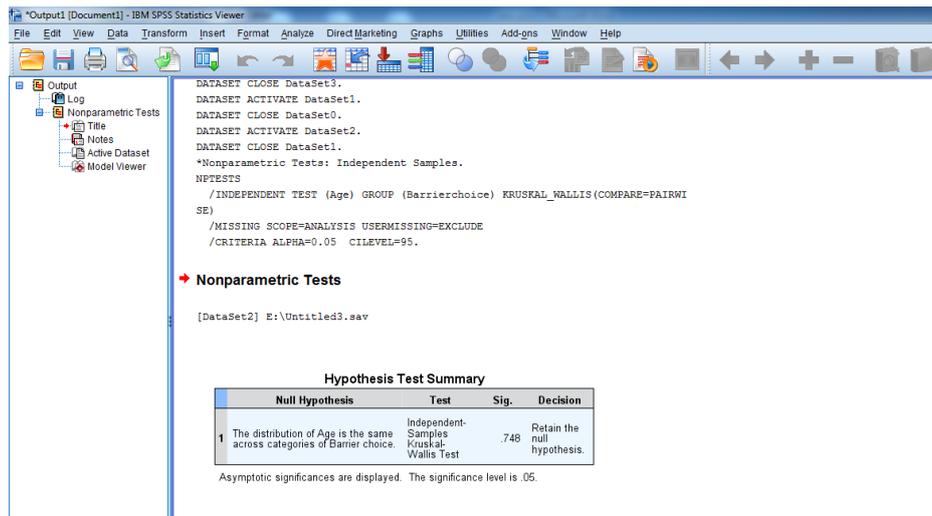


Figure 14: non-parametric test, example of output window