SIMPLE STATISTICAL ANALYSIS IN SPSS FOR WINDOWS (VERSION 8.0/9.0)

This document contains a series of exercises which give an introduction to analysing data in the SPSS for Windows statistics program (version 8.0/9.0).

AUTHOR Information Systems Services
University of Leeds

DATE August 1999

EDITION 2.0

TUT 53

50p

Contents

Format Conventions

In this document the following format conventions are used:

Commands that you must type in are shown in **bold Courier** font.	**WIN31**
Menu items are given in a **Bold, Arial** font.	**Windows Applications**
Keys that you press are enclosed in angle brackets.	**<Enter>**

Feedback

If you notice any mistakes in this document please contact the Information Officer. Email should be sent to the address **info-officer@leeds.ac.uk**

Print Record

This document was printed on 17-Nov-99.

Introduction

About Data Analysis

The end of any statistical process is always the results – what does the data mean? What do the graphs look like? What are the implications? This document aims to tackle the process of performing an SPSS analysis. The skills acquired from this set of exercises should enable you to start tackling other sorts of analyses on your own.

Before a computer can work on your data, you have to tell it how the data is structured. More to the point, SPSS itself often expects the data to be structured in a particular way in order to carry out the analysis correctly. It is important that you understand what SPSS is expecting before you give it the data to analyse. Sometimes this may involve restructuring your datasheet to get the information into the right format.

This document will take you through some of the basic tasks that you may need to accomplish when analysing your data in SPSS for Windows. This document will teach by demonstration with a single common sort of analysis, running a One-Way Analysis of Variance. It is not intended to be an exhaustive description of this or any of the other analyses SPSS can perform – that's what the manuals are for! The tasks have been designed in such a way that you are advised to complete a task and its exercises **before** proceeding to the next one.

Requirements

It is assumed that you already know how to login to the Novell network and run the Microsoft Windows operating system. It is also assumed that you know how to run SPSS for Windows from the Windows desktop, create a new data sheet, enter text into a cell, run a simple analysis and print it out. If you do not yet know how to achieve these operations it will be necessary for you to read and work through at least one, and preferably two, of the following:

Getting Started with SPSS for Windows	(BEG 14)
Introductory Exercises in SPSS for Windows	(TUT 51)
SPSS Tutorial, found under the Help menu.	

Documentation

If you require further information on the facilities in SPSS the following document is also available:

Data Operations in SPSS for Windows	(TUT 52)

References are made in this document to the SPSS manual *SPSS Base User's Guide* which is available for reference from the Information Systems Services Help Desk.

About Data Analysis

Analysing data is neither a simple nor a straightforward process. Any statistician will tell you that in order to do it right you should decide what analysis you are going to do *before* you perform the experiment. In fact, the whole business of experimental design and its subsequent analysis is so integrated that if you have not yet done your experiment you should go and chat to a statistician about your design, what you are hoping to achieve or demonstrate, and what analysis you should be doing afterwards.

Most of you will not be in that happy position. You will already have collected your data, and you may or may not know what analysis you should use on it. Here are some guidelines and how they apply to SPSS.

Data Errors

Getting your data from the experiment to the statistical package is a process fraught with dangers. If you're doing a survey people might lie, or be mistaken, or misunderstand the question, or you might have misheard them when they answered. If you're doing a scientific experiment something might have gone wrong with the procedure, you might have copied the results down wrongly or the equipment might be faulty.

Then you've got to copy the data into a computer. This is often done late at night and in a rush. Numbers may be mistyped or misread. Whole cases might be left out as your eye misses out a line.

Many of these errors are avoidable, or at least minimisable, but not all of them, so it makes sense to first look at your data for what it is. To do this we employ the SPSS `Explore` procedure. This is a collection of tests designed to help you look at your data. Essentially you want your data to show some sort of regular or consistent pattern. Any data point clearly outside this pattern might be a mistake and needs to be checked.

Assumptions

Most statistical tests rely on certain assumptions being made about the data being tested. The most common assumptions are:

The data is normal

When statisticians say this, they mean that if you plot it on a histogram it looks bell-shaped. Much statistical theory relies on your data looking like this. The simplest test, therefore, is to plot your data on a histogram. SPSS also allows you to superimpose a 'Normal Curve' which will give you an idea of whether the data is symmetrical or not. There are also a number of other *Normal Plots* available under `Explore...` which will help you decide whether your data is normal or not.

The variances are equal

Especially if you are testing one set of data against another, like in a t-test of Analysis of Variance. The most important test here is the *Levene Test* which is available through Explore and the Analysis of Variance procedures.

The relationship is linear

To test this, plot one variable against the other on a scatter plot. If they form a sort of straight line then you know you're safe. If they don't then you may need to do something called *Transforming* to one set of data to get it linear. See a good statistics textbook for details on this.

How SPSS Expects Data

One of the main problems people experience when using SPSS for Windows is knowing how SPSS expects you to present your data for it to analyse. Often the only way to discover this is to pore over the manuals and try a worked example that you already know the answer to. For more complicated analyses it may be a case of getting a good statistics textbook and typing in an example from there, then seeing if the results SPSS gives you are the same as the ones the text book gives you.

A confusion often arises when we want to compare two sets of data directly. Take the example of someone doing a simple experiment on some plants, one set of which are given one type of fertilizer, whilst the other set are given a different type of fertilizer. Let's say that we are using weight as the measurement of how well these plants are doing. The classic way to record this data is like this:

Fertilizer 1	**Fertilizer 2**
45	22
56	31
88	28
…	…

Depending on the analysis you want to do (probably a t-test of some sort in this instance) many statistics packages will accept your data in precisely this form. This is particularly true of statistics add-ons to spreadsheets, such as Astute for Excel. This is not true of SPSS.

SPSS, because it was written by statisticians for statisticians, doesn't see the measurement of weight as two separate variables in the way it is presented above. SPSS looks very much more at *cases*, and in this instance the *case* is the *individual plant.* Now plants have attributes, such as height, weight, colour, species, and what fertilizer they were grown in. In SPSS's terminology, these are *factors* that might or might not affect something or be affected by something. In our case, the type of fertilizer the plant was grown in is a factor of the plant that could affect the plant's weight.

Consequently when SPSS wants to see the data, it expects to see one variable for the data under analysis – in our case the weight variable –which is the d*ependent variable* in that its value for any one case is (we think) dependent on the value of another variable. SPSS then wants to see a second *independent variable* that groups the first one. Here it would group the plants into the two groups treated by the different fertilizers. The weight variable could have any value because it is a continuous variable, but the fertilizer variable will have certain discrete values – either the plant was in one type of fertilizer or it was in the other one – and these are your factors. Diagrammatically, SPSS expects the data to look like this:

Dependent (Weight)	**Factor (Fertilizer)**
45	1
22	2
56	1
…	…

Consequently when you move data into SPSS you may need to move your data around so that all the weight data is in one column, and you then create a fresh column to categorise what the data refers to. This will seem counterintuitive for simple t-tests, but overall it does make sense.

Task 1 Getting Started

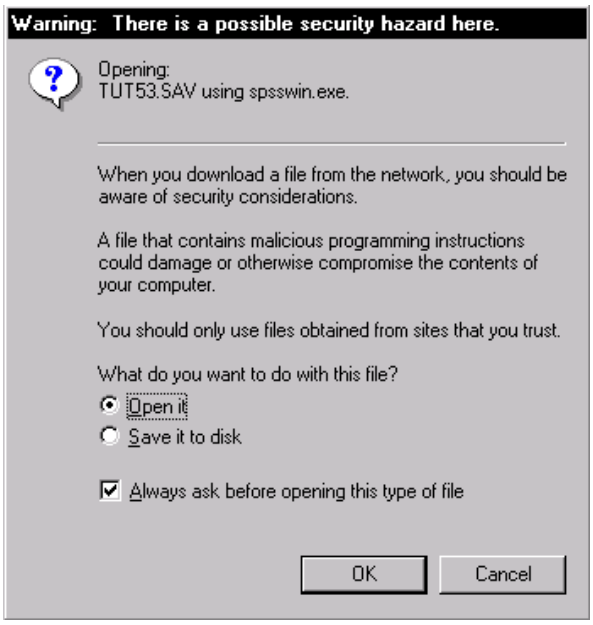
Objective An existing datasheet will be used for this exercise. This first task will copy it into your own home directory.

Instructions You will ftp the file from the ISS web site.

Comments You will need this file in order to complete these exercises.

Activity 1.1 Open a web browser such as Netscape and go to URL :
`http://www.leeds.ac.uk/iss/documentation/tut/index.html`

Activity 1.2 Scroll down to TUT 53 and click on the **`tut53.sav`** link, the dialog box below will appear. Select **Save it to disk** and click **<OK>**.



Activity 1.3 From the **Save As ..** dialog box choose a suitable directory to store the file in and click **<Save>**. Close the web browser.

Activity 1.4 Locate the SPSS icon from the **Statistics** menu and double-click the icon to open SPSS. After a short period the SPSS data editor window will be displayed.

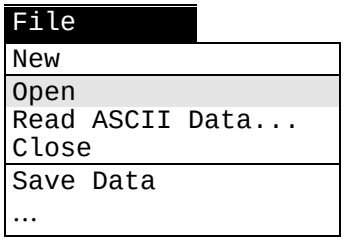
Task 2 Opening the Example File

Objective To open the example sheet from within SPSS.

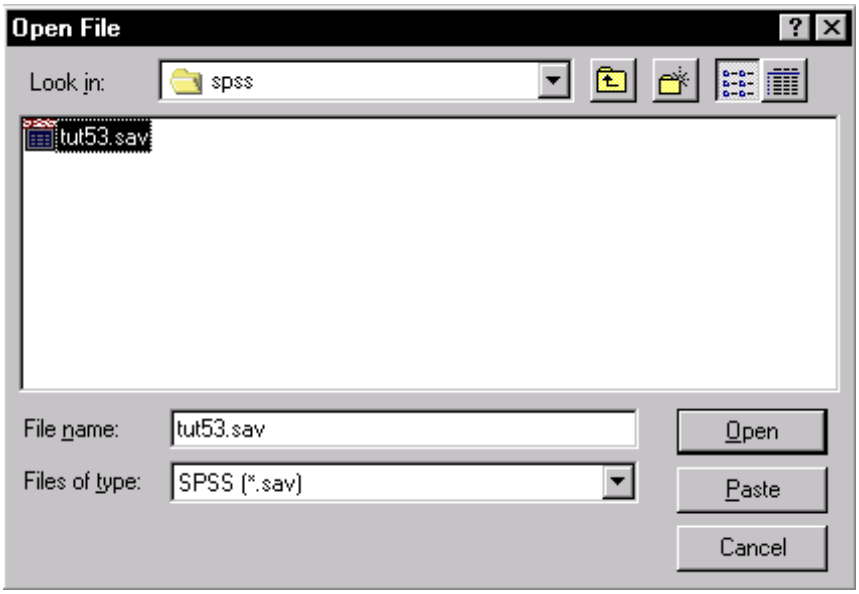
Instructions You will use the **Open** command from the **File** menu, or the **FileOpen** button on the toolbar.

Comment Opening existing files is more common than opening new ones. Fortunately the process for opening files in SPSS is the same as it is for practically every other Windows program.

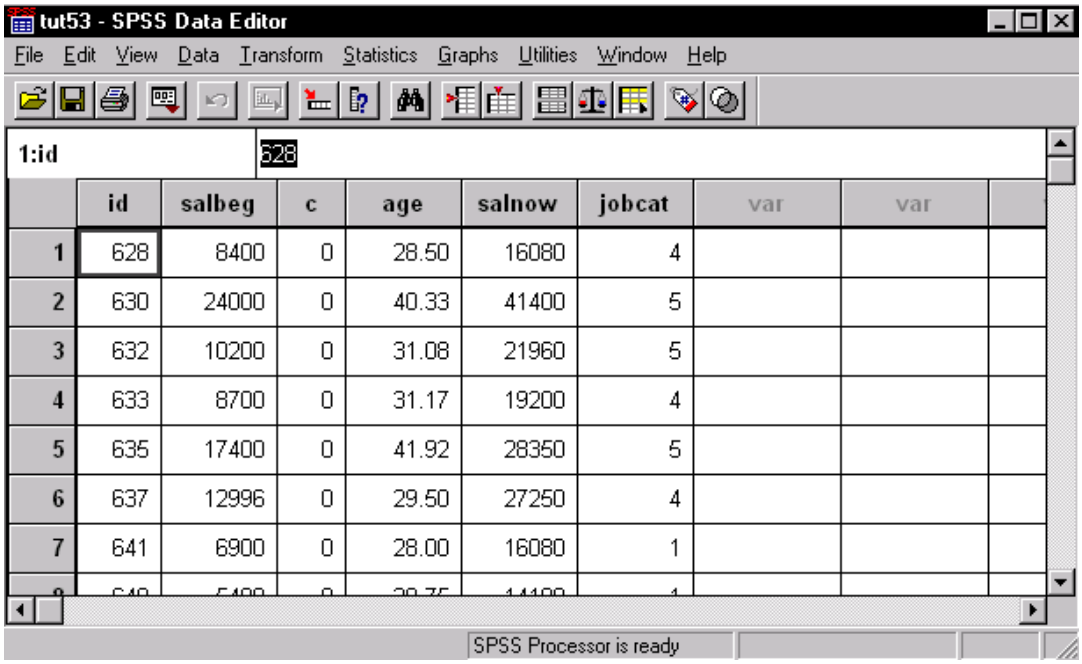
Activity 2.1 Select the **FileOpen** button on the toolbar  , or else click on the **Open** command under the **File** menu.



Activity 2.2 The Open dialog box is displayed as shown below:



Select the directory in which you saved tut53.sav in Task 1, select the file **tut53.sav** and click on **<Open>**. The file loads as shown.

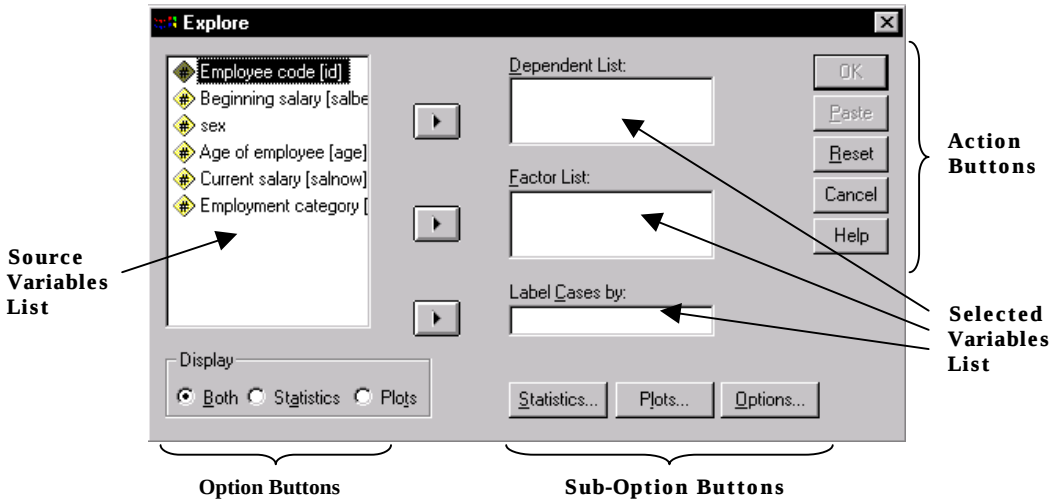

tut53 - SPSS Data Editor
File Edit View Data Transform Statistics Graphs Utilities Window Help
1:id
628
id salbeg c age salnow jobcat var var
1 628 8400 0 28.50 16080 4
2 630 24000 0 40.33 41400 5
3 632 10200 0 31.08 21960 5
4 633 8700 0 31.17 19200 4
5 635 17400 0 41.92 28350 5
6 637 12996 0 29.50 27250 4
7 641 6900 0 28.00 16080 1
SPSS Processor is ready

Task 3 Running Explore

Objective To check your data for accuracy and normality.

Instructions You examine your data using the **Explore** command.

Activity 3.1 From the **Statistics** (**Analyze** in version 9.0 of SPSS) menu click on the **Summarize** (**Descriptive Statistics** in version 9.0) command and then from the sub-menu choose the **Explore...** command. The following dialog box appears:



You will have encountered the basics of analyses boxes before in the document *Introductory Exersises with SPSS for Windows*.

The list on the left is your *Source Variable List* of all the variables which can be used in this analysis (here it only includes the numeric variables).

The three boxes in the middle entitled **Dependent List:** and **Factor List:** and **Label Cases By:** together consitute your *Selected Variable List*. These are the variables you will actually use in the analysis.

Finally you have a range of buttons around the box which will either modify your analysis or actually do something with it.

Activity 3.2 Select the variable **`Current salary [salnow]`**, and click on the pushbutton next to the **Dependent List:** box. The variable salnow should now appear in the **Dependent List:** box.

Activity 3.3 Select the variable **`Employment category [jobcat]`**, and click on the pushbutton next to the **Factor List:** box.

Click on the **<OK>** button in the top right corner of the dialog box to run the Explore procedure with the default options. You will see on the status bar the various procedures SPSS is running as it performs the analysis. You will also see how many cases it has processed for each procedure as it happens. When it has finished you should see the Output Window in the front of the screen.

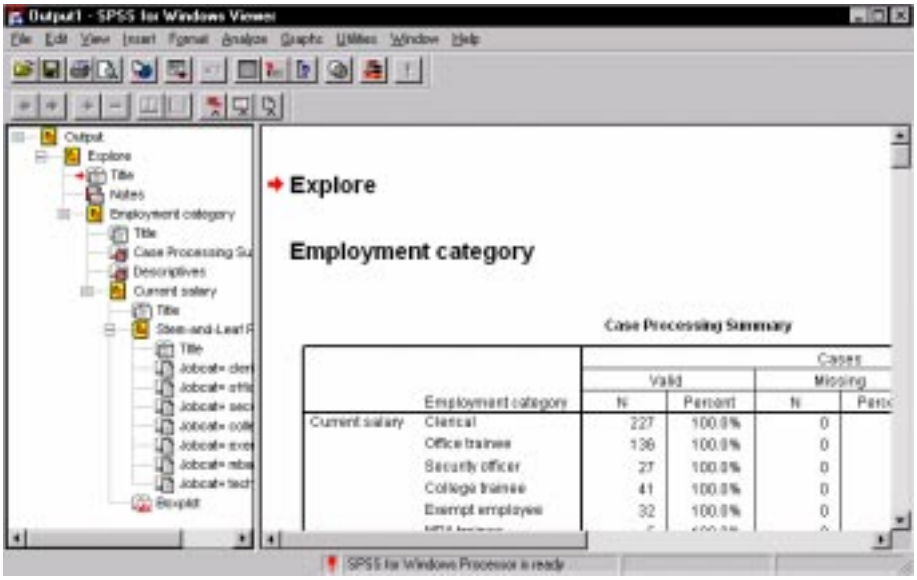
Task 4 Understanding Output

Objective To look at and understand the contents of the Output Window.

Instructions You will use some buttons from the Toolbar.

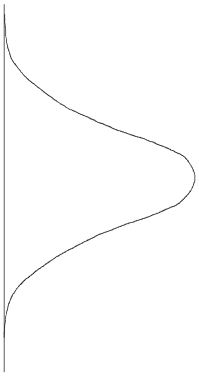
Comments The Output Window contains the results of your analyses. This is where you discover if it has all been worthwhile.

When your analysis has finished running your screen should look like this:



Activity 4.1 Now browse the Output Window using the scroll bar or clicking on the various items in the lefthand-side of the screen. You will see that SPSS has analysed the variable **`Current salary [salnow]`** separately for each grouping of the variable **`Employment category [jobcat]`**.

Activity 4.2 You will see that for each category of **`jobcat`** SPSS has printed out a variety of statistics –



various types of average, measures of dispersion plus measures to test the normality of the data, called *skewness* and *kurtosis*. In addition SPSS has printed out a *Stem-and-Leaf* plot for each group. This gives a rough and ready histogram of the data whilst retaining the individual data point values. If you turn your head sideways on and look at these plots you can get a quick idea of whether each group is normally distributed or not (i.e. is it symmetrical and looking like a bell?).

The statistics and stem & leaf plot give us a good idea of whether the data we want to analyse looks okay and meets the assumptions of the analysis we are going to use. The stem & leaf plot also gives us some indication of what values are outside the expected range – called *outliers*. With our attention drawn to these we may want to recheck whether the data has been entered correctly, or if there is a reasonable, but unusual, cause for this odd data, whether we want to include it in the main analysis or not.

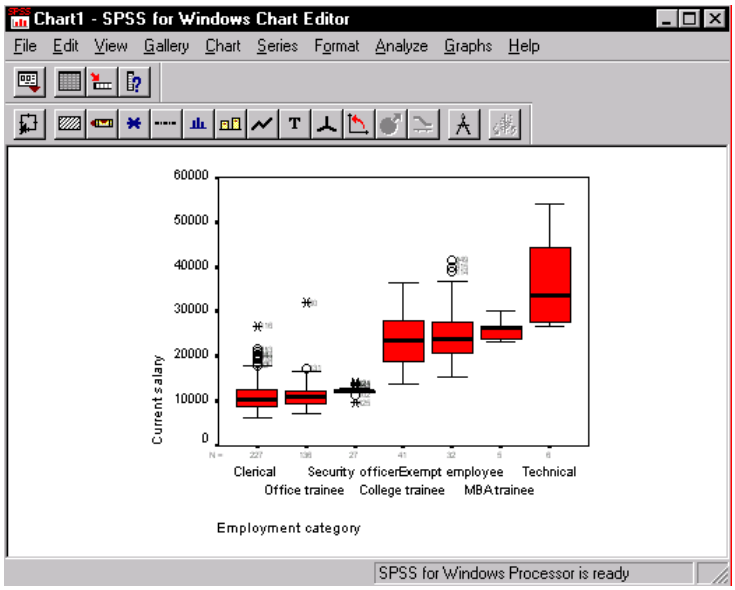
Task 5 Browsing Charts

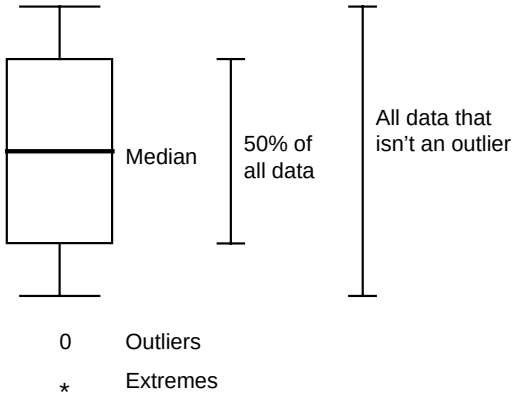
Objective To look at charts.

Instructions You will use some tool buttons from the Toolbar.

Comments It is much easier to check your data with a chart than with numbers.

Activity 5.1 Browse the Output window till you reach the boxplot produced from the Explore procedure done in Task 4. Double-click on the chart to open up the Chart Editor window. The window should look like this:





The chart is a boxplot chart of the variable **`Current Salary [salnow]`**, divided up into categories or groups by the variable **`Employment category [jobcat]`**. This shows us not only how the individual data is distributed but also how they compare against each other. If the median is very off-centre then the data is not normally distributed. You will also want to look at any outlier or extreme values.

Activity 5.3 You will see that there is a clerical worker who is earning nearly £30,000. This needs to be looked at. Click on the **Goto Case** button. A dialog box appears: type in the number **116** (printed on the boxplot) as the case to go to and click **<OK>**. SPSS will take you to that case in the Data Editor. Change the value for **`salnow`** to **`16750`**.

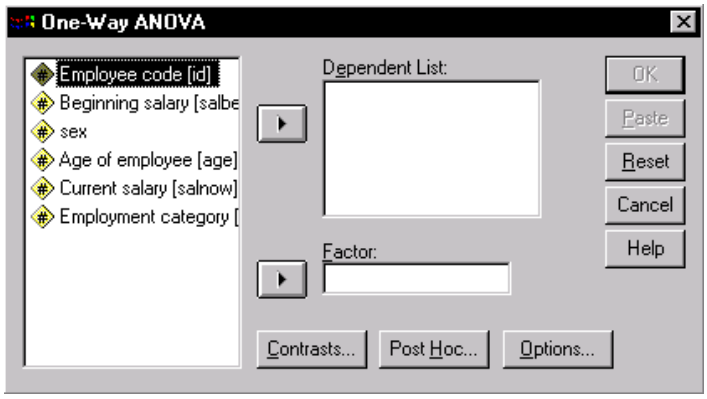
Task 6 Starting a One Way ANOVA

Objective To start a One-Way Analysis of Variance

Instructions You will use the **One-Way Anova..** command from the **Compare Means** submenu.

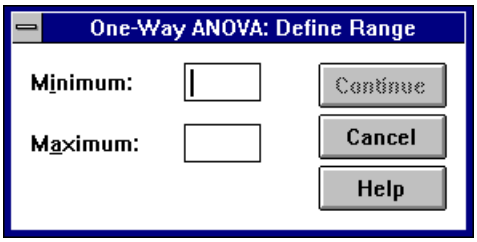
Comments One-way analysis of variance compares several groups of data and says whether they are from the same population or not.

Activity 6.1 From the **Statistics** (**Analyze** in version 9.0 of SPSS) menu click on the **Compare Means** command. From the sub-menu choose the **One-Way Anova...** command. The following dialog box appears:



Activity 6.2 Move the variable **salnow** to the **Dependent list** and the variable **jobcat** to the **Factor:** box.

Activity 6.3 Click on the button marked **<Define Range... >** underneath the Factor: box. The following dialog box appears:



In this dialog box we specify the values of the Factor variable that we want taken into account when the analysis is run. This dialog box makes it possible for us to investigate just a small range of values, and not necessarily all the values available.

Activity 6.4 Click in the **Minimum:** box and type the number **1**. Then click in the **Maximum:** box and type the number **7**. Now click on **<Continue>** to take you back to the One-Way ANOVA dialog box.

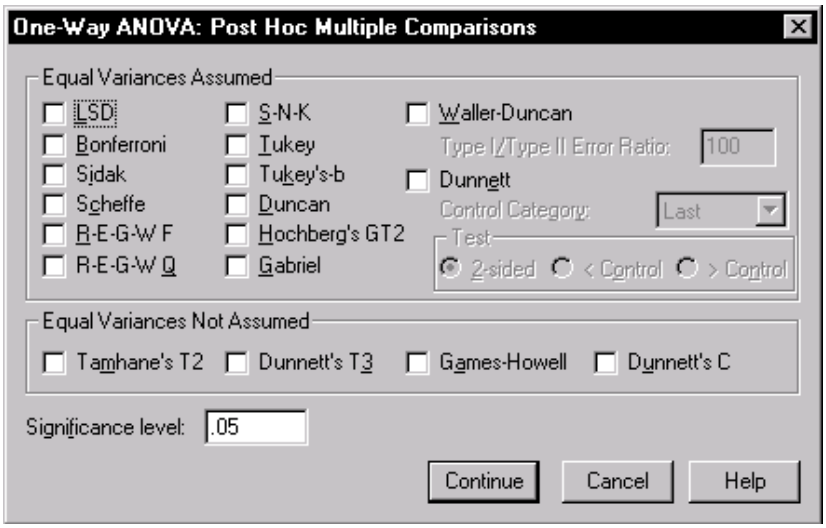
Task 7 Adding Analysis Options

Objective To finish defining your analysis and run it.

Instructions You will add some options from the **<Post-Hoc>** button and run the analysis.

Comment Every dialog box has a default set of options. However, these can be changed by means of sub-dialog boxes.

Activity 7.1 Click on the **<Post-Hoc...>** button. A dialog box appears listing all the optional tests you can perform on the data following an analysis of variance.



Activity 7.2 Click on the check box marked **Bonferroni**. This test does a comparison between all the groups being examined in the analysis and indicates where the differences highlighted by the analysis of variance might actually lie.

Activity 7.3 Click on **<Continue>** to take you back to the One-Way ANOVA dialog box.

You are now ready to run your One-Way Analysis of Variance on the variable `salnow`. The test we are performing is to see whether someone's job category determines what salary they get. Clearly we expect the answer to be "Yes".

Activity 7.4 In the One-Way ANOVA dialog box click **<OK>**. The analysis is run and you will be taken to the Output Window. Browse through the output using the mouse and scrollbars. Look firstly in the ANOVA table for something saying `Sig`. If the figure under this column is less that 0.05 then we can say that there is a difference at the 5% significance level. If the figure is less than 0.01 then there is a difference at the 1% significance level. Our figure of 0.0000 suggests quite a significant difference. Can you now class the 7 groups into 3 from the Bonferroni post-hoc analysis?

Task 8 Saving and Printing Output

Objective To print out the contents of the Output Window.

Instructions You will use the **Print** command from the **File** menu.

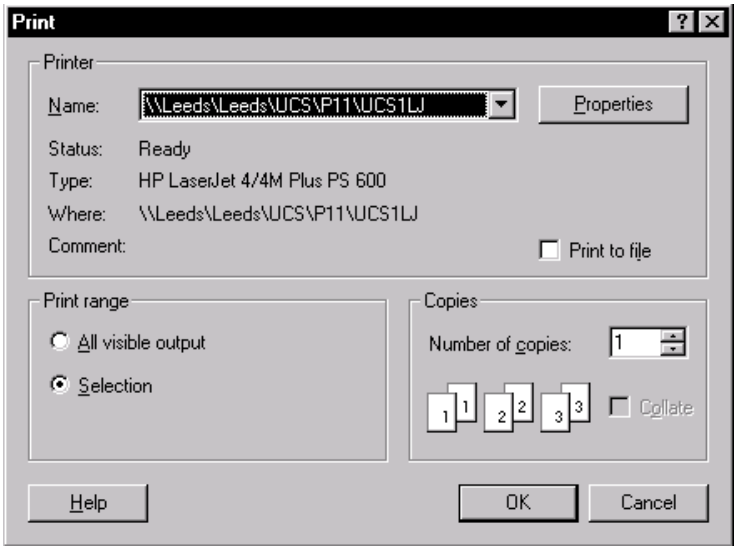
Comments The results of your analysis are all important, but you will probably not grasp the full implications of them immediately. It is always better to print the results off, grab a coffee and then review them at your leisure.

Activity 8.1 Before printing, you will need to ensure that SPSS is 'talking' the right printer language and attached to the correct printer queue. To do this you should consult the document *Getting Started with Microsoft Windows* (BEG 2).

The next activity is to select the output we want to print. We could print the entire contents of the Output Window but in this exercise we are just going to print a portion of it.

Activity 8.2 Click the **ANOVA Table** once with the left mouse button.

Activity 8.3 From the **File** menu select the **Print ...** command. Alternatively click on the **Print** button in the toolbar. You should get the dialog box as shown below:



The options are fairly self evident – you can either print the selection or all the output, and you can print multiple copies. Be aware that the font the Output Window will be printed in will be the same font you have asked it to be displayed. Clicking twice on a section of the Output allows formatting options to be changed.

Activity 8.4 Keep the options as they are displayed and press **<OK>**. The table selected will be printed out on whatever printer you are currently attached to.

Task 9 Finishing SPSS

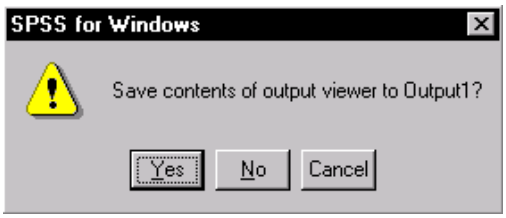
Objective To quit SPSS.

Instructions You will use the **Exit** command from the **File** menu.

Comments You should always quit any computer program when you have finished your session. Never switch off the computer when SPSS is still running unless absolutely necessary as this will corrupt your spreadsheet files. Also never leave a computer whilst you are still logged on to it as others may use your Novell account and could potentially damage your files.

Activity 9.1 Select the **Exit** option from the **File** menu.

Activity 9.2 SPSS will ask you if you want to save the contents of various windows before it lets you quit, so the following dialog box appears:



The numerical results of all your analysis work and also any error messages from SPSS if things don't go right are written in the Output window. You want to save this so click **<Yes>** and in the box that appears give your results a name. Once the datasheet has been saved it can be reopened by SPSS and edited.

Once you have done this SPSS will quit. If you click on the **<Cancel>** in any of the boxes you will be returned to your unsaved work to continue.