



DATA OPERATIONS IN SPSS FOR WINDOWS (8.0/9.0)

This document contains a series of exercises which give an introduction to manipulating 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
52**

50p

Contents

Contents	1
Introduction	1
About Data Management	1
Requirements	1
Documentation	1
Task 1 Getting Started	2
Task 2 Opening the Example File	3
Task 3 Defining Variables - Names	4
Task 4 Defining Variables - Labels	6
Task 5 Defining Variables - Formats	8
Task 6 Generating New Variables	9
Task 7 Recoding Data	10
Task 8 Data Display Options	12
Task 9 Printing the Datasheet	13
Task 10 Finishing SPSS	14

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



Copyright

This document is copyright University of Leeds. Permission to use material in this document should be obtained from the Information Officer (email should be sent to the address info-officer@leeds.ac.uk)

Print Record

This document was printed on 18-Nov-99.

Introduction

About Data Management

The end result of any statistical process is always the interpretation - what does the data mean? What do the graphs look like? What implications does this have? However, at the start there is the dealing with the raw data. It is the less glamorous aspect of data handling in SPSS that this document is intended to tackle.

Before a computer can do any work on your data, you have to tell it what your data is. This either means typing it in or reading it in from an existing source. You also need to tell the computer how the data is organised - maybe what units the data is in, what type the data is (numerical or text, categorical or value). It may be necessary to generate your own data from existing results - create a 'sort order' for them, or simply put a random sequence of numbers against them for selection purposes. It may be necessary to reduce the detail in the data, to break it down into groups in order to make the results easier to analyse.

It's not only the computer that needs additional information about the data in its data sheet. Humans are not very good at just looking at raw figures. They often want some explanatory text to say what the raw figures mean so that when the results are printed out they can be interpreted easily.

This document will take you through some of the basic tasks that you may need to accomplish when entering and formatting data in SPSS for Windows. The concepts learnt in this document should enable you to set up your own data files in SPSS. 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, 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:

Simple Statistical Analysis in SPSS for Windows (TUT 53)

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.

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.

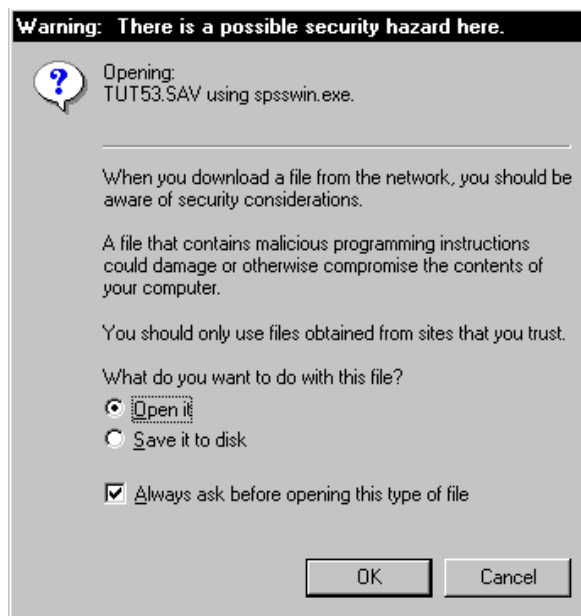
Instructions You will ftp the file from the ISS web site.

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 52 and click on the **tut53.sav** link, the dialog box below appears. 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.

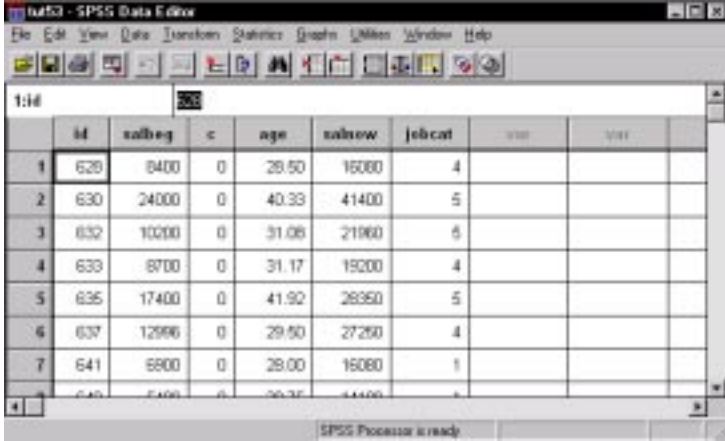
Comment Opening existing files is more common than opening new ones. Fortunately the process for opening files in SPSS is the same as it is for practically every other Windows program.

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 below.



	M	salbeg	c	age	salnew	jobcat	year	year
1	628	8400	0	28.50	16080	4		
2	630	24000	0	40.33	41400	5		
3	632	10000	0	31.08	21960	6		
4	633	8700	0	31.17	19200	4		
5	635	17400	0	41.92	28350	5		
6	637	12966	0	29.50	27250	4		
7	641	6900	0	28.00	16080	1		
8	645	5400	0	26.33	14400	1		

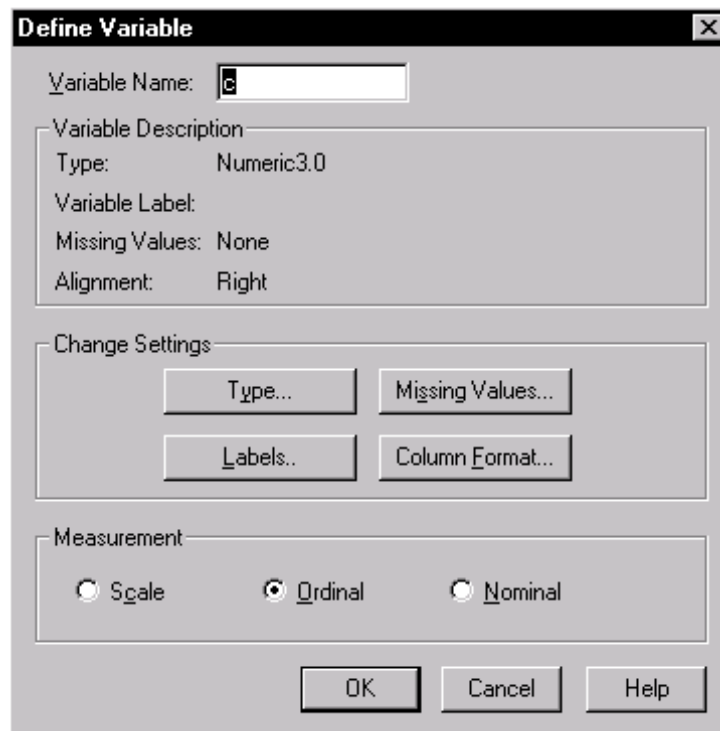
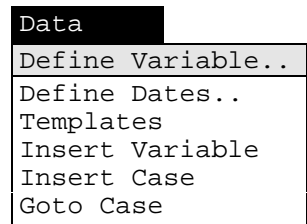
Task 3 Defining Variables - Names

Objective To label your variables and make their names more meaningful.

Instructions You will use the **Define Variable** command on the **Data** menu.

Comments The first step for you after (or during) entry of your data is to say what it means so that whenever it is displayed or reported the results are immediately obvious.

Activity 3.1 Move the mouse over the third column label, c, and double-click the left mouse button. Alternatively, click the mouse button once and then go to the **Data** menu and select the command **Define Variable..** . The following dialog box appears:



This dialog box controls all the definitions of the variable you are working with. At the top is the current variable name, highlighted.

Activity 3.2 Type the word: **sex**. This is now the variable name and is the name by which you will know the variable within the SPSS program. Later on we will add a fuller description of the variable which will get used in all results tables and charts.

Variable names are subject to some restrictions

- they can only be up to eight characters long
- they must start with a letter
- the rest of the variable name can consist of letters, numbers, full stops or the following symbols: @, #, _, or \$. No other characters are allowed (including spaces)
- the variable name must not end with a full stop or the underscore character
- variable names must be unique. Be warned that SPSS is case insensitive with variable names, so that the names `newvar`, `NEWVAR` and `NewVar` are all the same.

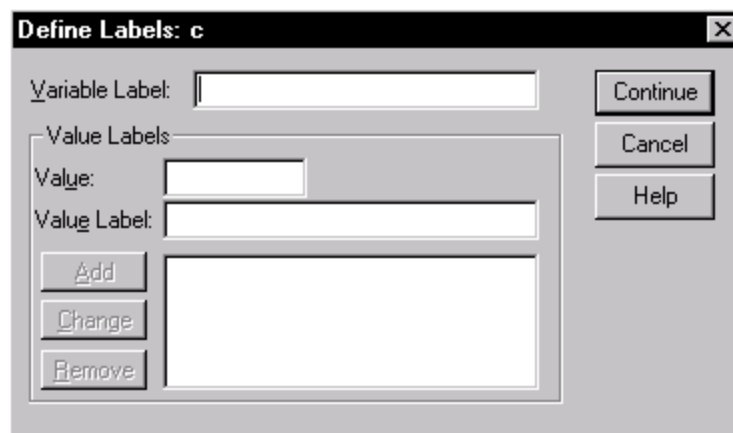
Task 4 Defining Variables - Labels

Objective To give the variables a more descriptive label and also to label the values.

Instructions You will use sub-dialog boxes from the Define Variables dialog box.

Comments The variable name `sex`, although more informative than the rather functional `c`, is still not great. Does it mean “gender” or does it mean something else? Some extra information is required.

Activity 4.1 Now click on the button labelled `<Labels..>`. Another dialog box appears:



In this dialog box you can add a fuller description of the variable and add labels for the different numbers in your data.

Activity 4.2 Type the label `: Sex of employee` in the **Variable Label:** box.

This particular variable has three potential values: 0, 1 and 9. This in itself is not very informative news and will not help anyone reading our results to interpret them correctly. Therefore we need to label the values of our variable `Sex` to say what value corresponds to what sex.

Activity 4.3 Click on the **Value:** box in the **Value Labels** section and type 0.

Activity 4.4 Now click on the **Value Label:** box below and type `Male`.

Notice that the moment you start typing, the `<Add>` button changes from being greyed out to becoming active.

Activity 4.5 When you have finished typing, click on the `<Add>` button. The box next to the `<Add>` button should contain the text:

```
0 = "Male"
```

Activity 4.6 Now repeat the exercise to label the value 1 as **Female**.

This now leaves the number 9. What does this mean? Usually this sort of number means that the information is missing. Often this is because the respondent has refused to fill the answer in, or else it could be that the respondent isn't sure of the answer. This latter reason is unlikely in our case, so we'll just define the value as "missing data".

Activity 4.7 Label the value 9 as **Missing**.

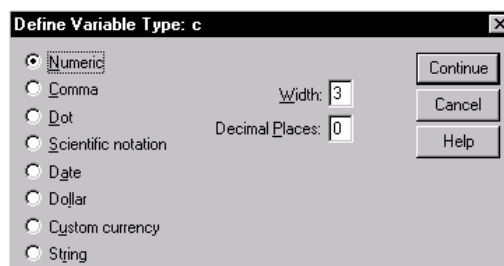
Task 5 Defining Variables - Formats

Objective To specify the format of a variable for display purposes.

Instructions You will use the <Type> button in the Define Variable dialog box.

Activity 5.1 From the Define Labels box click the <Continue> button. Notice that under the Variable description section SPSS has recognised the new variable label "Sex of employee".

Activity 5.2 Click on the <Type..> button in the box. The following dialog box appears:



This dialog box defines how the data in your variable is displayed both in the Data Editor and also in your tables of results. It does not change the way the underlying data is stored.

Activity 5.3 Click down each radio button in turn to see the options that are associated with each one.

- The top four options all have the same two parameters - **Width**, as in total width of the column, and **Decimal Places**. The width setting must take into account not only the decimal places but also the decimal point. For example, if you want 2 decimal places and you expect numbers to go up to 1000 you need to define the width as $4+2+1 = 7$.
- **Date** doesn't give you a width option but does give quite an extensive list of options for date formats. All the common ones are available.
- **Dollar** and **Custom Currency** both give you the **Width** and **Decimal Places** options together with special lists of formats for currency work.
- Clicking the **String** option just gives you the one parameter to change - **Characters**, which to all intents and purposes is the same as width.

Activity 5.4 Select **Numeric** and alter the format of the column so that it is 1 character wide with 0 decimal places.

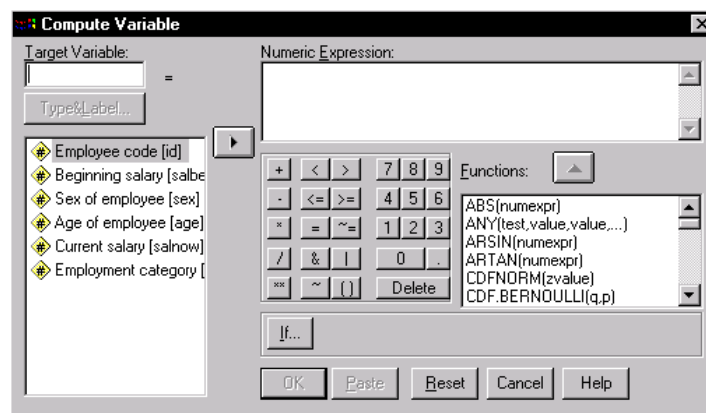
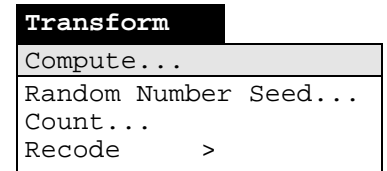
Activity 5.5 Now click <Continue> to get back to the Define Variables dialog box, and from there click <OK> to get back to the Data Editor. The variable name **sex** should now appear at the top of the third column.

Task 6 Generating New Variables

Objective To generate new variables either from scratch or on the basis of existing ones.

Instructions You will use the **Compute** command from the **Transform** menu.


Activity 6.1 From the **Transform** menu choose the **Compute** command. The following dialog box appears.



This box allows you to create a new variable either by generating one from fresh (for instance, by generating a sequence of random numbers) or as a result of a calculation involving one or more variables.

Activity 6.2 Click in the box labelled **Target Variable:** and type the variable name: **salrise**. We are going to create a variable that is the difference between people's salaries now and when they started.

Activity 6.3 Click on the variable **Current salary [salnow]** in the **Source Variables** list on the left hand side. Click on the pushbutton to move it across to the **Numeric Expression:** box.

Activity 6.4 Now click on the minus button  in the calculator pad underneath the **Numeric Expression:** box. A minus sign appears.

Activity 6.5 Click on the variable **Beginning salary [salbeg]** in the **Source Variables** list on the left hand side. Click on the pushbutton to move it across to the **Numeric Expression:** box.

We have now set up our numeric expression to create the variable, **salrise=salnow - salbeg**.

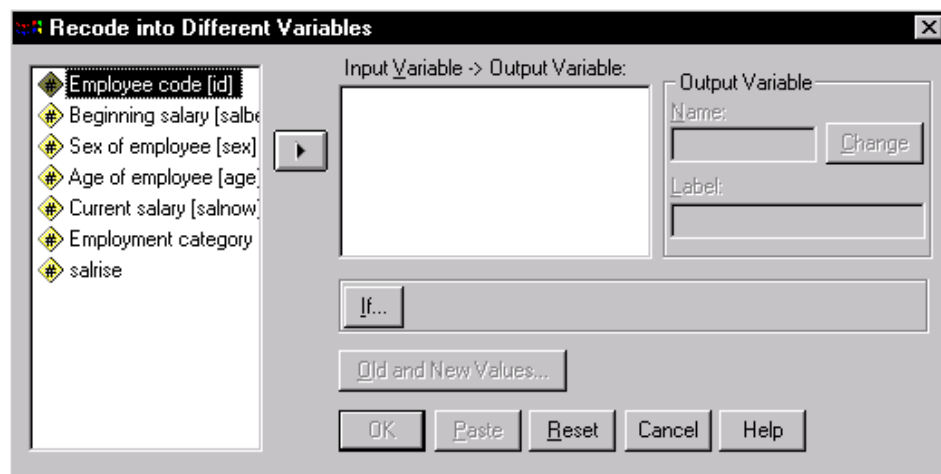
Activity 6.6 Click on the **<OK>** button. The status bar briefly shows the **Running COMPUTE** and **Running EXECUTE** messages, and you are then returned to the Data Editor. **Salrise** will be a new variable in the next available column.

Task 7 Recoding Data

Objective To recode a block of data in order to group some categories together.

Instructions You will use the **Recode** command from the **Transform** menu.

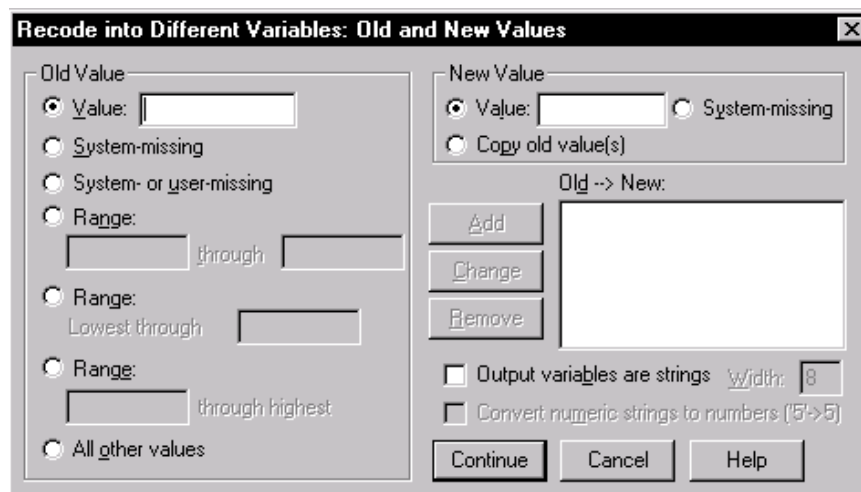
Activity 7.1 From the **Transform** menu select the command **Recode**. From the submenu select **Into Different Variables**. The following dialog box appears:



Activity 7.2 Click on the variable **Age of employee [age]** in the **Source Variables** list on the left hand side. Now click on the pushbutton to select the **age** variable and put it into the central box.

Activity 7.3 Click in the box labelled **Name:** under the **Output Variable** section on the right. Type the name **agegrp** and click on the **<Change>** button. The word **agegrp** should now appear on the other side of the arrow in the central box.

Activity 7.4 Now click on the button **<Old and New Values...>** towards the bottom of the dialog box. The following screen appears:



This box allows us to say what values are in the old variable (*age*) and how we want them to relate to values in the new recoded variable (*agegrp*).

Age values have a range of 23.00 to 64.50. We are going to classify them in groups, or *bins*. SPSS works much better if you assign things fairly arbitrary numeric codes which you then label meaningfully, so the following is a table of the groups with the code we shall attach to them.

20-30	2
30-40	3
40-50	4
50-60	5
60-70	6

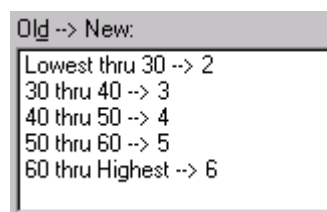
Activity 7.5 In the Old and New Values dialog box click on the radio button labelled **Range: Lowest through**, the fifth option down. In the text box type: 30. This means that our first group will include any values below the value of 30.

Activity 7.6 In the **New Value** section at the top right corner of the dialog box click in the box labelled **Value:** and type: 2. The **<Add>** button directly below now becomes highlighted, so click on it to add the first value recode to the list for this variable. The **Old -> New** box should now have one entry in it reading **Lowest thru 30 -> 2**.

Activity 7.7 Next, click on the first **Range:** option which gives you two boxes and the word **Through** in between them. This is now the first full range to recode.

Type 30 in the first box and 40 in the second. How do we know which group the value 30 itself will fall into – group 2 or group 3? SPSS puts inclusive values into the first recode it comes across – in this case into value 2. Now click on the **Value:** box as before and enter the number 3. Click on **<Add>** to add it to the **Old -> New** list.

Activity 7.8 Complete the recoding according to the values shown in the table above. For the last group, 60-70, you may as well define it as 60 thru highest. When you have finished, your **Old -> New** table should look like this:



Lowest thru 30 --> 2
30 thru 40 --> 3
40 thru 50 --> 4
50 thru 60 --> 5
60 thru Highest --> 6

Activity 7.9 Click on the **<Continue>** button. Then in the Recode into Different Variables dialog box click on the **<OK>** button.

The dialog box disappears and you are taken back to the Data Editor. After a short pause a new variable, *agegrp*, appears in the first spare column of the Data Editor with age group values. You may now like to define this variable properly and add descriptive labels for the values.

Task 8 Data Display Options

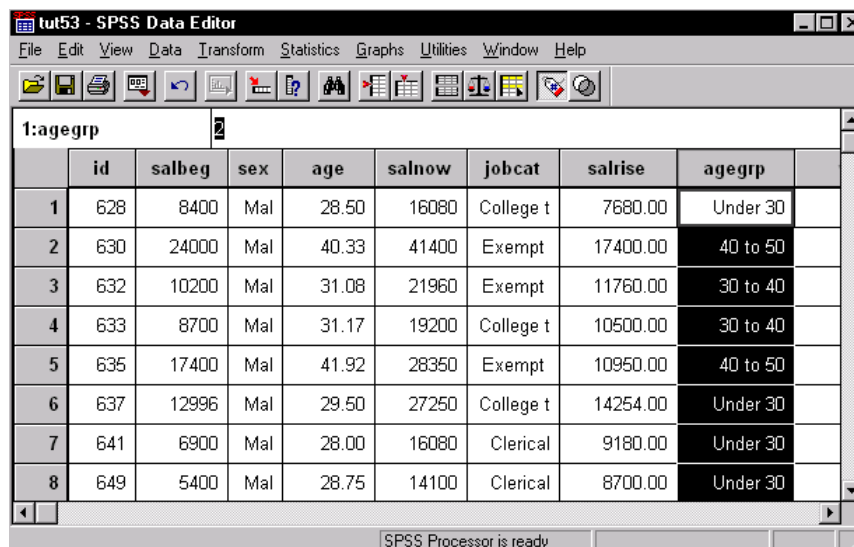
Objective To enable you to alter the way SPSS for Windows displays the information in the Data Editor.

Instructions You will use some commands from the **View** menu.

Comments SPSS allows us to look at our raw data in its labelled form to help us analyse individual cases.


The sheet as it stands is just showing us the number values. We will now get it to show labels.

Activity 8.1 From the **View** menu select the **Value Labels** command. The contents of the Data Editor change to show the labels for the individual values rather than the values themselves. Your Data Editor window should look like that shown below:



	id	salbeg	sex	age	salnow	jobcat	salrise	agegrp
1	628	8400	Mal	28.50	16080	College t	7680.00	Under 30
2	630	24000	Mal	40.33	41400	Exempt	17400.00	40 to 50
3	632	10200	Mal	31.08	21960	Exempt	11760.00	30 to 40
4	633	8700	Mal	31.17	19200	College t	10500.00	30 to 40
5	635	17400	Mal	41.92	28350	Exempt	10950.00	40 to 50
6	637	12996	Mal	29.50	27250	College t	14254.00	Under 30
7	641	6900	Mal	28.00	16080	Clerical	9180.00	Under 30
8	649	5400	Mal	28.75	14100	Clerical	8700.00	Under 30

Not all the labels actually fit into the columns at their present widths, so we might like to change the column widths to accommodate them.

Activity 8.2 Move the mouse cursor slowly over the variable names at the top of the columns from left to right. Notice that as the cursor crosses the boundary from one variable to another it changes shape briefly from a crosshair to a double-headed white arrow. 

Move the cursor between the two variables **jobcat** and **salrise** so that it becomes a double-headed white arrow. Click and hold the mouse button and move the mouse to the right until the arrow is over the border of **salrise** and **agegrp**. Let go the mouse button, and the column should resize itself.

Activity 8.3 Go to the **View** menu again and choose the command **Fonts**. A typical Windows fonts dialog box appears. Select a font of your choosing and click **<OK>**. The entire Data Editor will be displayed in this font.

Task 9 Printing the Datasheet

Objective To print out the contents of the Data Editor.

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

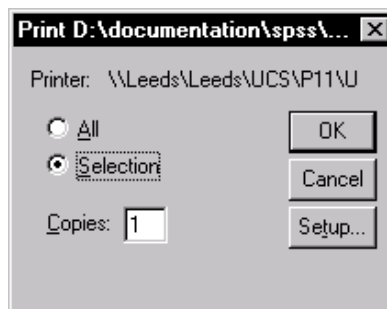
Comments Having worked so hard on your data you will want to take away a printed copy. It is also much easier to see errors on printed sheets than on screen.

Activity 9.1 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 area of the sheet we want to print. We could print the whole sheet (and normally you probably would) but in this exercise we are just going to print a portion of it.

Activity 9.2 Click the mouse button in the top left hand cell of the **Data Editor (1:id)**. Hold the mouse button down and drag the mouse down until you reach the 20th case, or row. Keep the button held down as you drag the mouse across to include all the variables in the Data Editor. You should now have selected everything from 1:id to 20:agegrp. Let go the mouse button.

Activity 9.3 From the **File** menu select the **Print ...** command. Alternatively, click on the **Print** button in the toolbar. You should get the dialog box shown below:



The options are fairly self-evident – you can either print the selection or the entire worksheet, and you can print multiple copies. Be aware that the font the datasheet will be printed in is the same font it is displayed in.

Activity 9.4 Keep the options as they are displayed and press **<OK>**. Your datasheet will be printed out on whatever printer you are currently attached to.

Task 10 Finishing SPSS

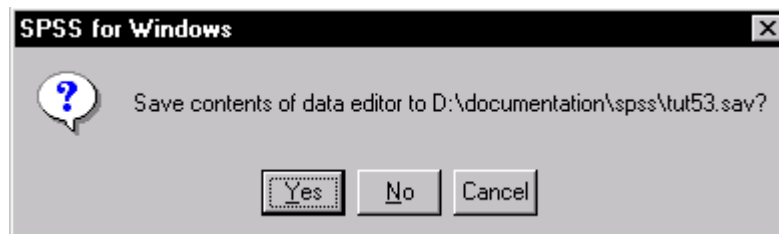
Objective To quit SPSS.

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 10.1 Select the **Exit** option from the **File** menu.

Activity 10.2 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:



Click either **<Yes>** or **<No>** depending on whether you want the data saved. If you choose to save it the file is saved to the same name on disk. Once the datasheet has been saved it can be reopened by SPSS and edited.

You will also be asked if you want to save the contents of your Output Window – usually the numerical results of all your analysis work. Normally you would want to save this, but here you should click **<No>**.

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