TRAINING PROGRAMME ON UTILISATION OF SPSS SOFTWARE PACKAGE FOR QUANTATIVE EDUCATIONAL DATA FOR THE FACULTY OF REGIONAL INSTITUTE OF EDUCATION (NCERT), MYSORE

Report

Dates:

30th November to 4th December 2004

Team:

Dr. G. Viswanathappa Coordinator

Dr. D. Basavayya, Co-coordinator

Regional Institute of Education (NCERT)

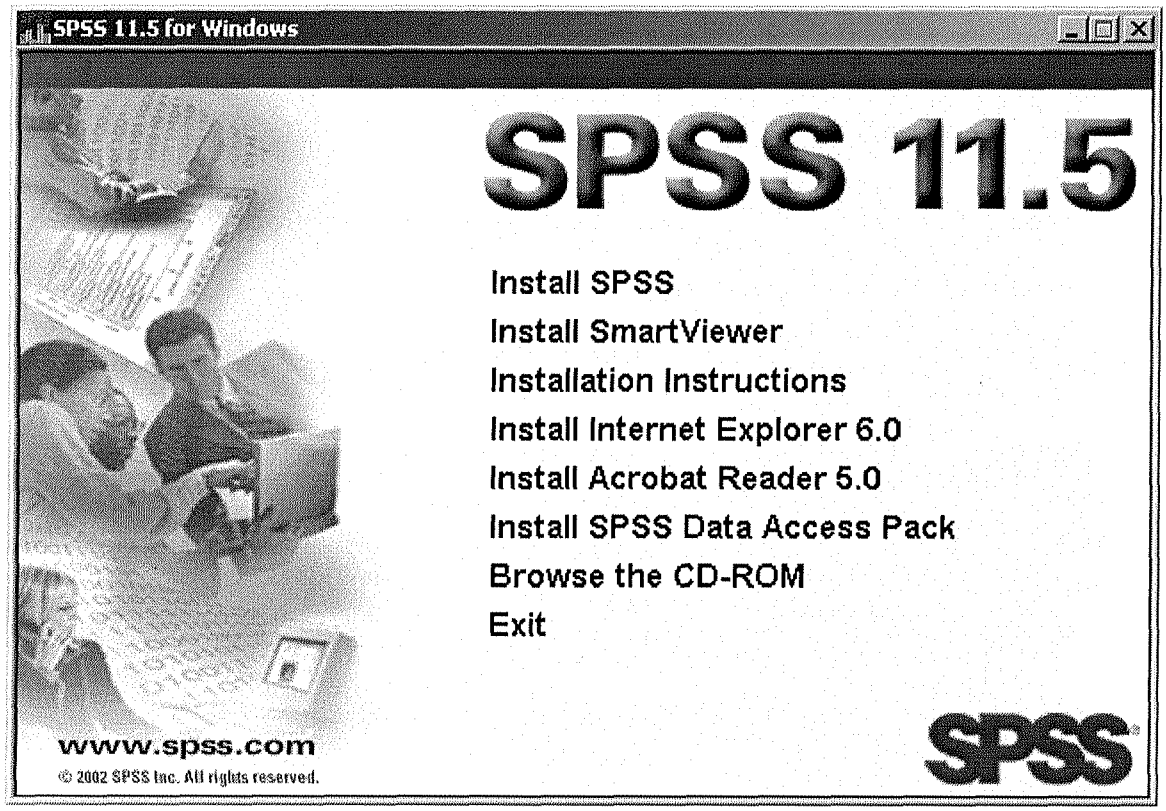
MYSORE - 570 006

SPSS PC Tour contents

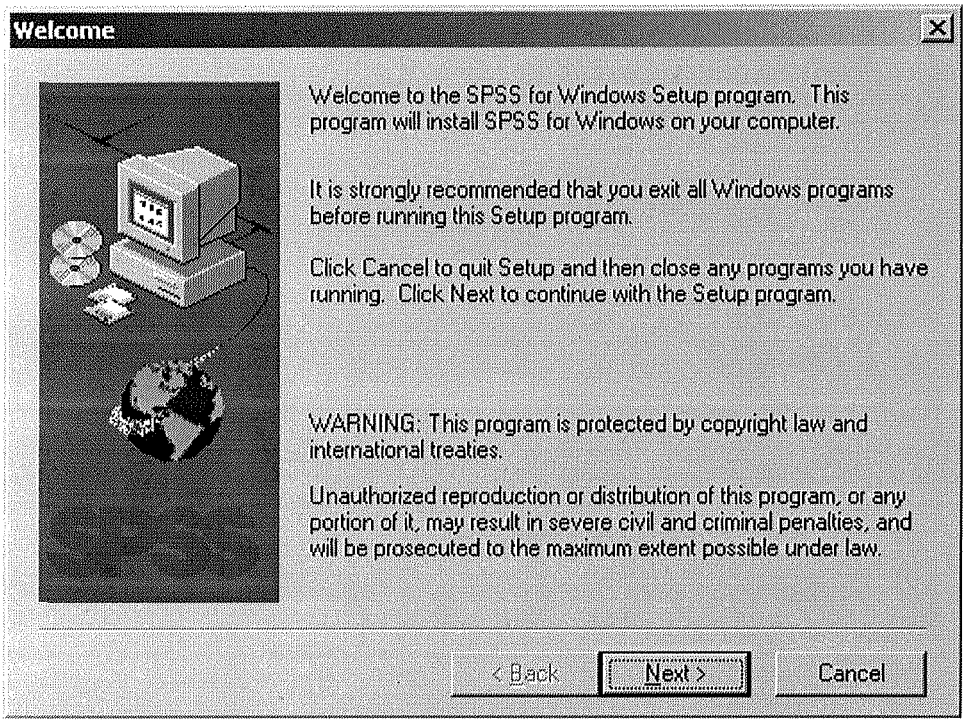
SPSS 11.5 for Windows Installation

SPSS 11.5 for Windows 98/ME/XP/2000/NT
Installation instructions

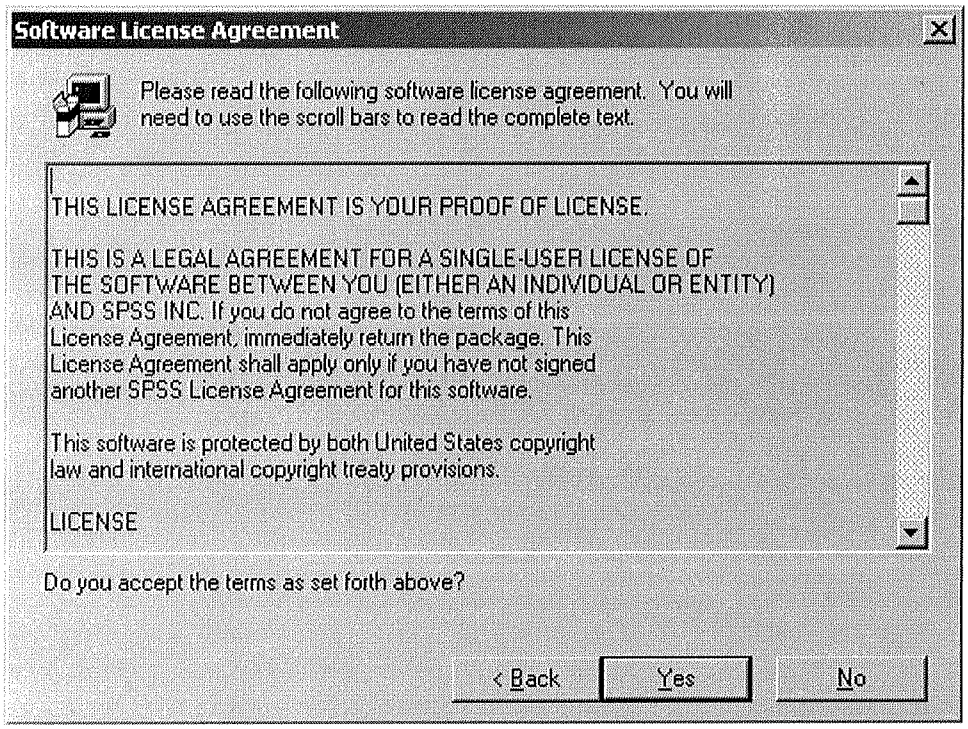
1. Make sure there is 84 megabytes of available space on the hard drive before installation of this program.
2. Make sure you have the "**serial number**" and "**license code**" for SPSS. This information would have been emailed to you. Keep a printout of this email handy so you can type in the "**serial number**" and "**license code**" when prompted during the installation.
3. Insert the "**SPSS 11.5 for Windows**" CD into the disk drive. After several seconds the install splash screen will appear. If the install wizard fails to activate then press "**Start | Run**" menu items. Navigate the Run dialog box to the CD-ROM containing the software and select the "**setup.exe**" executable file. Press the "**Open**" button to select the program and return to the run dialogue box then press the "**Ok**" button to activate the setup program.



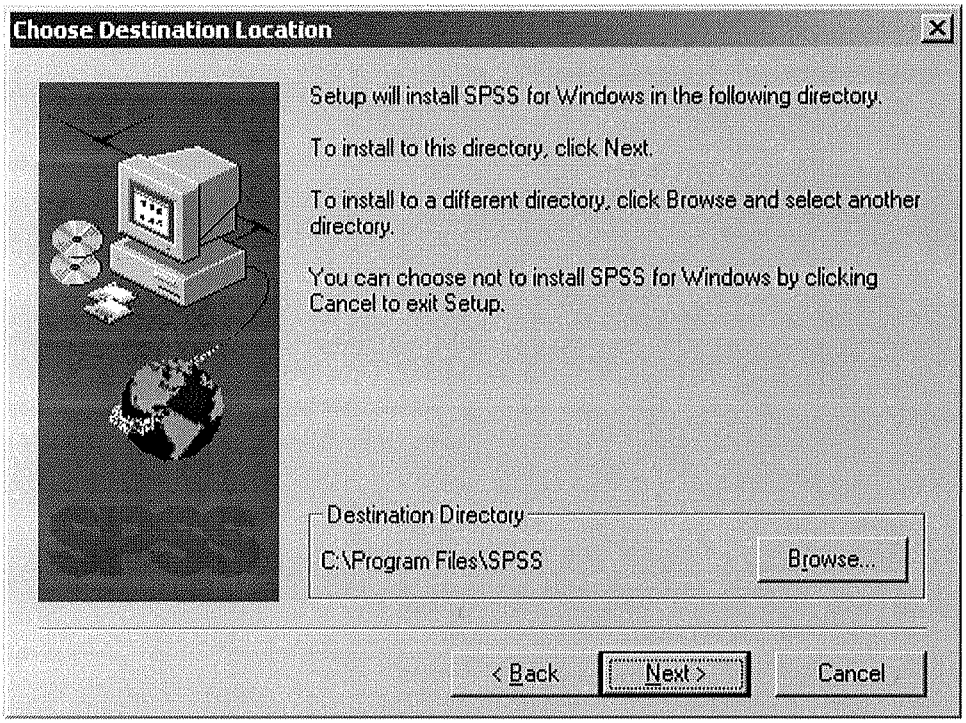
Click on '**Install SPSS**'.



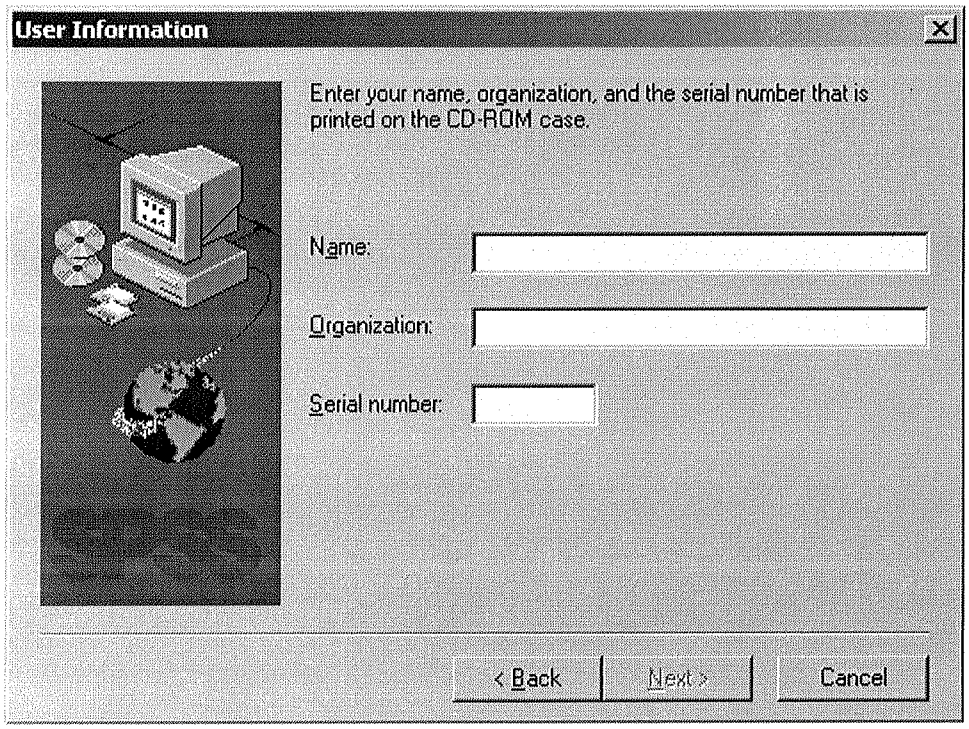
Click on '**Next**'.



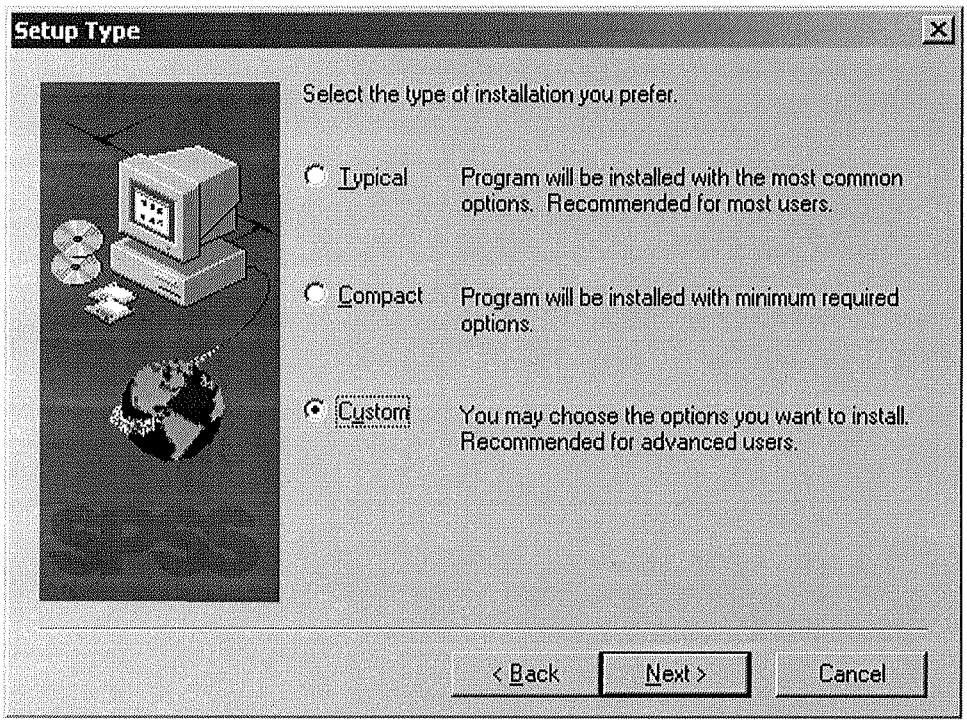
6. Click on '**Yes**'.



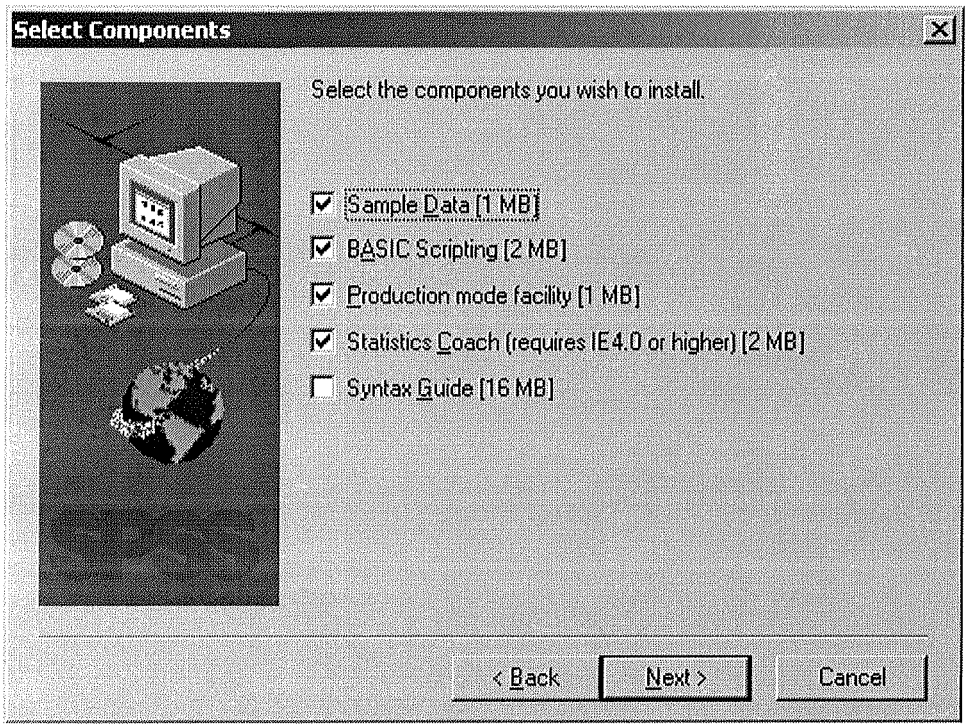
7. Click on '**Next**'.



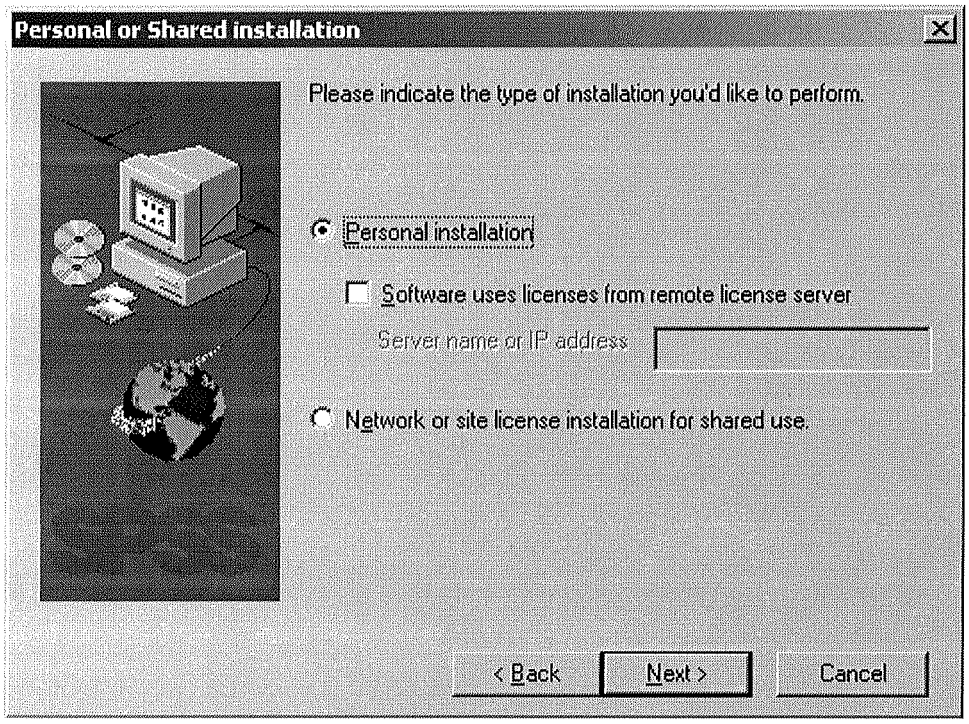
Under '**Name**', enter **your name**, under '**Organization**', enter '**University of Cincinnati**', and under '**Serial number**', enter the serial number supplied to you in the e-mail. Once this is done, click '**Next**'.



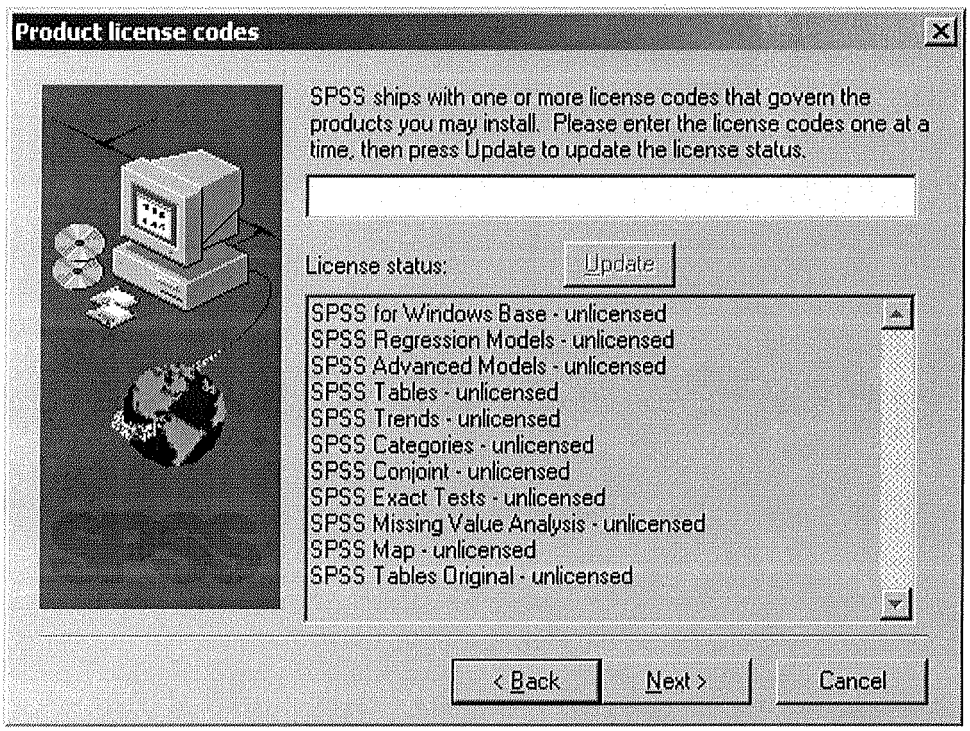
Choose '**Custom**' and click on '**Next**'.



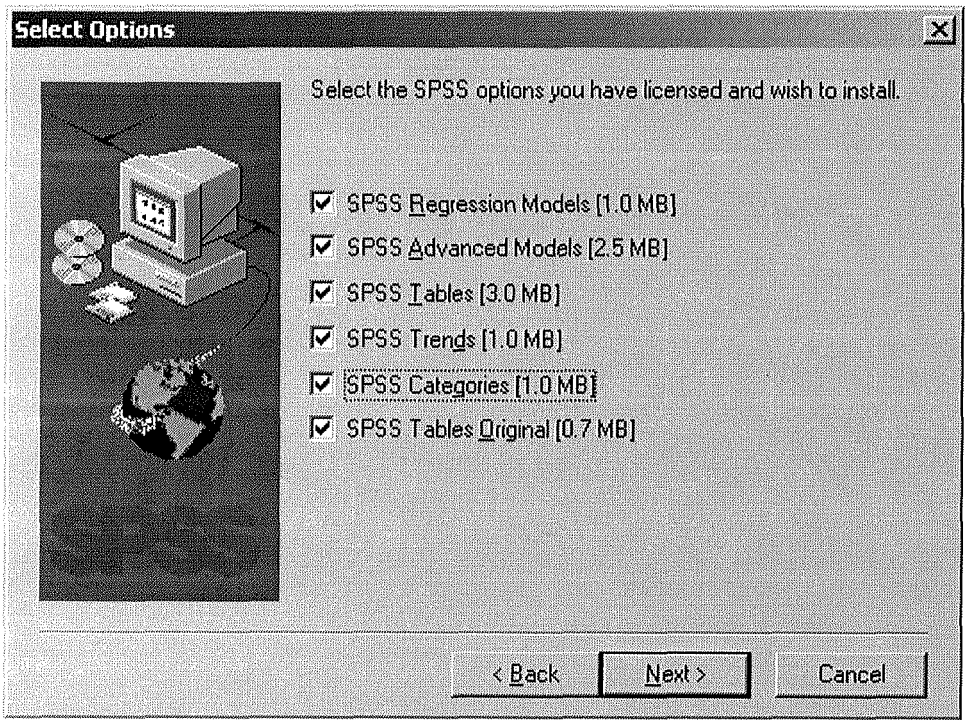
Check the '**Syntax Guide** (16 MB)' check box and click on '**Next**'.



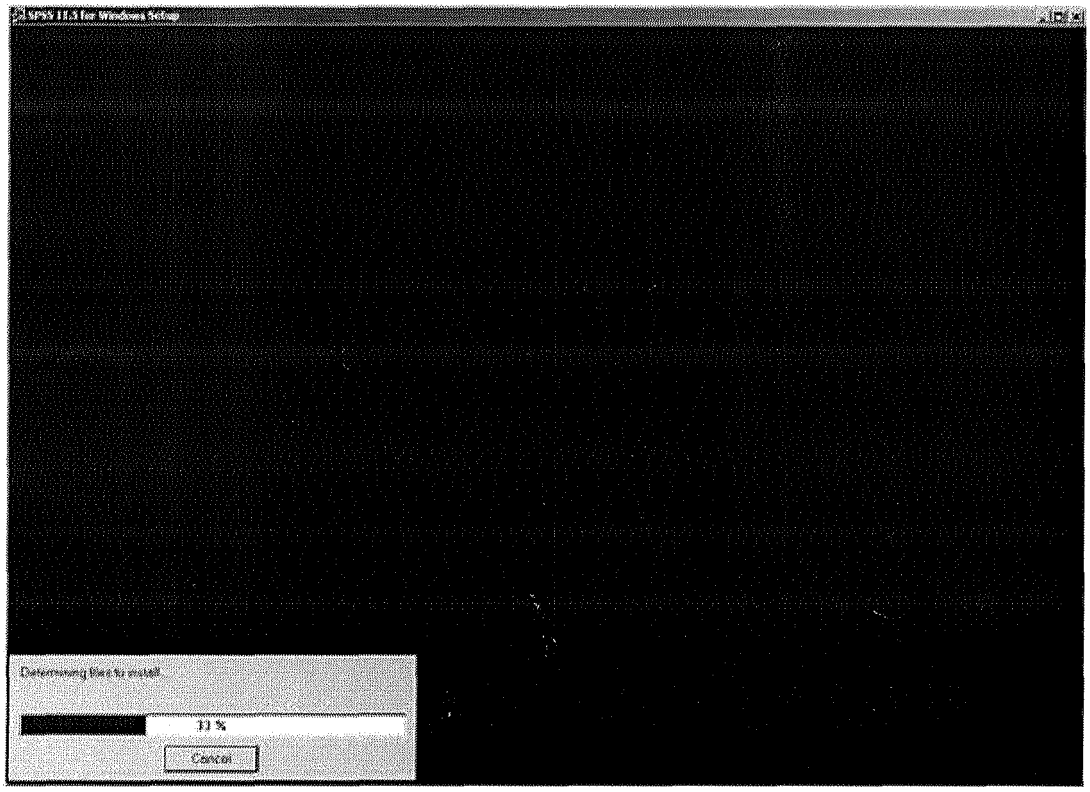
Click on '**Next**'.



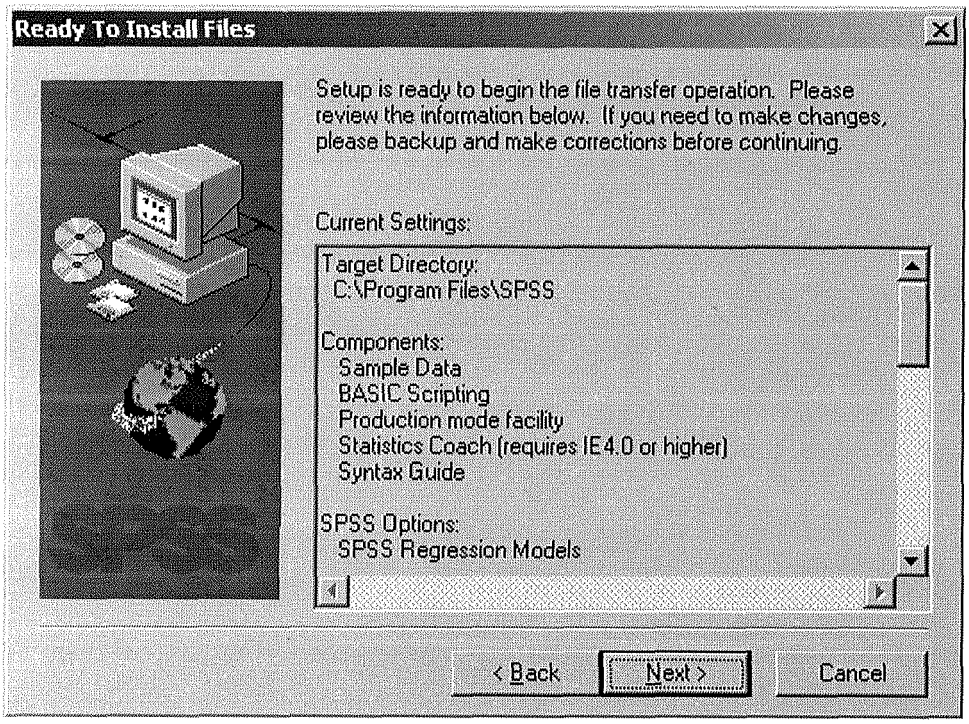
Enter the license number which has been e-mailed to you in the box above. Be sure to enter the license number without any dashes. Then click '**Next**'.



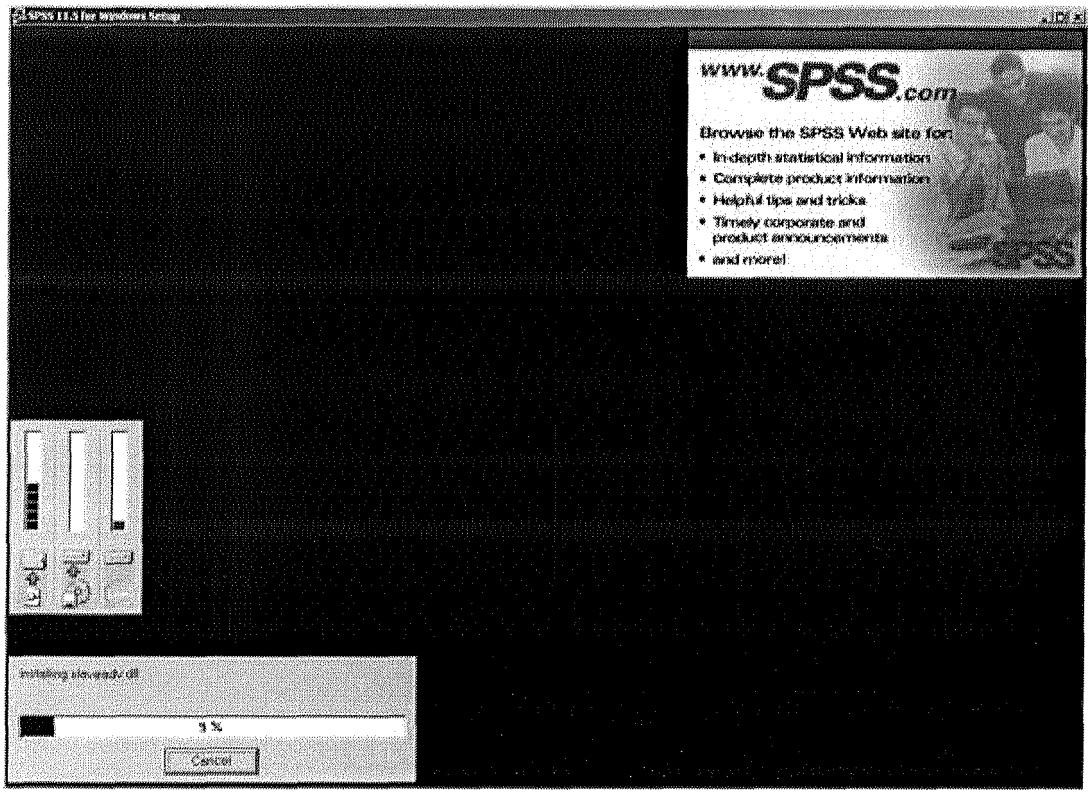
Click '**Next**'.



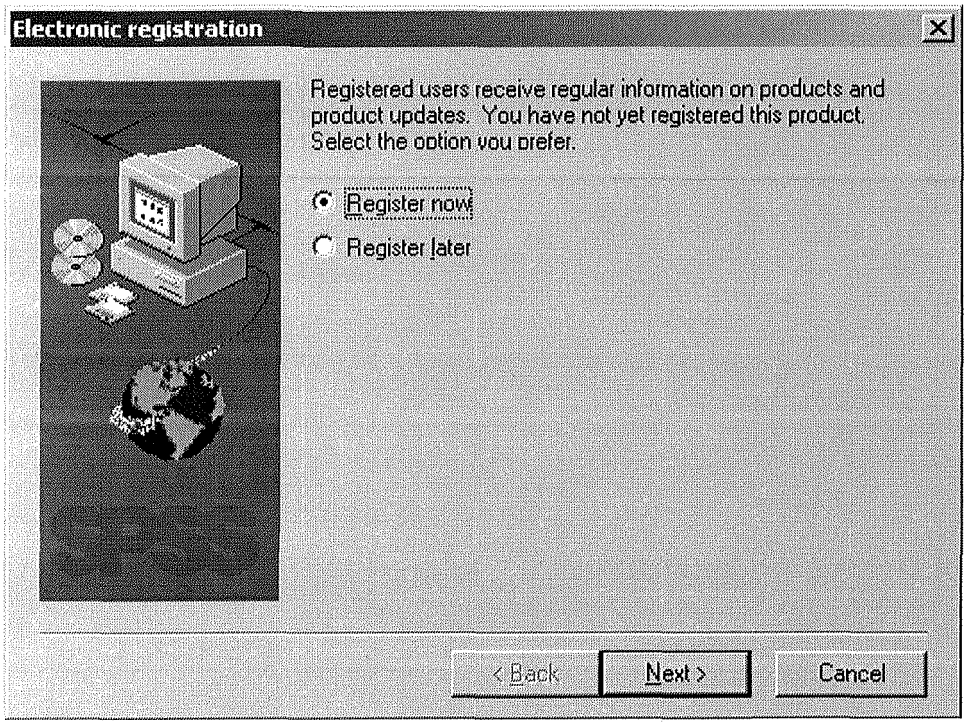
At this point the installation begins and you will see a screen like the screen above. Wait till the screen changes to the one shown below.


Ready To Install Files
Setup is ready to begin the file transfer operation. Please review the information below. If you need to make changes, please backup and make corrections before continuing.
Current Settings:
Target Directory:
C:\Program Files\SPSS
Components:
Sample Data
BASIC Scripting
Production mode facility
Statistics Coach (requires IE4.0 or higher)
Syntax Guide
SPSS Options:
SPSS Regression Models
< Back
Next >
Cancel

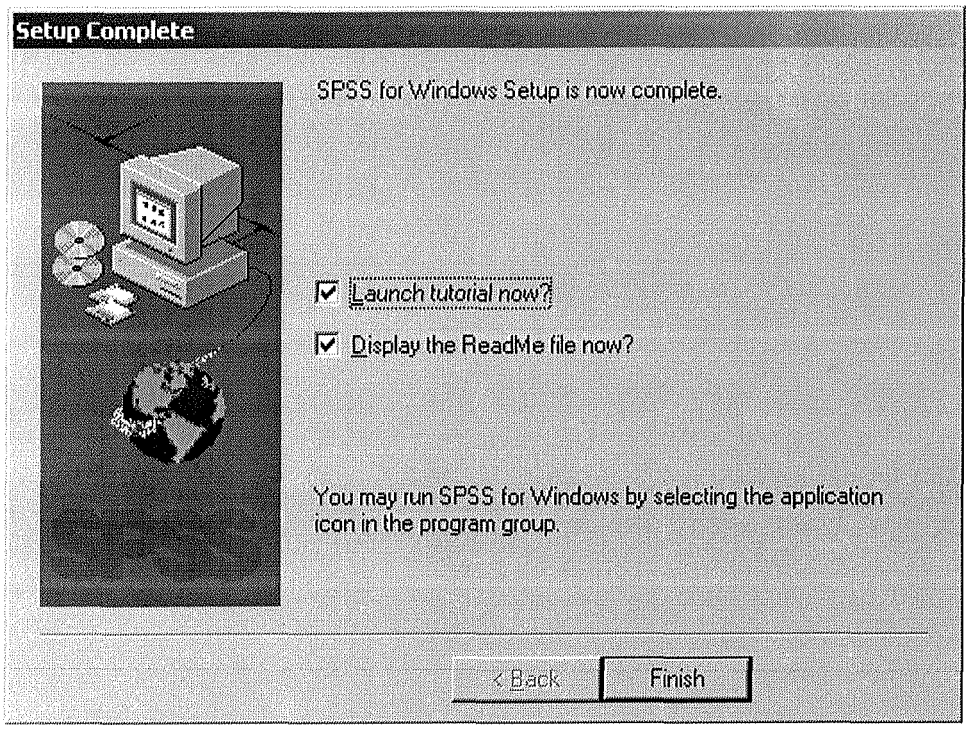
Click '**Next**'.


www.SPSS.com
Browse the SPSS Web site for:
• In-depth statistical information
• Complete product information
• Helpful tips and tricks
• Timely corporate and product announcements
• and more!
Cancel

You will see an install screen like the one shown above. Once the installation is complete, you get a screen like the screen below.



Select the option '**Register later**' in the screen above and click '**Next**'.



Uncheck the '**Launch tutorial now?**' and the '**Display the ReadMe file now**" checkboxes and click on '**Finish**'.

SPSS 11.5 for Windows
Uninstall instructions:

1. Select the following menu items to arrive at the dialogue box where you can uninstall the software package "**Start | Settings | Control Panel | Add/Remove Programs**".
2. Scroll down the listing and select the entry for "**SPSS 11.5 for Windows**"
    a. Click the Change/Remove button to active the "Remove programs from your Computer" wizard. The following components will be removed ...
        i. Shared program files
        ii. Standard program file
        iii. Folder items
        iv. Program folders
        v. Program directories
        vi. Program registry entries
3. A progress bar indicates the progression of the removal task.
4. You will receive uninstall completed message. Some elements could not be removed. You should remove items related to the application.
5. Click on the "**Ok**" button. The entry in the "**Add/Remove Programs**" listing will disappear and disk space on the hard drive should increase by 82 megabytes.
6. Restart the computer.

Starting off

This is brief tour to whisk you round the main features of SPSS 10 to 11.5. Like any tour nothing is covered in great detail, the aim is to let you see how SPSS works.

ⓘ *The information symbol will be used to flag interesting areas and where to find out more. Go tothe* `Information` *section at the end of the tour to find out about:-*

- ⓘ **Things to do...** including SPSS courses.
- ⓘ **Places to visit...** such as useful web sites and where to find SPSS.
- ⓘ **People to see...** about getting help with SPSS and computers.

All you will need to follow this tour is a computer with SPSSfor Windows installed, any version from 10 to 11.5. If you have an Apple computer with SPSS 10 for Macintosh installed, there is separate tour for that.

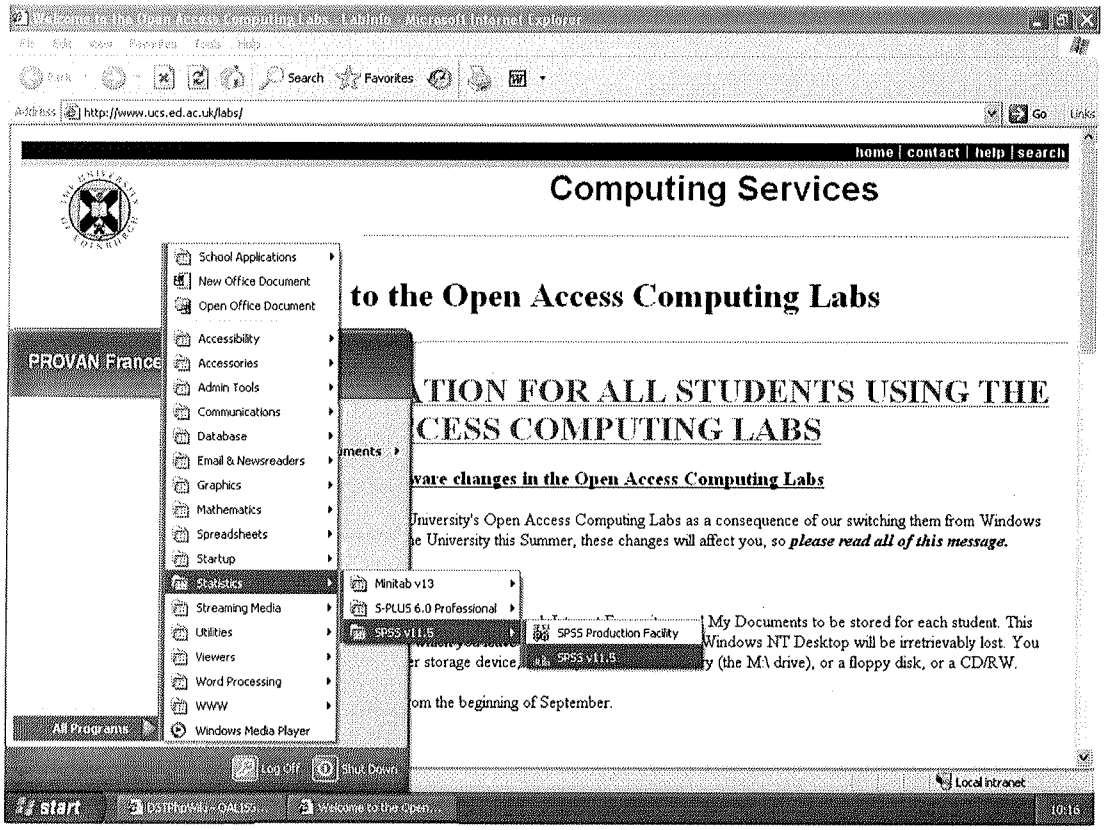
So grab your hat and a computer and let's visit the wonderful world of SPSS!

The mouse symbol will be used when you have to do something on the computer to follow the tour. In this case, switch on/log into the computer you are going to use.

Starting up SPSS

SPSS can be opened from **Start** menu. The button for the **Start** menu is usually at the bottom left hand corner of the desktop. Pic 1 is a picture of a PC desktop running Windows XP in the public labs. You can see from the **Start** menu, **All Programs** has been selected, from there **Statistics** has been chosen, and then from **Statistics**, **SPSS 11.5** has been chosen and then **SPSS 11.5** has been chosen from its menu.

Pic 1: Choosing SPSS from the SPSS directory



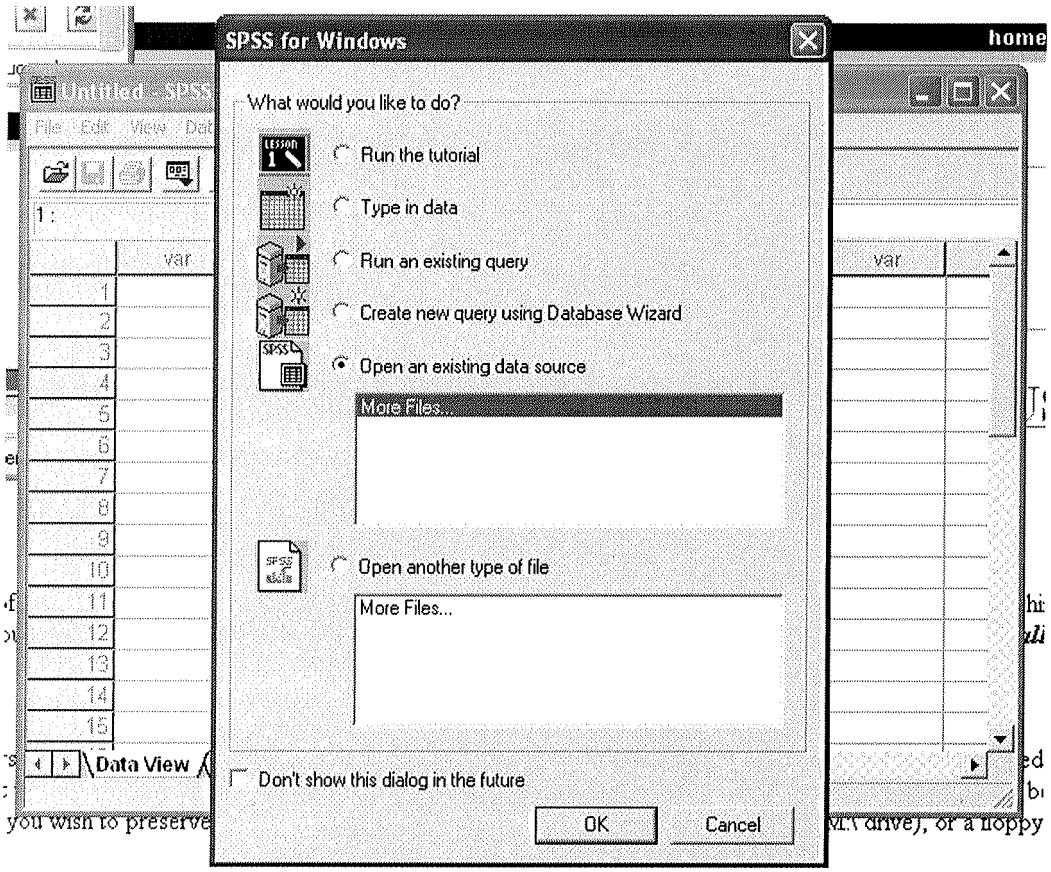
In Public labs, click the Start button and choose All Programs, from there choose Statistics, then choose SPSS 11.5 and SPSS 11.5 again to open SPSS.

*The exact sequence might vary depending on how the computer has been setup but SPSS should be available through the **All Programs** or **Programs** menu if it is there. If you can't find it, search for the SPSS directory using the **Start** menu choose **Search** or **Find**, then select **All files or folders** or **Files or Folders** and search the computer for the word "spss".*

SPSS for Windows dialog

Starting up SPSS will open the **SPSS for Windows** dialog box shown in **Pic 2.**

Pic 2: First view of SPSS



SPSS should look like this in the Public labs and Computing Services training suite and when it is first installed (unless SPSS has been customised on the computer you are using). There are several choices of what to do next: -

Run the Tutorial runs a slide show tutorial to show SPSS's main features.

Type in data presents an empty untitled data window SPSS Data Editor.

Run an existing query will run a query to bring data in from a database.

Create new query using Database Capture Wizard takes you through creating your own query to extract data from a database or Excel.

Open an existing data source to select an SPSS or other type of data file to analyse.

Open another type of file to open any other sort of file.

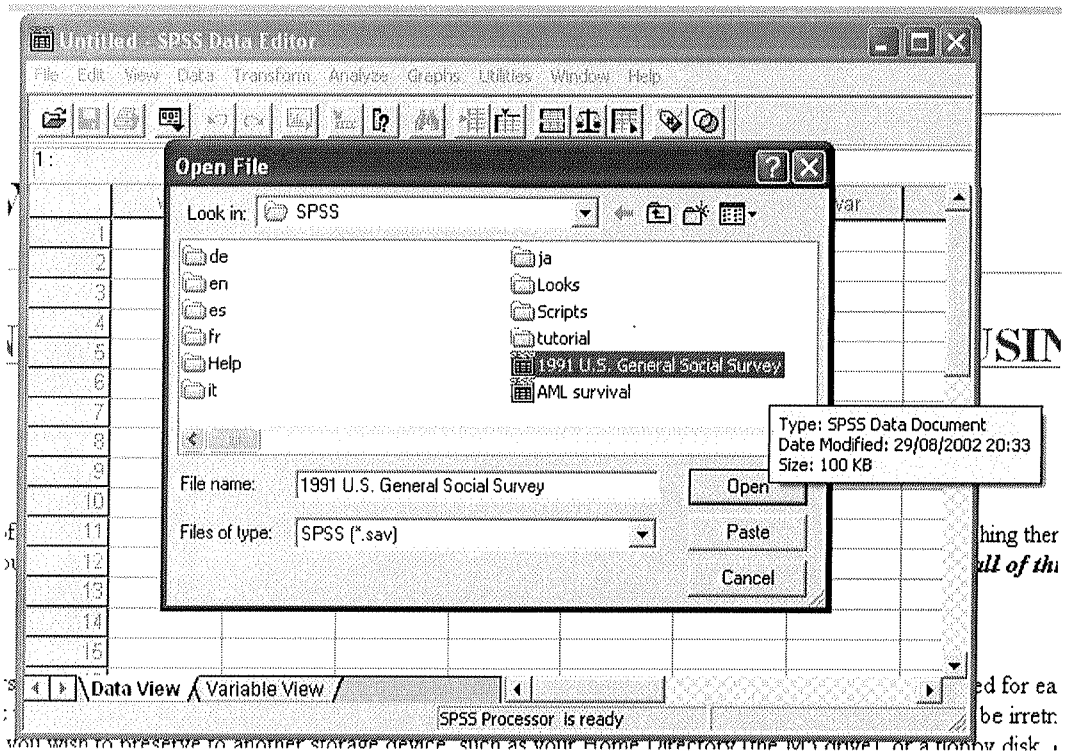
Don't show this dialog box in the future will stop the dialog box appearing in the future.

Click OK to choose Open an existing data source.

If you don't see the dialog box shown in Pic 2 then use the **Open File** button (first left on the tool bar, see Pic 9) or choose **File/Open/Data** from the menus to activate the **Open File** dialog box. Starting up SPSS without the dialog box should open the SPSS Data Editor window, as in Pic 4 but without any data in it.

We will use a data set called **1991 U.S. General Social Survey.sav** that comes with SPSS and in general there should be a copy of it in the folder where SPSS is installed. In the labs it is drive T, in the Skills centre there is a copy in the spsswork directory as well. On your own computer it is likely to be on the C or D drive in the **Program Files** or **Programs** folder. You might have to scroll to the right past lots of folders before you get to the data files.

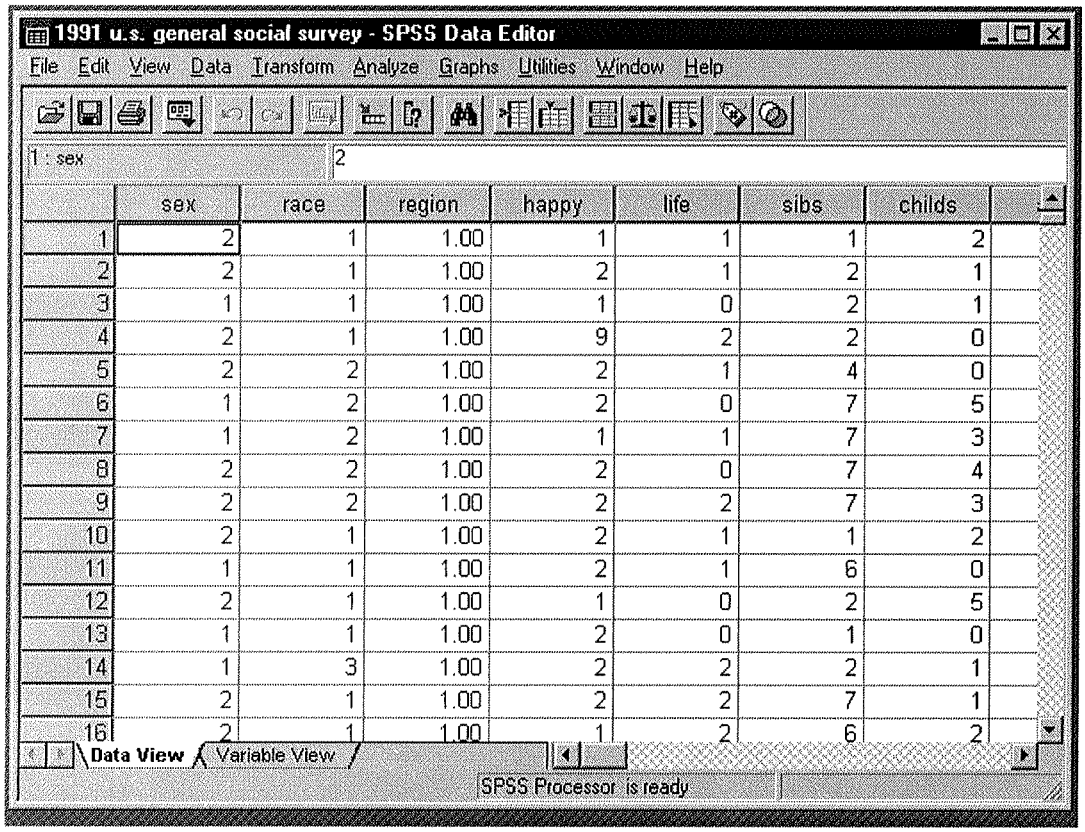
Pic 3: *Open File* dialog box showing data (.sav) files in the Spss11 folder



Select the file 1991 U.S. General Social Survey.sav and click Open to read the data file into the DataEditor window.

Data in the SPSS Data Editor

Pic 4: Survey data without value labels



	sex	race	region	happy	life	sibs	childs
1	2	1	1.00	1	1	1	2
2	2	1	1.00	2	1	2	1
3	1	1	1.00	1	0	2	1
4	2	1	1.00	9	2	2	0
5	2	2	1.00	2	1	4	0
6	1	2	1.00	2	0	7	5
7	1	2	1.00	1	1	7	3
8	2	2	1.00	2	0	7	4
9	2	2	1.00	2	2	7	3
10	2	1	1.00	2	1	1	2
11	1	1	1.00	2	1	6	0
12	2	1	1.00	1	0	2	5
13	1	1	1.00	2	0	1	0
14	1	3	1.00	2	2	2	1
15	2	1	1.00	2	2	7	1
16	2	1	1.00	1	2	6	2

Now you should be able to see the Data Editor window dominated by the grid used to display and change existing data or enter new data. Use the vertical and horizontal scroll bars to see more of the data, there are over 1500 rows representing the opinions of people from all over the states. Above the grid and below the title bar are the menu headings as seen in Pic 5.

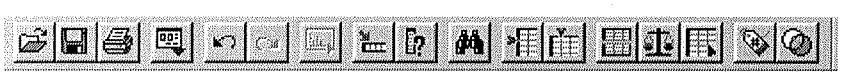
Pic 5: SPSS Data Editor detail, Menu Headings

Click each heading to look at the sort of items on each menu.

Most menu items are greyed at and cannot be selected until there is some data in the Data Editor. The **Help** menu contains lots of useful information about using SPSS including what all the menus and windows are for. Try **Help/Topics** and look at the various topics, including **SPSS at a glance**. **Help/Tutorial** guides you through the basics of running SPSS – use it to supplement these notes.

Just below the menu headings is the **Tool bar**, its buttons are shortcuts for many common tasks in SPSS rather than using the menus.

Pic 6: SPSS Data Editor detail, The Tool bar

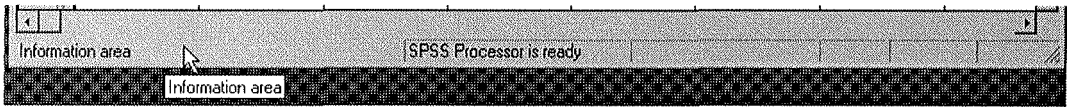


At the very bottom of the window you can see the **Status bar**. It shows the status of SPSS at the time, including the progress of any procedure and any special transformations in place (i.e. filters, splits or weights).

Pic 7: SPSS Data Editor detail, The Status bar

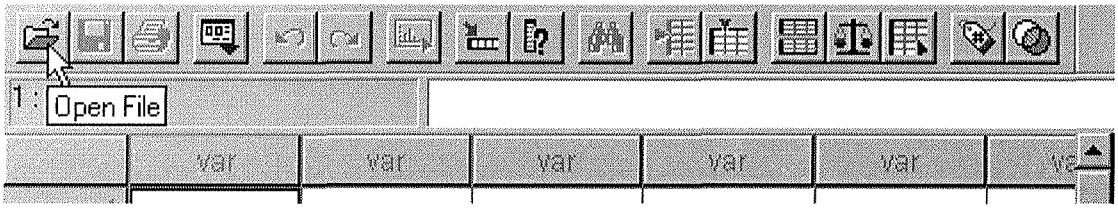
The message **SPSS Processor is ready** is showing in our example, indicating that SPSS is waiting for something to do. The first space on the status bar is known as the information area. It provides a phrase and sometimes a pop up label to describe whatever the mouse is pointing to.

Pic 8: Information area detail with pop-up label

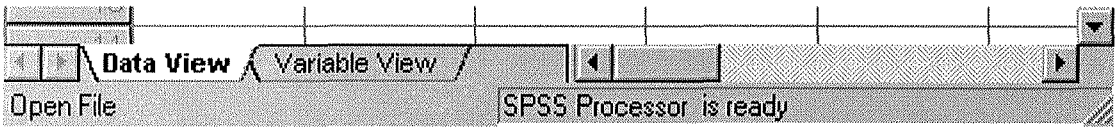


Place the mouse pointer over each button in the tool bar and watch the information area on the status bar at the bottom of the window as it changes.

Pic 9: The mouse over the Open File button and the Status bar



Data Editor grid in between.



The contents of both the tool bar and the menus will change depending on what type of SPSS window you are looking at. Toolbars can be changed or new ones can be created in the **Show Toolbars** dialog box, opened by choosing **View/Toolbars**. Menus can be customised using the **Menu Editor** opened by choosing **Utilities/Menu Editor**.

Data can either be typed directly into the **Data Editor**'s **Data View** or an existing data file can be read in. The Data Editor looks like a spreadsheet, but has a much more rigid data structure.

SPSS phrase book

CASE: Each row of the data editor represents a **case**. Cases are the units or individuals that make up the study, e.g. survey responses, products in an inventory, schools in a district, countries, plots in an agricultural experiment, etc.

VARIABLE: Each column of the data editor represents a **variable**. A variable is a particular type of measurement for each case, e.g. it could be someone's height, a code representing their social class or a recording of their opinion on some subject.

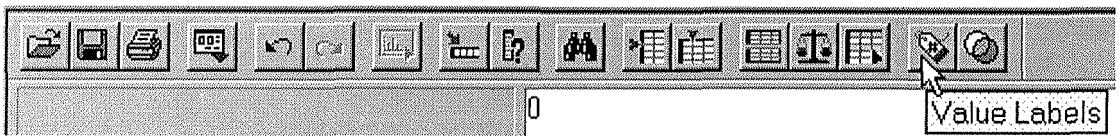
VARIABLE NAME: Above the data in the Data View, each variable has a **variable name.** In Pic 4, you can see the variable names sex, race, region, happy, life and sibs above the data. Variable names can be up to eight characters in length and must start with a letter. There must be no spaces, full stops, apostrophes or other punctuation marks in the variable name.

VARIABLE LABEL: Variables can also have a **variable label.** It is a longer bit of text to describe the variable. SPSS uses the variable label in the output when the variable is used and you can see what the label is when you point the mouse at the head of the variable as in Pic 13.

VARIABLE VALUE: Each cell in the Data View is known as a **variable value**. It represents the variable measurement for that particular case or row. These variable values can be numbers, numbers representing codes (nominal/ordinal) or text. SPSS assumes that the data is numeric unless it is told otherwise.

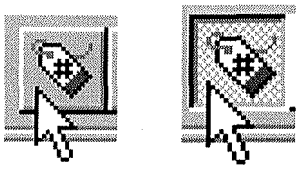
VALUE LABEL: Sometimes a piece of text called a **value label** is associated with each variable value. You should see the data with the value labels switched off. They can be displayed by using the **Value labels** button situated second from the right on the tool bar.

Pic 10: Value labels button



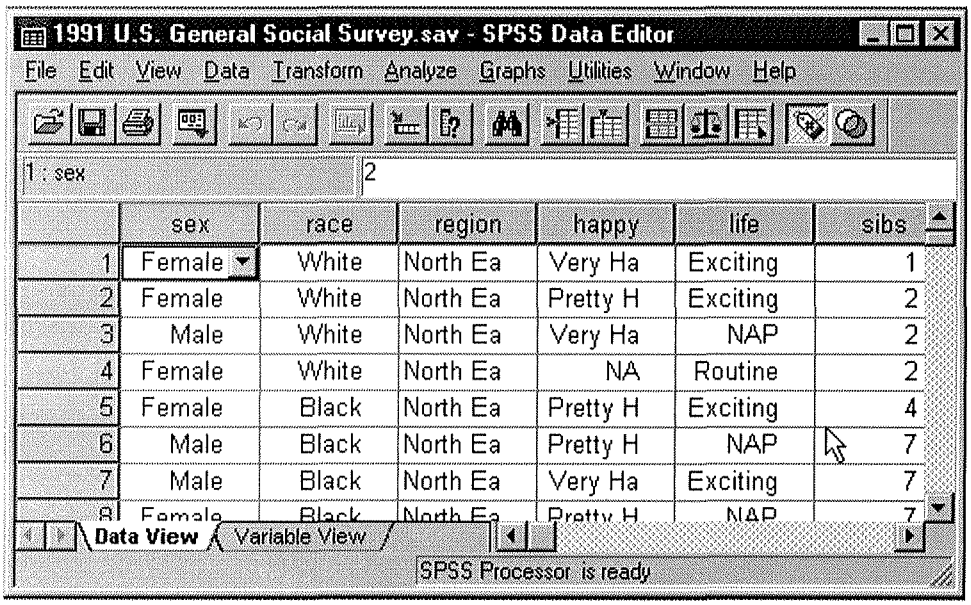
Click the **Value labels** button on the tool bar to display the labels if you cannot see them.

Pic 11: Value labels button **OFF** **ON**



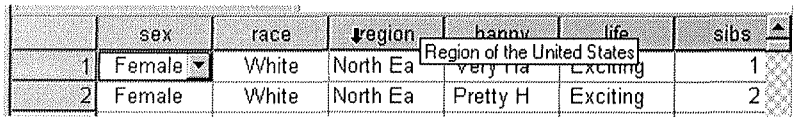
Click on **Value Labels** button again so you can see the labels.

Pic 12: Survey data with value labels shown



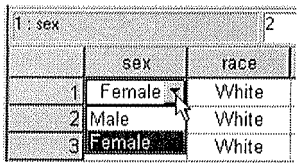
Point at a variable name at the top of the variable column to see its variable label, as shown below in Pic 13.

Pic 13: Detail of Variable label for region in the Data Editor



You can change a variable value just by selecting it and typing in a new code. If **Value Labels** are switched on, you can also change it by choosing another label from the list of value labels or more tediously by typing in the whole value label exactly as it appears in the list.

Pic 14: The list of value labels for the sex variable



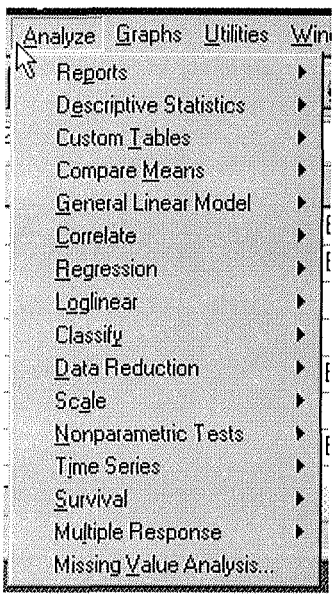
Select a cell in the Data Editor and click on the small arrow beside it to see the list of labels defined for that variable.

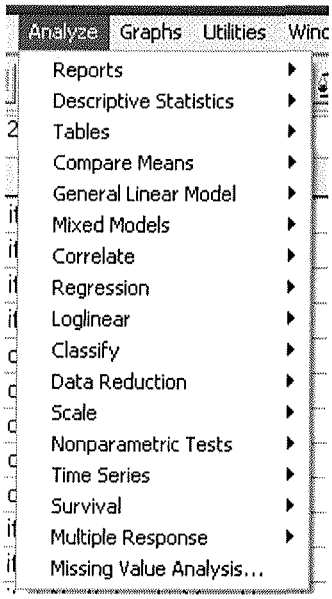
*If you change the data remember to return the data to its original state before going on or use the **File /Open** menu to open the original system file, clicking on **No** when it asks you to save changes to the data.*

Running an analysis

Now to look at how you would run an analysis in SPSS. The **Analyze** menu contains the statistical analyses available. We will run through one procedure that deals with variables individually (univariate analysis) and another that deals for a tabulation using two variables together. From there you should be able to apply the same principles to other analyses.

Pic 15 & Pic 15a: Different sorts of statistical analysis available for SPSS 10 and SPSS 11.5





Choose Analyze from the menus to see what is available.

An arrow beside a menu item indicates that it has a sub menu. Notice all the items in the **Analyze** menu have submenus. The list of menu items shown here is a fairly standard SPSS installation. Move the mouse over each arrow to see a list of the procedures available in each sub menu. Other items may appear on this menu depending on what modules are installed and whether the menus have been customised.

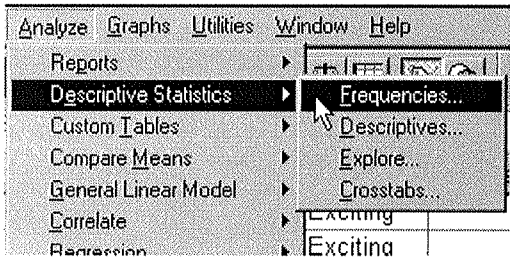
From the Analyze menu, choose Descriptive Statistics to see a list of the descriptive procedures.

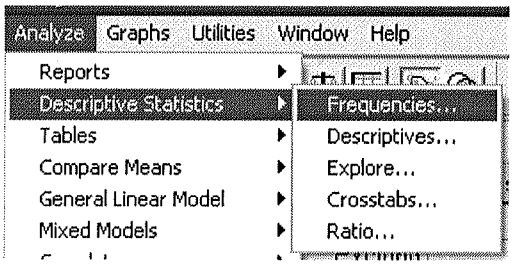
*These notes are intended to show you how to use SPSS, not to teach you statistics. There is a **Statistics Coach** in the **Help** menu to help you choose the correct analysis for your data. The coach is not very detailed so get take advantage on any statistics courses your department may run or get a good introductory statistics book.*

Frequencies

The **Frequencies** procedure generates summary tables of frequencies of each value for a variable. It can also be used to calculate summary statistics and charts for each variable individually as well.

Pic 16 & Pic 16a: Items from the Analyze/ Descriptive Statistics submenu for SPSS 10 and SPSS11.5

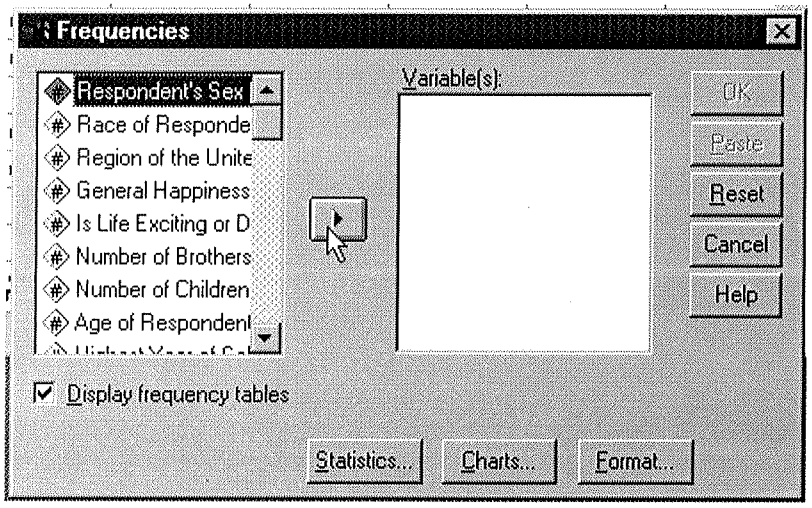




Now choose Analyze/Descriptive Statistics/Frequencies.

The **Frequencies** dialog box consists of two boxes of variable lists, below them are buttons to customise the procedure results and to the right of the lists are the five standard dialog buttons. To produce frequencies tables for a variable simply select it from the left hand list of all the variables in the data set and click on the arrow button to put it in the **Variable(s):** box.

Pic 17: Frequencies dialog box



The three buttons along the bottom of the dialog box are for making additions or changes to the default output. **Statistics...** is used to add summary statistics for each variable in the **Variable(s)** list. **Charts...** is used to produce a graph for each variable in the **Variable(s)** list. You can choose between a bar chart, pie chart or histogram. **Format...** is used to change the format of the standard frequency table and the order objects are displayed in.

The five buttons down the right hand side of the box are common to most dialog boxes.

Clicking **OK** tells SPSS that the dialog box is complete and is to be acted on.

Paste will paste the underlying SPSS command into a syntax window - the SPSS commands course deals with this in more detail.

Reset can be used to clear the dialog box of any selections made so you can start specifying the procedure from scratch.

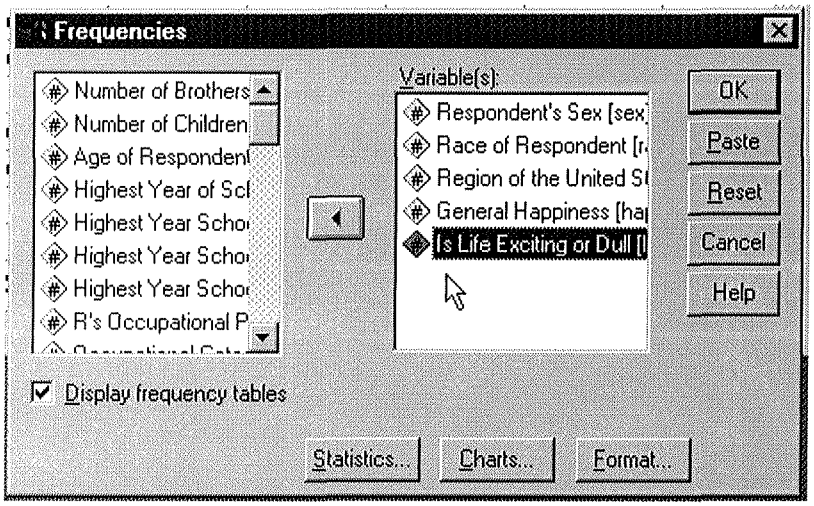
Cancel is used to exit from the dialog box without running the procedure.

Help has information about the procedure and how to use the dialog box.

Above in Pic 17 the first variable in the list, Respondent's Sex (gender), has been selected and is about to be put in the **Variable(s):** box using the arrow button.

Put a few categorical variables in the Variable(s) list, e.g. variables sex, race, region, happy and life were chosen in Pic 18.

Pic 18: Selecting variables for Frequencies[1]



Double clicking on the variable will also transfer it into the **Variable(s):** box. Once there, if you decide not to use a variable select it in the **Variable(s):** list and move it back using the arrow button which should be pointing back to the original list or double click on it again.

Notice that the **OK** button (and **Paste** button) is active now in Pic 18 – as soon as SPSS has enough information to run a procedure it will activate the **OK** button. By comparison the OK button in Pic 17 is inactive and looks faded or "greyed out". You could run the **Frequencies** procedure just now if you wanted to[2]. However we will look at the procedure in a bit more detail before we run it, just to heighten the sense of anticipation!

---

[1] Each variable has a small icon beside it, showing # for a numeric variable (i.e. values are numbers). A string or text variable will show $A_<$ or $A_>$ for ones over 8 characters wide.

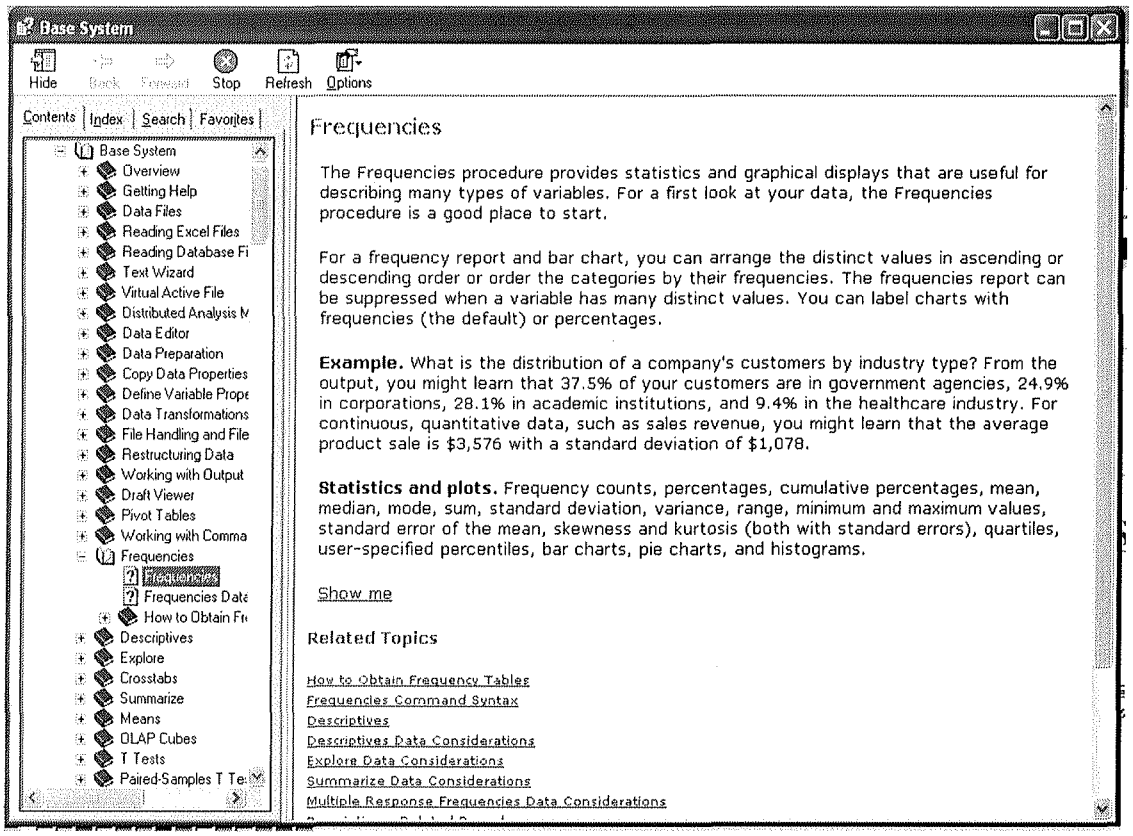
[2] If you do run Frequencies at the moment, then choose **Analyze/ Descriptive Statistics/ Frequencies** from the menus to continue with the next section. Page 19 will tell you about the output.

More information on procedures

Most Windows applications have a Help facility where you can get more information. It is usually accessible from the **Help** menu or by a **Help** button in SPSS dialog boxes. The **Help** button is "context sensitive", i.e. in the **Frequencies** dialog box it points directly to the information about **Frequencies** in the Help system. Any time you lose track of what you are doing in a dialog box, use the Help button or the Help menu to find out more about where you are.

Click Help button and SPSS will tell you more about Frequencies.

Pic 19: Frequencies information in the Help window in SPSS 11.5



Click Show me to open a tutorial about Frequencies or click a Related Topics link to find out more. In SPSS 10 the See Also button in the Help window will show you related topics.
Once you've finished reading, close the Help (and Tutorial) window and get back to SPSS and the Frequencies dialog box.

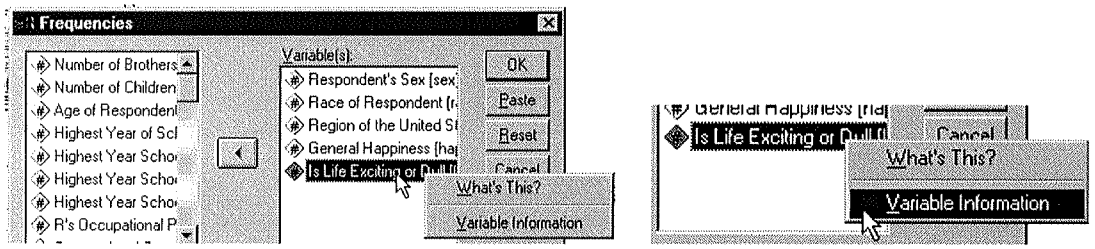
Context Sensitive Menus

A "context sensitive" menu can be activated for an object in Windows by selecting it and clicking on it with the right hand mouse button. The contents of menu will depend on the thing you selected, showing what you can do in that context. Normally one of the items is called "**What's This?**" or "**Properties**" and it will tell you about the selected object.

Information about variables

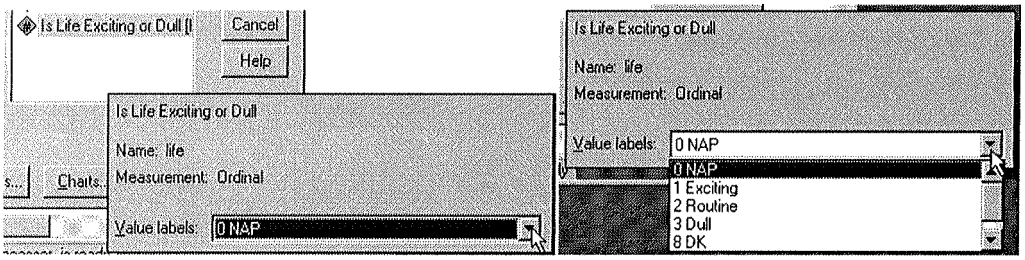
A variable information box is available via a context sensitive pop-up menu from SPSS dialog boxes. To do the following make sure the **Frequencies** dialog box is open.

Pic 20 & Pic 20a: Pop-up menu & choosing Variable Information



Click on one or two variables in either list using the *right* mouse button to display a pop-up menu similar to Pic 20. Choose Variable Information from the menu to see information about it.

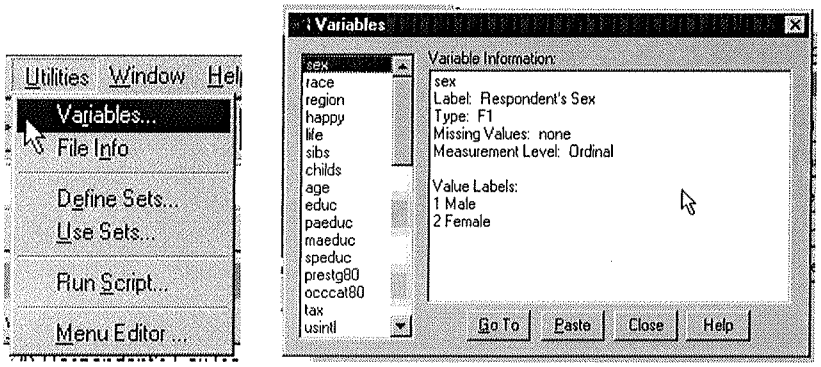
Pic 21 & Pic 21a: Variable information for life variable & value labels



Notice as well as the rest of the variable definition there is a drop-down menu of defined value labels. If you click (left mouse click this time) on the dialog box in the background then the variable information will disappear. The other way you can get information about the variables is by selecting **Variables...** from the **Utilities** menu to open the **Variables** window.

Choose Utilities/Variables... to see the Variables window.

Pic 22 & Pic 22a: Choosing Utilities/Variables & Variables window



The **Variables** window contains the same sort of information in Pic 21, but in a window rather than a pop-up menu. The window can remain open in the background while you work elsewhere in SPSS. It can also be opened at practically anytime for reference in SPSS.

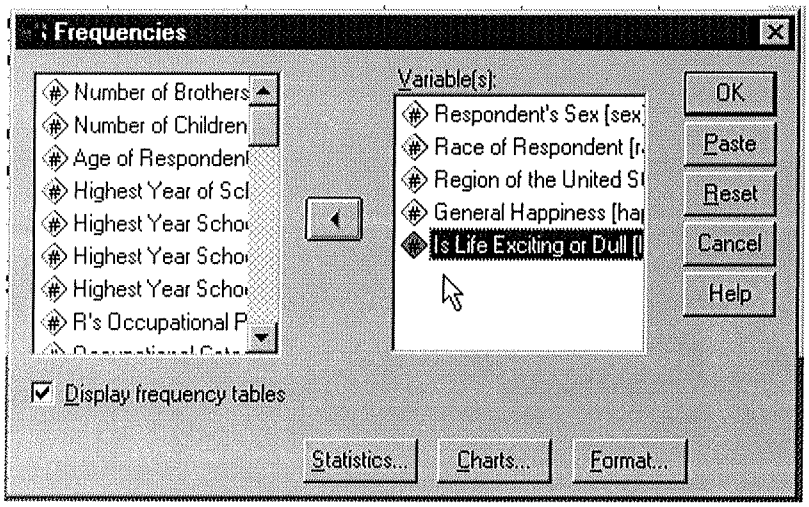
Click the Close button to close the Variables window, so the Frequencies dialog box is at the front.

We will now look at some ways we can customise the Frequencies output using the **Charts...** and **Format...** buttons. Most of the statistics available via **Statistics...**button are more suitable for scale variables than the categorical ones we have chosen – so we will not bother with them at the moment.

*If you've lost the **Frequencies** dialog box, choose **Analyze/ Descriptive Statistics/ Frequencies** from the menus again and SPSS will bring it back to the front.*

Just in case you have forgotten what it looks like, Pic 23 (a copy of Pic 18) shows the dialog box before we got diverted.

Pic 23: Showing the variables selected for Frequencies

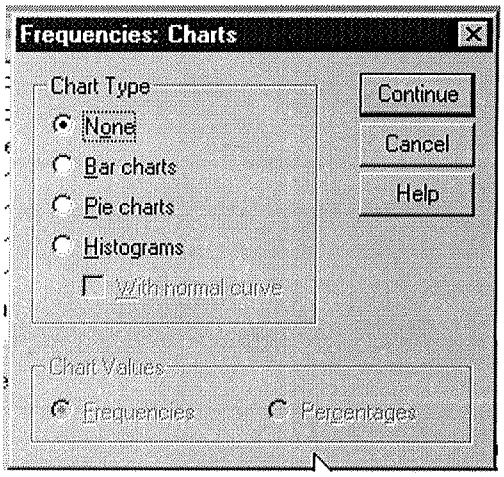


Frequencies: Charts, Specifying a bar chart

The **Charts...** button on the **Frequencies** dialog box opens the **Frequencies: Charts** dialog box where you can choose to produce graphical equivalents of the frequency tables. There are three sorts of chart available. **Bar charts** and **Pie Charts** are best for variables with few distinct values, i.e. categorical (nominal or ordinal) variables. **Histograms** are best for variables with lots of values measured on a scale, i.e. quantitative variables.

Click on the Charts... button to bring up the Frequencies: Charts box.

Pic 24: Frequencies: Charts dialog box



Click on Bar Charts to produce bar charts with the tables.
Then click Continue to go back to the main Frequencies dialog box.

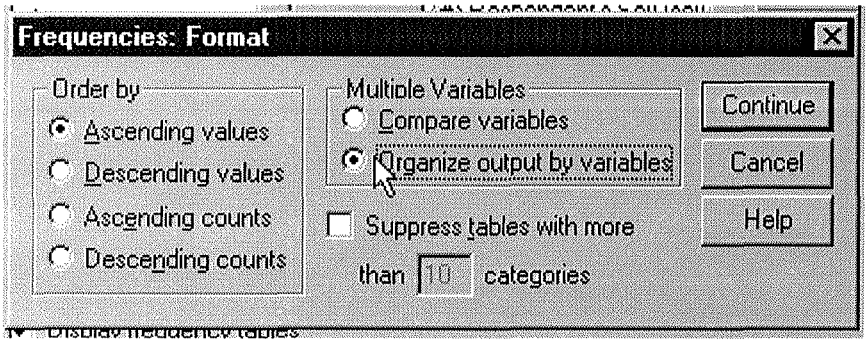
*Make sure you selected **Bar Charts** and then clicked **Continue**. If you click **Cancel** or otherwise close the dialog box SPSS will close the **Frequencies: Chart** box without saving your selection and no chart will be specified.*

Frequencies Format

The **Format...** button opens the **Frequencies: Format** dialog box. It contains options to change the order of the Frequencies output or even suppress the output under certain conditions. We will change the setting and use **Organize output by variables** so we can see the frequency table for a variable beside its pie chart.

Click the Format... button and choose Organize output by variables button. Click Continue to get back to the Frequencies dialog box to run it.

Pic 25: Selecting Organize output by variables



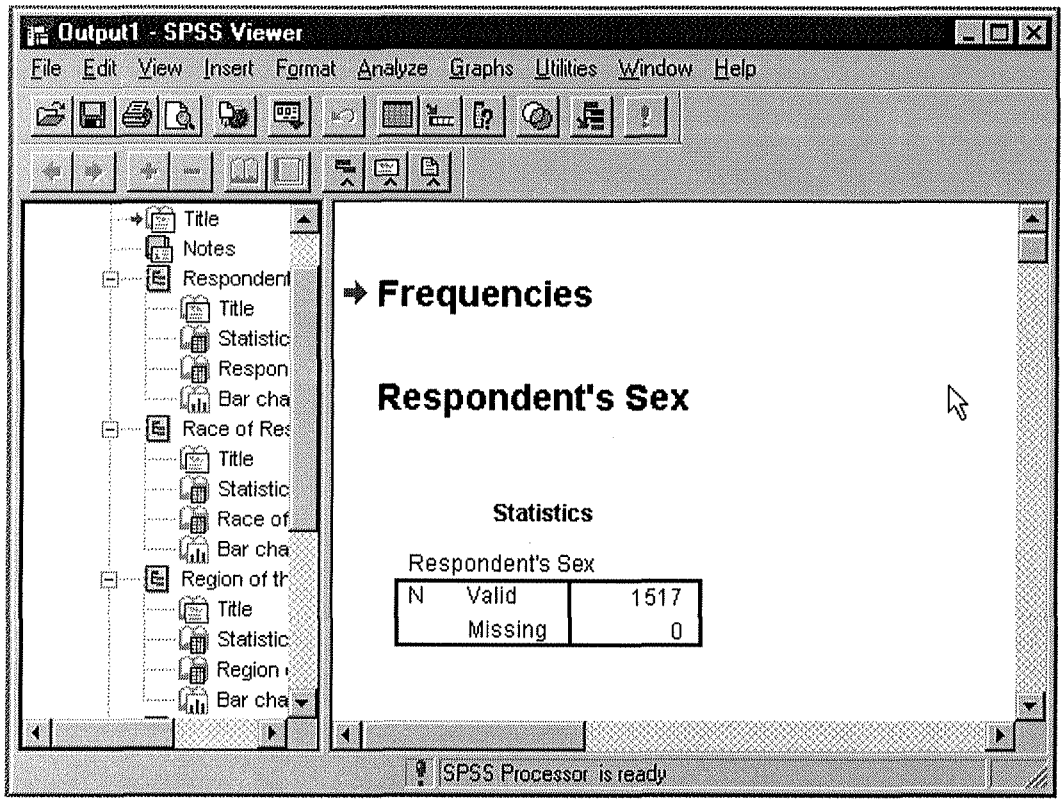
Click the OK button in the Frequencies dialog box to run it.

*Again make sure you click **Continue** after making your selection to save your selection.*

The output viewer

As soon as you click **OK** you should see the SPSS Output viewer come to the front. And depending on how fast the computer is, you may see the status bar at the bottom change to **Running FREQUENCIES....** before the results are written to the Output window. It should change back to **SPSS Processor is ready** as illustrated below when the procedure has finished.

Pic 26: *A first view of Frequencies output*



The **SPSS Output Viewer** will usually come to the front as soon as a statistical procedure is run. The Output window is split vertically into two panes. Above the split window, the menus have changed and different buttons are active on the tool bar to reflect the fact they are for use with different sort of window. The two "panes" of the viewer each have separate scroll bars which means you can scroll down the results or you can scroll through the outline or click on an icon to jump to that bit of the results.

The **left** side of the output window is the viewer outline, consisting of small icons representing the different parts of the SPSS output.

The **right** hand side contains the output itself.

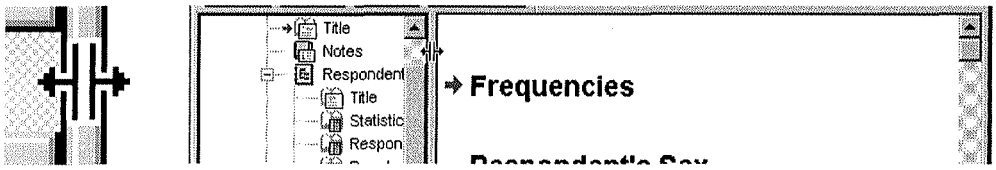
In Pic 26 you can see the start of frequency output with the default summary statistics table for the first variable from our list, **Respondent's Sex. N Valid** is the number of cases with valid values, i.e. there are 1517 where gender has been recorded and **N Missing** is the number of missing cases, where there is no valid value, none in this example. If we had used the **Statistics...** button to add more summary statistics they would appear in this table.

Hiding the outline

The outline side of the output window can be easily hidden to give you a more space on the screen to inspect the results.

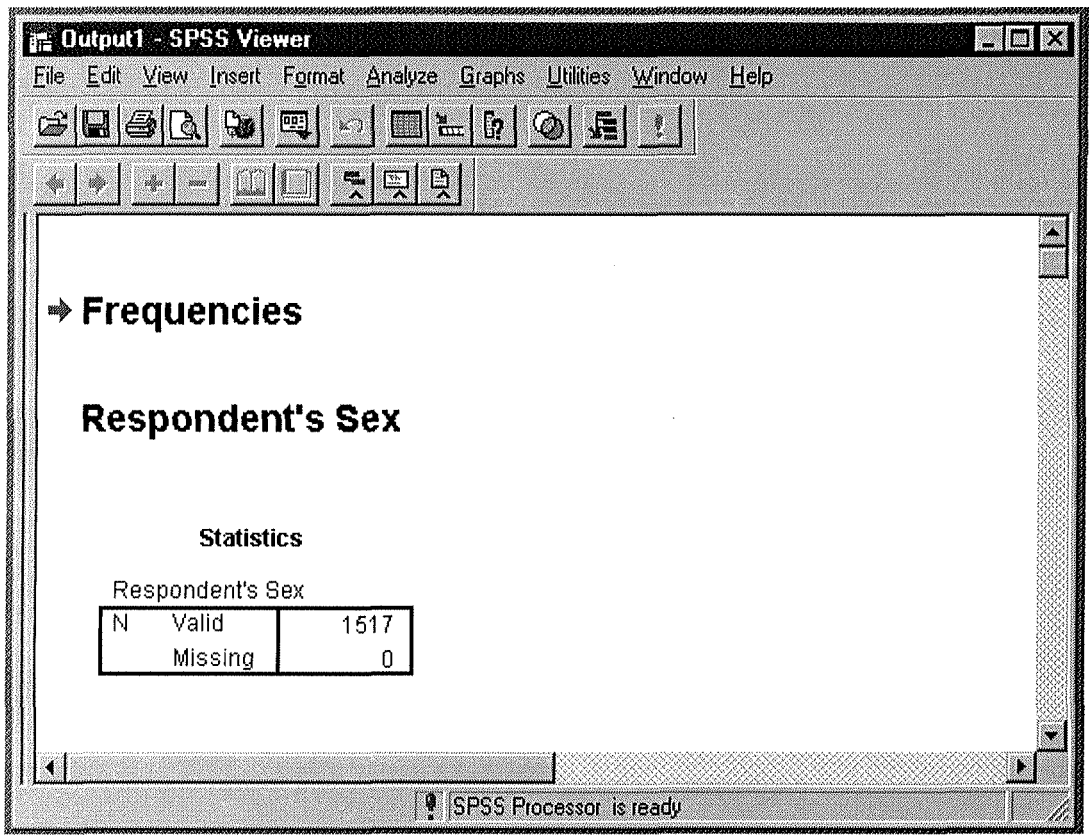
Place the mouse pointer over the divide between the two parts of the window. You should see the pointer change to a double-headed arrow.

Pic 27& Pic 27a: Double headed arrow & dragging to hide the Outline



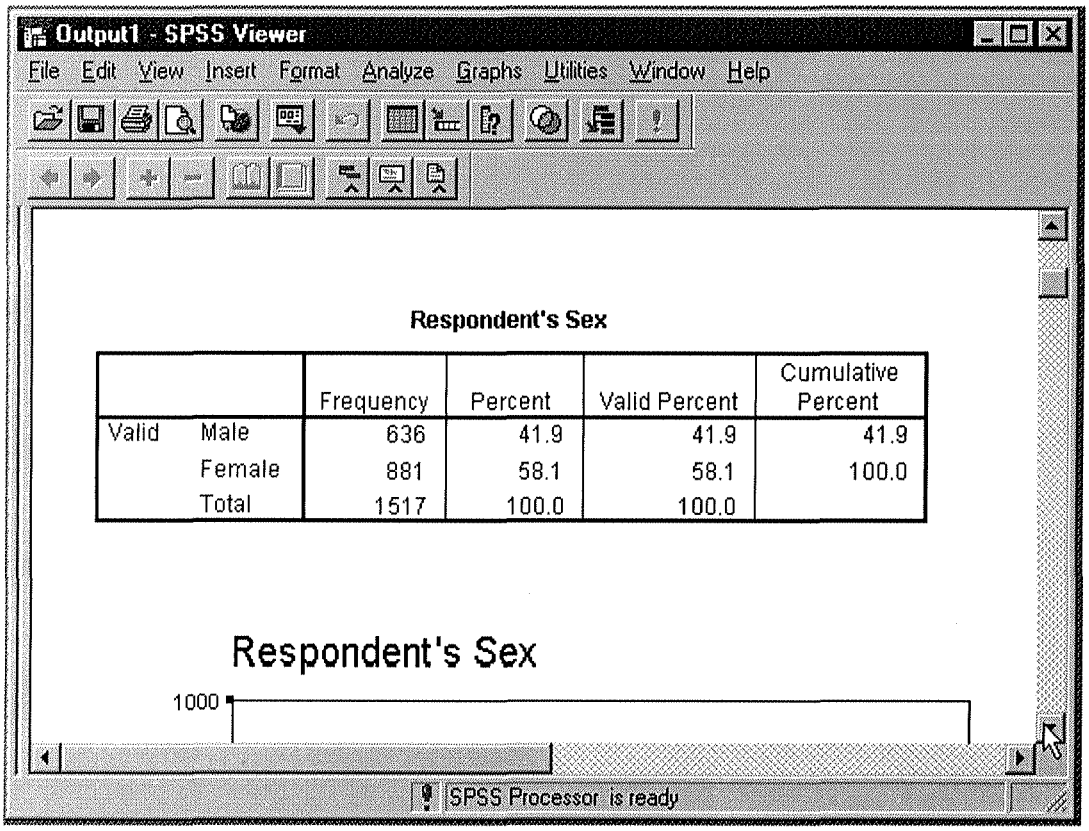
Click and drag the division over to the left most point and the outline will be hidden.

Pic 28: Results window with no outline



Use the scroll bars to inspect the rest of the output.

Pic 29: Scrolling down - showing Frequency table for sex variable



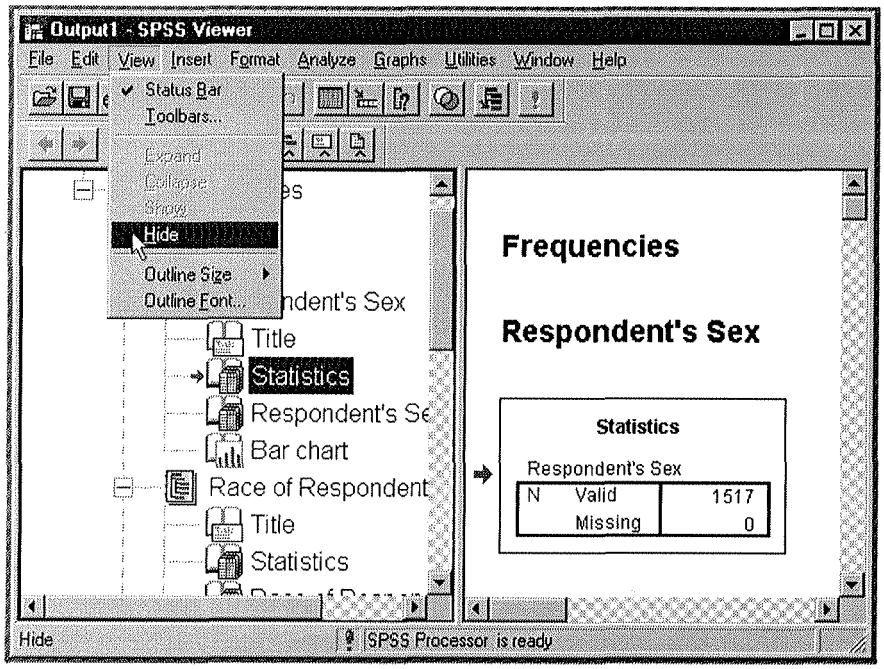
The outline can be seen again by clicking on the left hand side of the window and dragging the division bar towards the middle of the window. (mouse should be the double headed arrow again) This method can also be used to change the relative size of the panes by dragging the bar between them but just not all the way across.

Restore the outline by clicking on the left hand side of the window and dragging the division bar into the middle again.

Using the outline

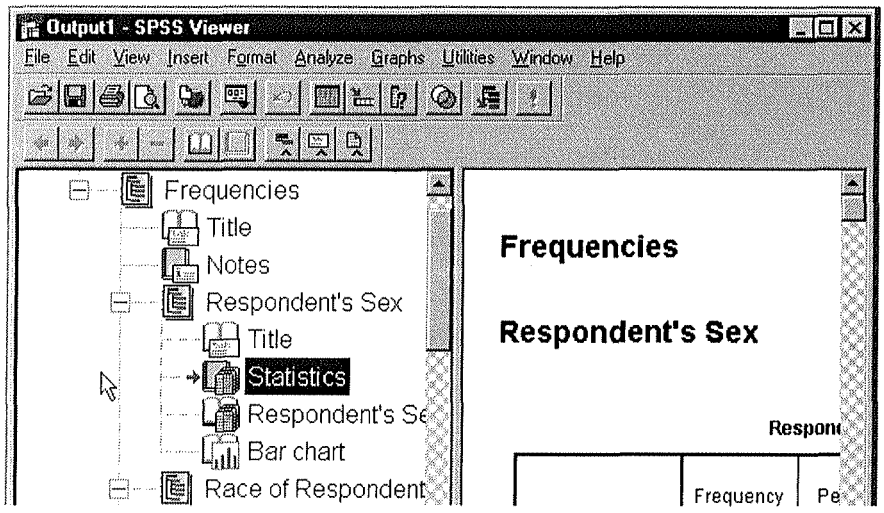
The outline can be used to organise the results you want to see to jump to a particular section of output. It is possible to hide items in the results by selecting it in the outline and "hiding" them. Hide a selected icon by choosing **View/Hide** from the menus or double clicking on the outline icon. Click once to go to the output the icon represents and double click to hide the output.

Pic 30: Using the View menu to hide some output



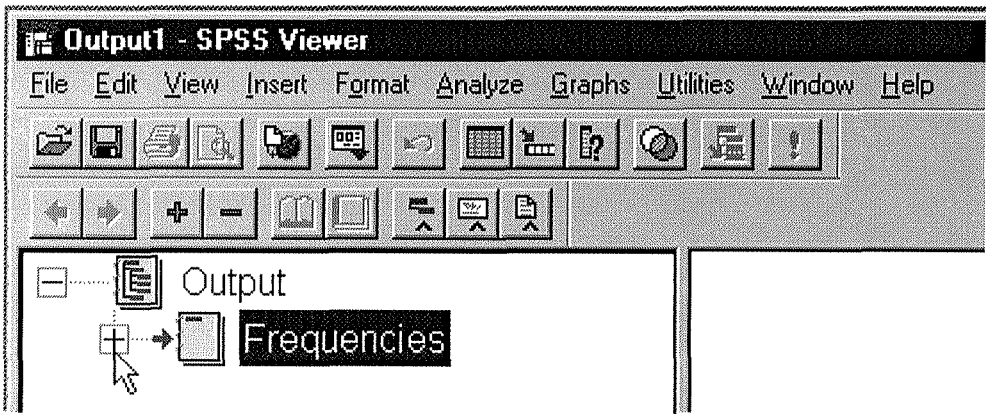
Select an item from the outline and choose Hide from the View menu.

Pic 31: The result of hiding the Statistics output



Double clicking on an icon in the outline will hide it, so the item is hidden in the results window and the icon looks "closed". Double click on the icon again and it will open. Pic 32, the figure below shows the whole Frequencies output block closed, just by double clicking on the icon.

Pic 32: Frequencies block icon closed



You can see in Pic 32 what the output icon representing an output block looks like when it's open, labelled **Output**, and when it is closed, the one labelled **Frequencies**. Also the small minus, -, and plus, +, buttons can be pushed to close or open an output block respectively.

About the Frequencies output

There are three different parts in Frequencies output, the summary statistics table, the frequency table itself and a chart if one has been specified.

The frequency table list the number and percentage of cases with each variable value, e.g. in Pic 33, the number in the sample who consider life dull was 41 and they were 2.7% of the whole sample and 4.2% of the people asked.

Pic 33: Frequency table for the life variable

Is Life Exciting or Dull

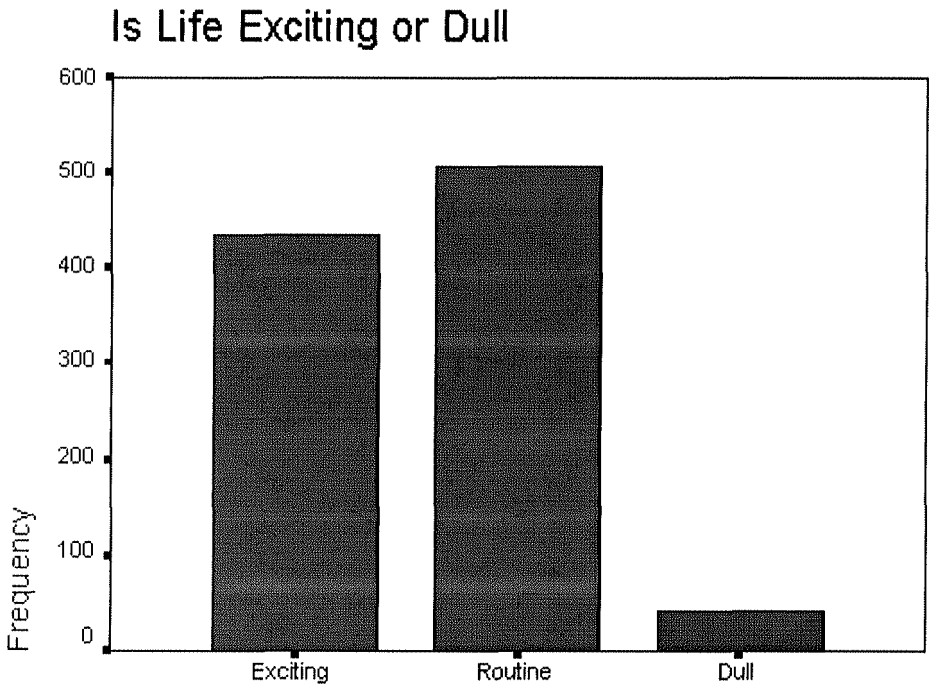
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Exciting	434	28.6	44.3	44.3
	Routine	505	33.3	51.5	95.8
	Dull	41	2.7	4.2	100.0
	Total	980	64.6	100.0	
Missing	NAP	524	34.5		
	DK	8	.5		
	NA	5	.3		
	Total	537	35.4		
Total		1517	100.0		

Look at the frequency tables for other variables in the Frequencies output and write down the number and percentage in the sample who:

1. **lived in the North East region**
2. **were very happy at the time of the survey.**

The charts are visual illustrations of the frequency table. In Pic 34 you can see the proportion of people who find life dull from the bar chart in relation to those who found life exciting and routine.

Pic 34: Barchart for the life variable



Notice that there are no bars representing missing values, the charts created with the Frequencies procedure only has the valid values in it.

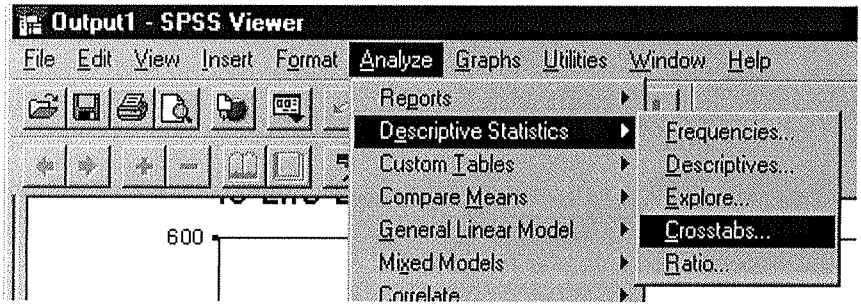
*Choosing **Graphs/Bar...** from the menus will permit you to create a chart showing the relative number of missing values. Look out for the chart tour coming soon which will have more on how to do this.*

Crosstabs

Frequencies deals with variables separately but **Crosstabs** displays two or more categorical variables in association with each other in a table. In fact it's called a crosstabulation and it's in the same menu as Frequencies, so choose **Analyze/ Descriptive Statistics/ Crosstabs....** will open the Crosstabs dialog box.

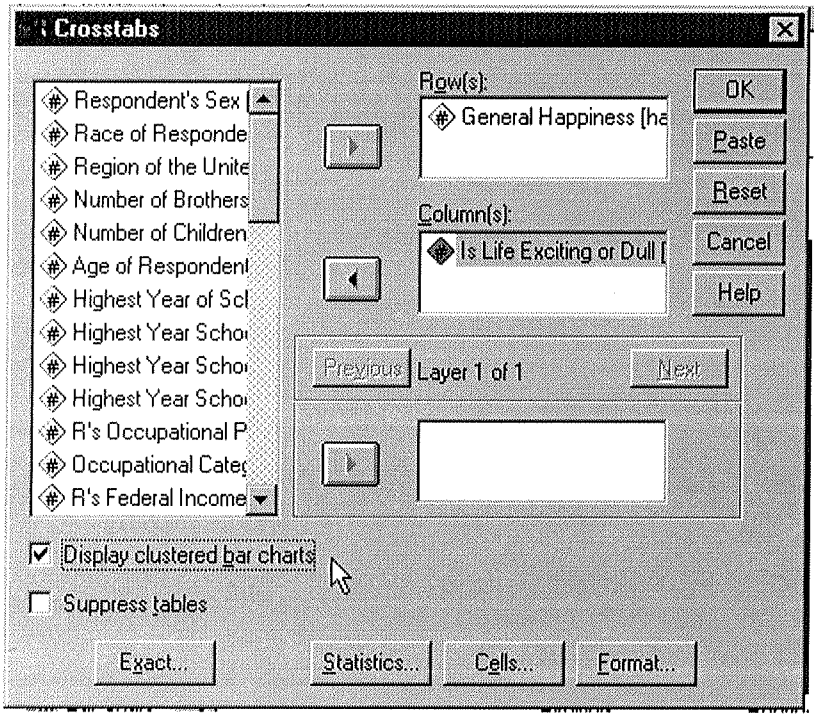
Choose Analyze/ Descriptive Statistics/ Crosstabs....

Pic 35: Selecting Crosstabs



Select the General Happiness (happy) variable and put it in Row(s):
Select the Is Life Exciting or Dull (life) variable put it in Column(s):
Click the Display clustered bar charts option, as shown above.
Click OK to run the Crosstabs procedure.

Pic 36: Crosstabs dialog box



When it is finished you should be looking at the Output Viewer again. You can see in the outline that there are several blocks to the Crosstabs output including the table below in Pic 37 and the bar chart shown in Pic 38.

Pic 37: *Crosstabs table*

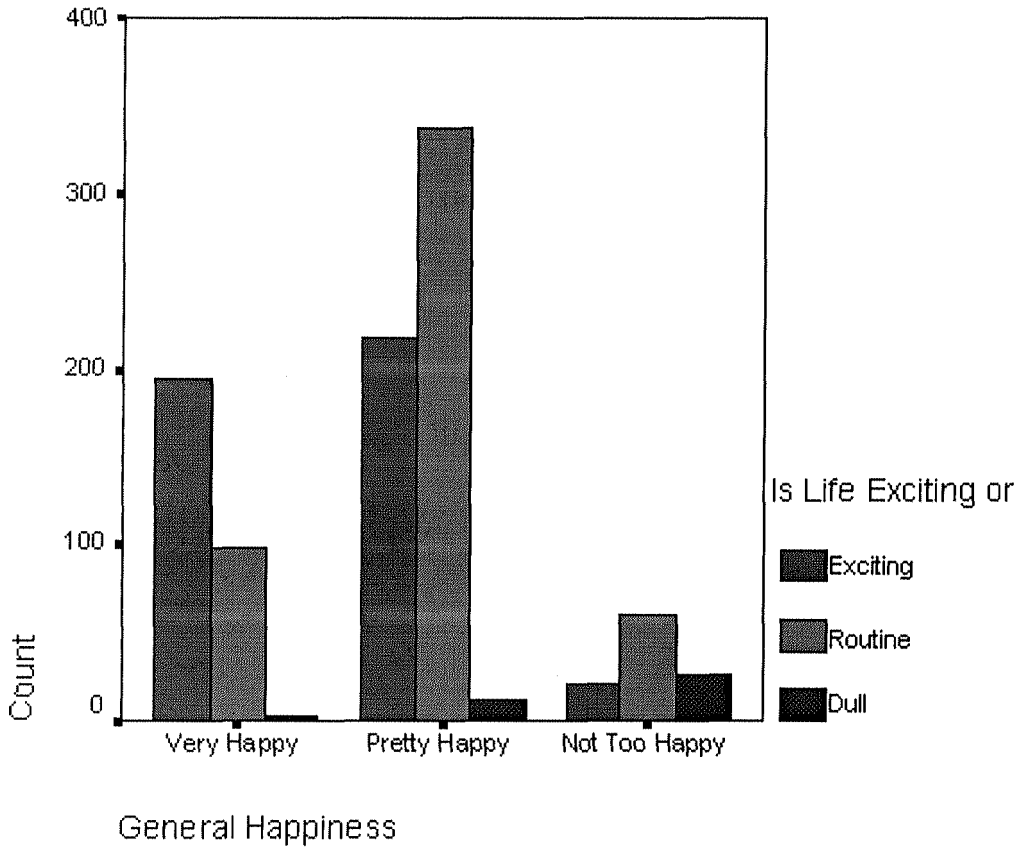
General Happiness * Is Life Exciting or Dull Crosstabulation

Count

		Is Life Exciting or Dull			
		Exciting	Routine	Dull	Total
General Happiness	Very Happy	195	98	2	295
	Pretty Happy	218	338	12	568
	Not Too Happy	21	61	26	108
Total		434	497	40	971

Each valid value of the row variable will define a row and each valid value of the column variable will define a column of the table.

Pic 38: *Crosstabs Clustered bar chart*

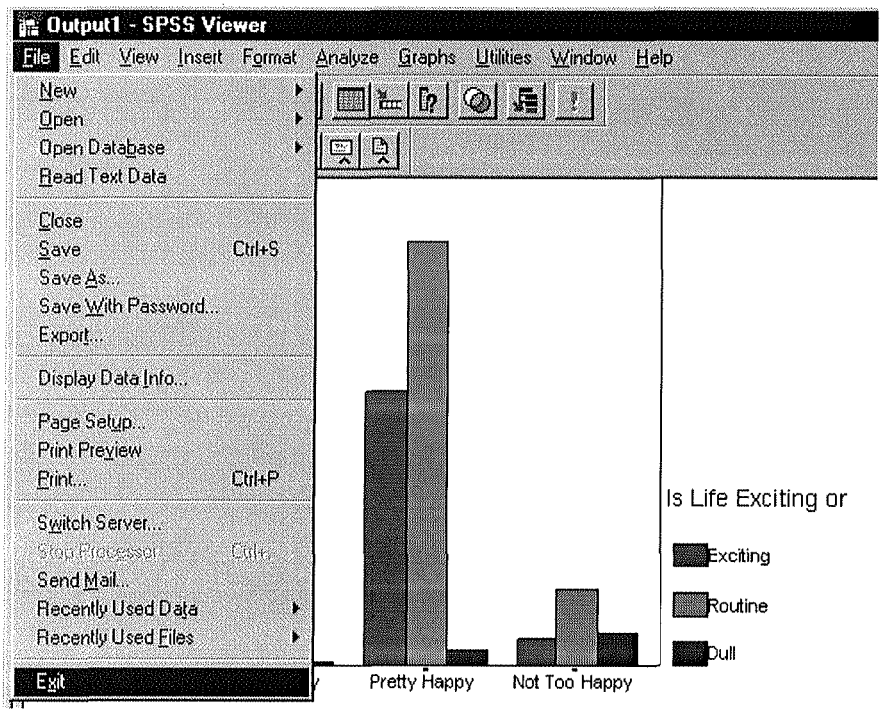


Tables and charts can be customised to suit your own taste. Look out for the table tour and the chart tour coming soon!

Leaving SPSS

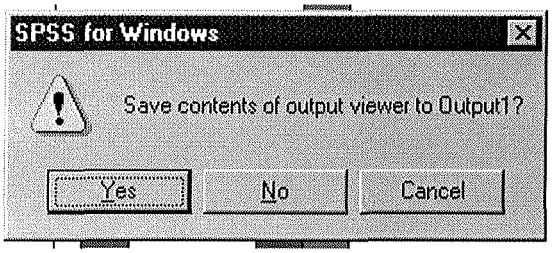
To finish the SPSS session, select **File /Exit** from the menus.

Pic 39: Selecting Exit from the File menu



Before SPSS shuts down you will be asked about saving the contents of each window in turn, in this case the output viewer window and the data window.

Pic 40: Warning to save the contents of the Output viewer



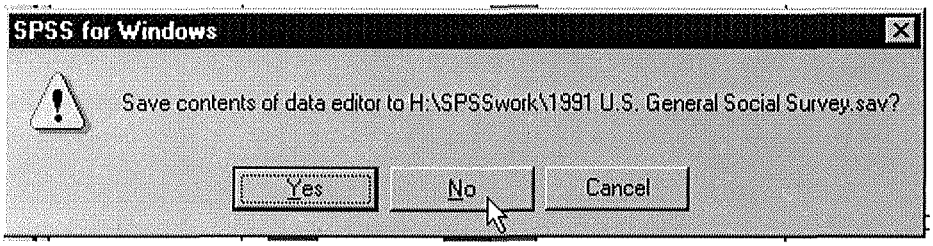
For example in Pic 40, SPSS is asking whether to save the contents of the output viewer. The three buttons have similar functions whatever the type of window that is about to be closed.

Yes saves the contents to a file, either to a named file or a dialog box is opened to name a file where that window's contents are to be stored.

No closes the window and none of the changes you have made during the session will be saved (unless you've used the save option in the file menu).

Cancel stops SPSS closing and leave the current window open.

Pic 41: Warning to save the contents of the data file



Click No for each of these dialog boxes, unless you wish to keep the data file or the output you made.

It is not essential to keep the contents of any of the windows created during this session – unless you want to keep a souvenir of the tour.

Each saved window can be opened up in SPSS, using the **File /Open** from the menus. Shortcuts for the most recently used files are available in the File menu under **File/Recently Used Data** and **File/Recently Used Files**. Choosing the shortcut from the menu will open the file in SPSS as long as it hasn't been deleted in the mean time.

SPSS will associate a default file extension with different types of file. You can see these extensions in the drop down menu, **Files of type:** in **Open File** dialog box. The two most common extensions are `.SAV` and `.SPO`. **SPSS data windows** are saved to files with **.SAV** extensions and **Output windows** are given **.SPO** extensions.

ⓘ Information

This section contains some hopefully helpful nuggets of information.

Things to do... including SPSS courses and on-line notes.

Places to visit... such as useful web sites

People to see... about getting help with SPSS and computers.

ⓘ *If any of the web links don't work try searching from the homepages for the University of Edinburgh at,* `http://www.ed.ac.uk`
or Computing Services at, `http://www.ucs.ed.ac.uk/`

Things to do...

Go on Computing Services courses

There are three SPSS courses offered once a term by Computing Services: **1195** Introduction to SPSS, **1196** Intermediate SPSS and **1203** SPSS commands.

These courses are primarily aimed at postgraduates and staff but all the details are at,
`http://www.ucs.ed.ac.uk/usd/iss/courses/`

Go on courses in your department

If you are an undergraduate or taught postgraduate please check with your department (or course supervisor) – there maybe courses run for you there.

Look at SPSS notes on-line

There is a whole suite of on-line notes from the SPSS company in Adobe Acrobat format and the Statistical Analysis section will give you lots of examples of things you can do. Follow the link from the SPSS at Edinburgh web page at
`http://www.ucs.ed.ac.uk/usd/stats/spss.html`

Go on another tour

We hope that this will be the first of many tours, find out about them at
`http://www.ucs.ed.ac.uk/usd/stats/spsstours.html`

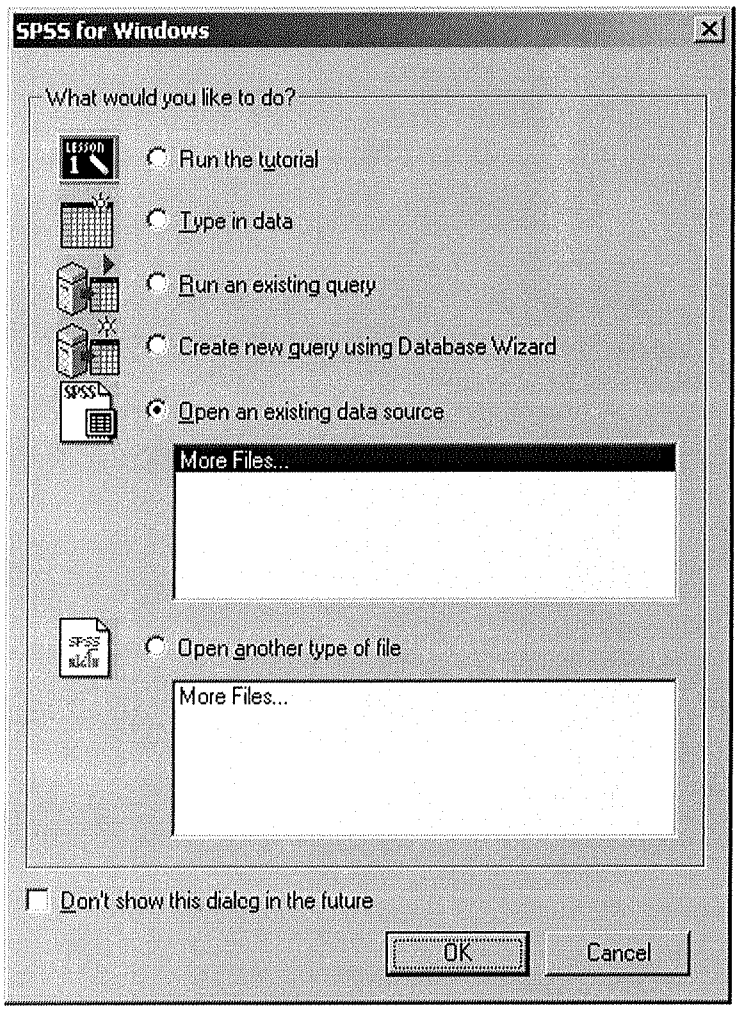
1. Introduction

SPSS is a package developed originally for Social Scientists using large mainframe computers. Since then it has been refined and redeveloped for different types of architecture including Windows. It is an extensive package with facilities for data entry, data manipulation and statistical analysis in a graphical environment. It has modules for survey analysis, graphical display and time series. The package has been considerably improved to include logistic regression and repeated measures analysis and much more beyond the scope of this tutorial.

2. Getting Started

This tutorial document assumes that you have a basic knowledge of Windows. The Information Technology Services (ITS) provide introductory courses and documents for this purpose. This tutorial is based on the default installation of the software at Reading where your Unix **home** directory will be available to you as the **N:** drive and shown on the desk-top in the Windows environment as **My Documents**. Access the package from the **Start** menu and select **Statistics** and then **SPSS 12 for Windows**. Ignore, by selecting **OK**, any error messages that may appear on the screen until you see Figure 1. Select the option **Don't show this dialog in the future** and this dialogue-box will not appear again. You can then remove the screen by selecting the **Cancel** button.

Figure 1

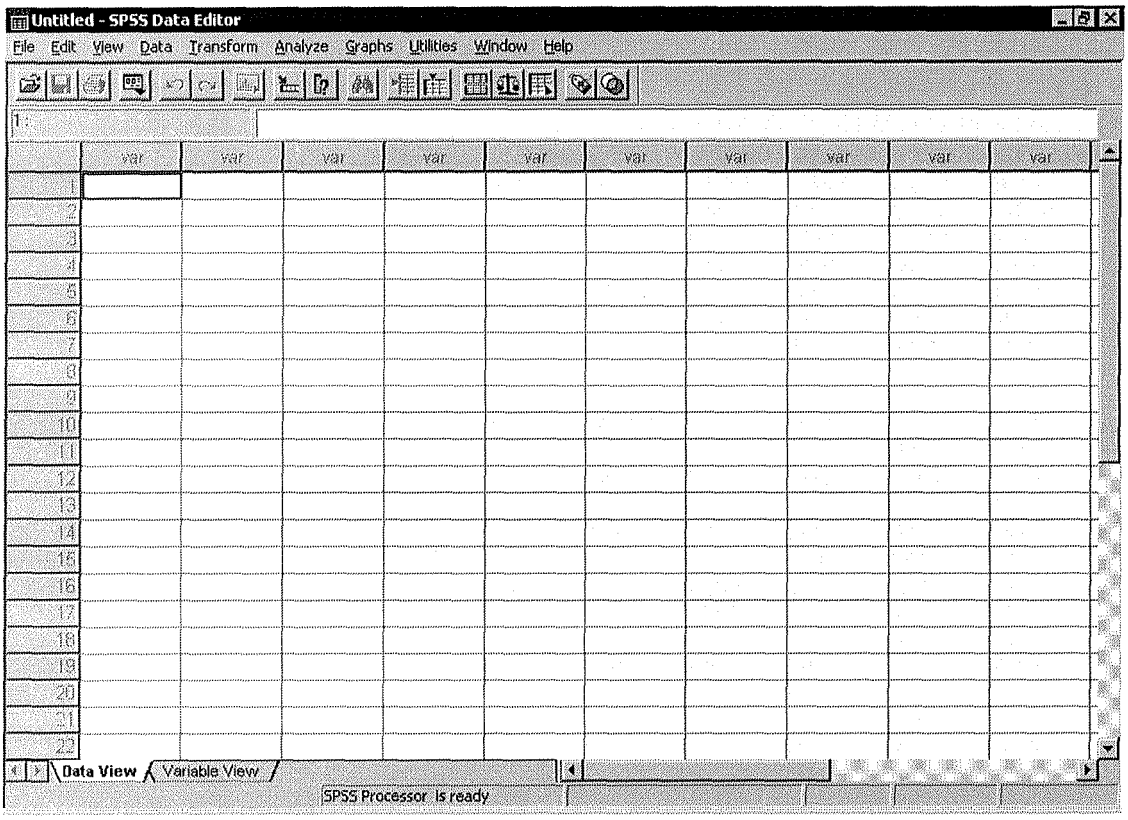


Wait until Figure 2 then appears on the screen. [For information - to exit SPSS, click on the **File** option on the menu bar and choose **Exit** and confirm your intention]

3. Default SPSS Windows

The default image, which should be visible, is shown in Figure 2. The main window, **Untitled - SPSS Data Editor,** with the **Data View** tab highlighted allows numeric data entry direct into the spreadsheet using the default options. If you are not in this view then select the correct **Tab** at the bottom of the screen before proceeding further. The **Variable View** tab will be discussed later in the tutorial for inputting of non-numeric data and other uses. Maximize this window image if you prefer.

Figure 2



4. Notation

In the sections that follow, the actions you must perform are shown in **bold**, as too are the default values chosen by **SPSS 12.** ***Italic*** is used to denote items whose names/values you may choose. It is preferred that variable names start with a letter and are kept as short and simple as possible.

For information:
A sequence of actions using the drop-down menus will be denoted by →. For example, **Data→Insert Variable...** would mean "Click-left on the **Data** option on the drop-down menu and select the option **Insert Variable...** (by clicking-left again)".

5. Data Entry

If **SPSS 12** has been set up correctly the first empty column in the first row will be outlined in the **Untitled - SPSS Data Editor** window. Enter the following three columns of data into the spreadsheet.

11	110	22
15	129	250
104	90	195
168	102	177
60	145	297
125	86	186
111	109	188

By default the names given to these three columns are **var00001, var00002** and **var00003**. Each row of this data represents a ***farm*** number, the number of ***trees*** on that farm and the total yield of ***apples*** (in unspecified units) from each of the farms. It would, therefore, be more useful to have suitable names for each column.

Move to the **Variable View** window (click-left on the tab at the bottom of the window) to see Figure 3 and replace the default names by those suggested above in the **Name** column. Move back to the **Data View** window to see Figure 4. Do not worry about the contents of the other columns, shown in **Variable View**, they will be discussed as the need arises.

Figure 3

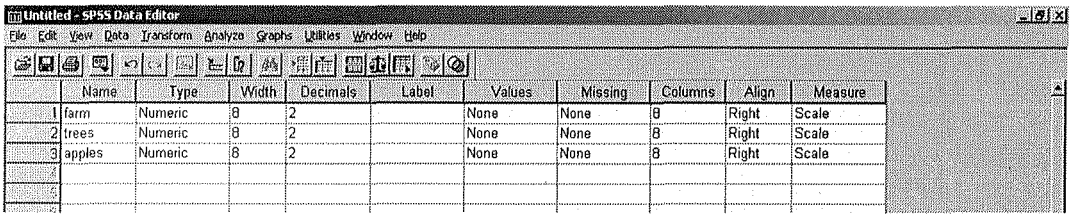
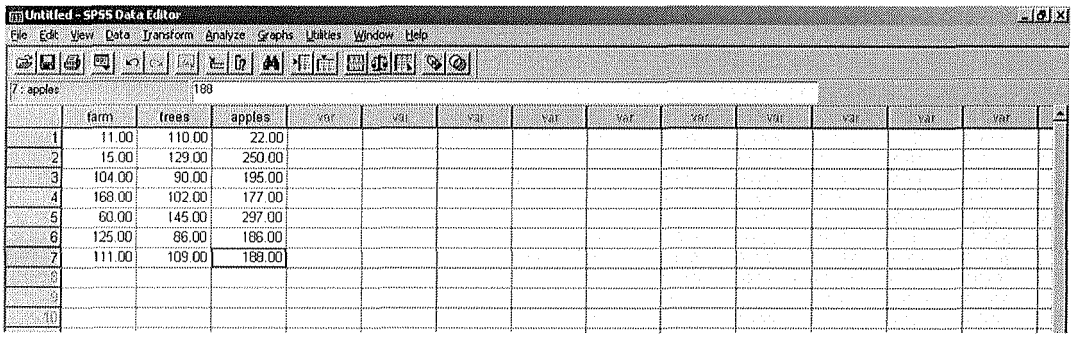


Figure 4



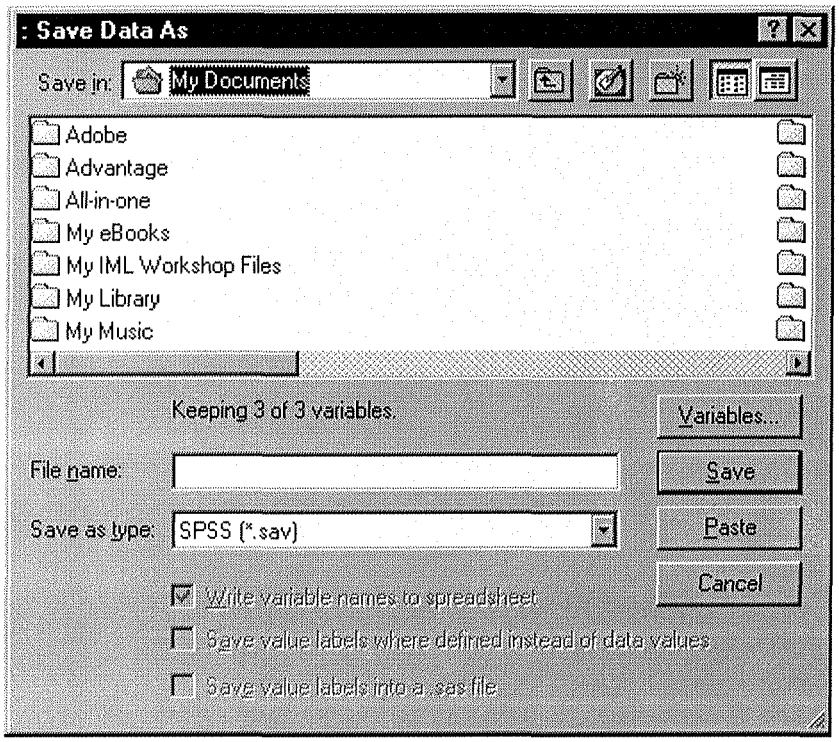
6. Saving Your Work

Although entering the data in Section 5 did not take too long it is always wise to save any work you have done. Saving your work should be a frequent operation whenever you use a PC.

Exercise 1

Several types of files, subsets of variables, or commands can be saved. Here you will be creating a **SPSS Save** file (***.sav***) which will create a file in compressed form containing the complete set of data, names and formats.
Choose **File→Save As...** from the main drop-down menu. The **:Save Data As** dialogue box appears (as shown in Figure 5). If necessary, choose the correct folder (**My Documents\SPSS_Tut**), then type a relevant filename e.g. ***apples*** in the **File name** box. [Note that the default file extension is given as ***.sav***] Click on **Save** and the data will be saved.
The dialogue box will close automatically after the file has been saved.

Figure 5



7. Reading in a SAVED data set

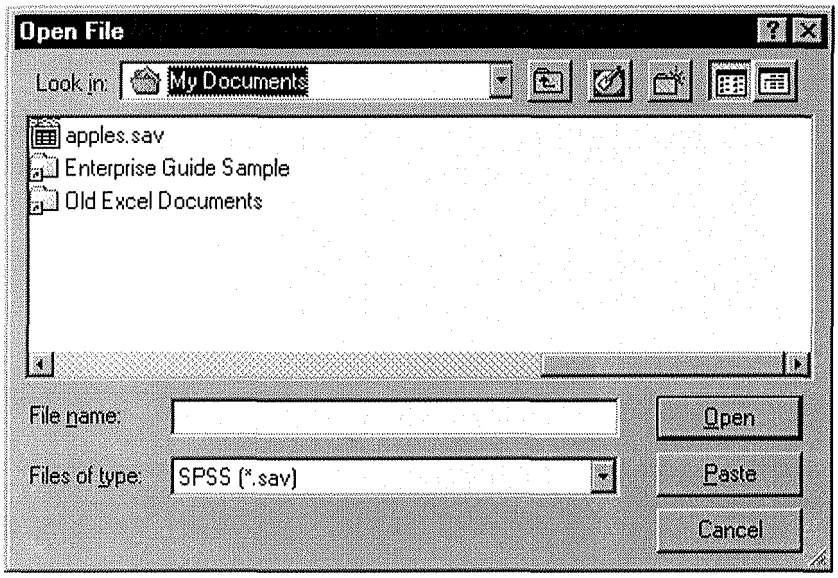
As we are dealing with a very small set of data it is a good time to learn how to read in a previously saved SPSS data set. Follow Exercise 2 and exit from SPSS, check that the file exists and then restart the program and retrieve the saved file you have just created.

If you have **SPSS 12** installed on your personal PC you would be able to open the file **My Documents\SPSS_Tut*apples.sav*** from within the file manager. However, this facility has been removed from the centrally controlled PC labs on campus.

Exercise 2

- Choose **File→Exit** from the main menu. Using any method you know (ie look at the contents of the folder **My Documents\SPSS_Tut** on the desktop) check the file ***apples.sav*** exists. If necessary you may need to repeat the tutorial from the beginning.
- **Invoke the package again and wait for the software to be reloaded.** (described in Section 2)
- Select **File→Open→Data...** from the main menu bar. This will show the **Open File** dialogue box similar to Figure 6. Check you are looking at the correct folder and filename **My Documents\SPSS_Tut*apples.sav*** and select **Open**
- Check that the ***apples.sav* - SPSS Data Viewer** is as shown in Figure 4.

Figure 6



8. Listing the Data

One of the first things that you should always do with the values you have input is confirm that the program will be working on the data values exactly as you intended. You can easily make a mistake by typing **100** in place of **1.00** and this would make your results very strange. As your data is in a spreadsheet you can do this by checking the values straight from the screen. However, you may wish to create a paper listing and take it away with you to check. Follow through the next exercise in order to list and check your data.

Exercise 3

From the main menu bar choose **Analyze→Reports→Case Summaries** to produce the dialogue-box shown in Figure 7.
Highlight the word ***apples*** from the selection on the left-hand side and then click on the  button to select the variable as required to be listed. Repeat for ***farm*** and ***trees***. Click on the **OK** button. The maximised **Output1 - SPSS Viewer** window will become visible. Scroll up and down through it to check that it looks like Figure 8. Note that the data is listed in the order in which you selected the variables e.g. alphabetic. This window could then be printed – do **not** attempt this now.
Correct any mistakes if necessary, and re-save the data file using **File→Save**. This will automatically overwrite the file **My Documents\SPSS_Tut*apples.sav***.

Figure 7

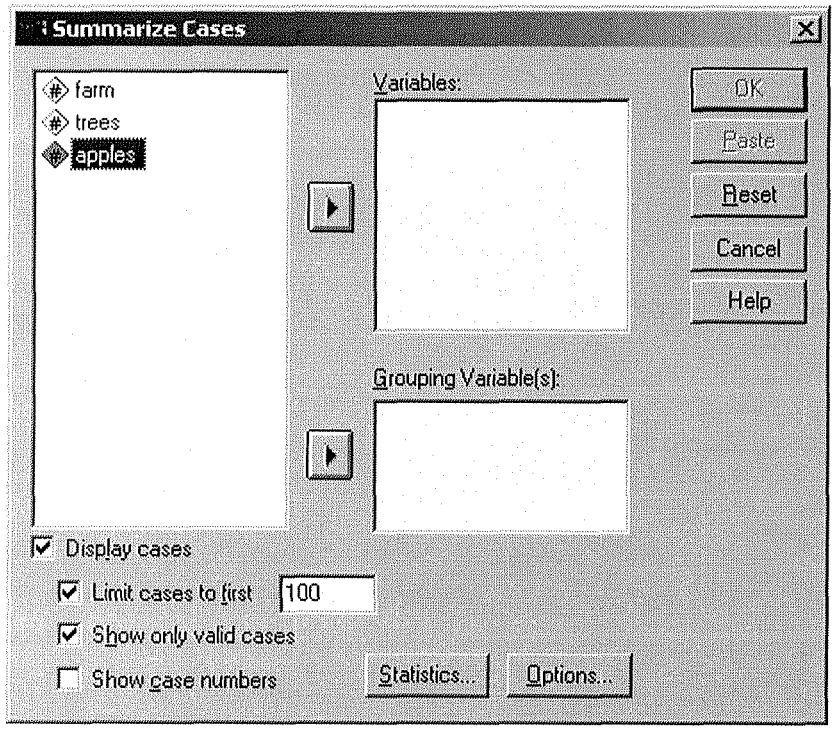
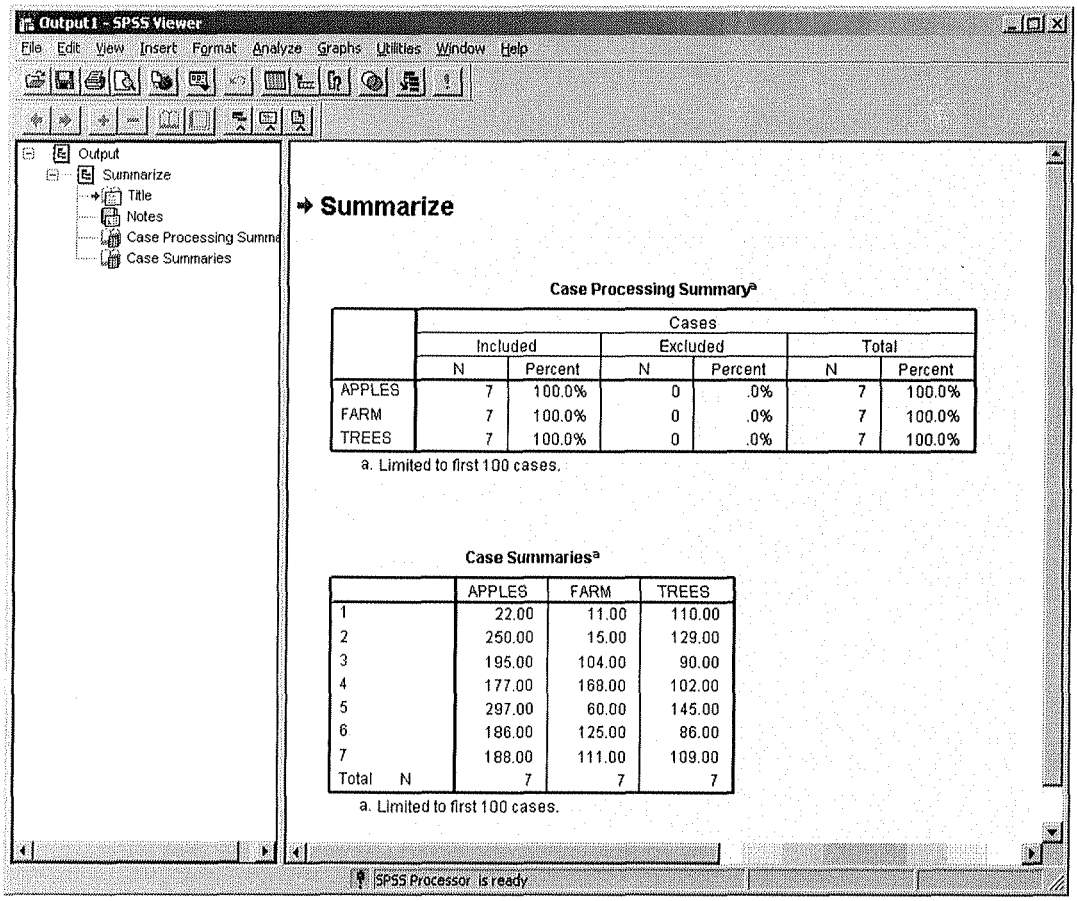


Figure 8



9. Exploring the Data

There are several techniques for helping you to summarise your data. Checking data means or plotting appropriate pairs of variables quickly give you an image of the spread of the data. Both of these checks can be made by following Exercises 4 & 5.

Exercise 4

1. Data Means:
Choose the **Analyze→Descriptive Statistics→Descriptives** menu to get a dialogue-box as shown in to Figure 9 below. Select the variables ***apples*** and ***trees*** in the same way as you did in the previous exercise and then click on **OK**. Check the contents of the **Output1 - SPSS Viewer** window against Figure 10. The summary statistics produced are the default actions of the command but as you will see from the syntax (command language) printed in Figure 33 the underlying command is getting quite complex.

Figure 9

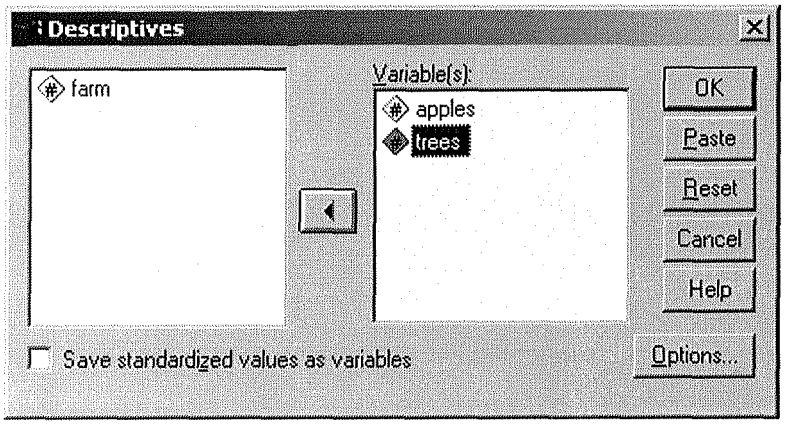
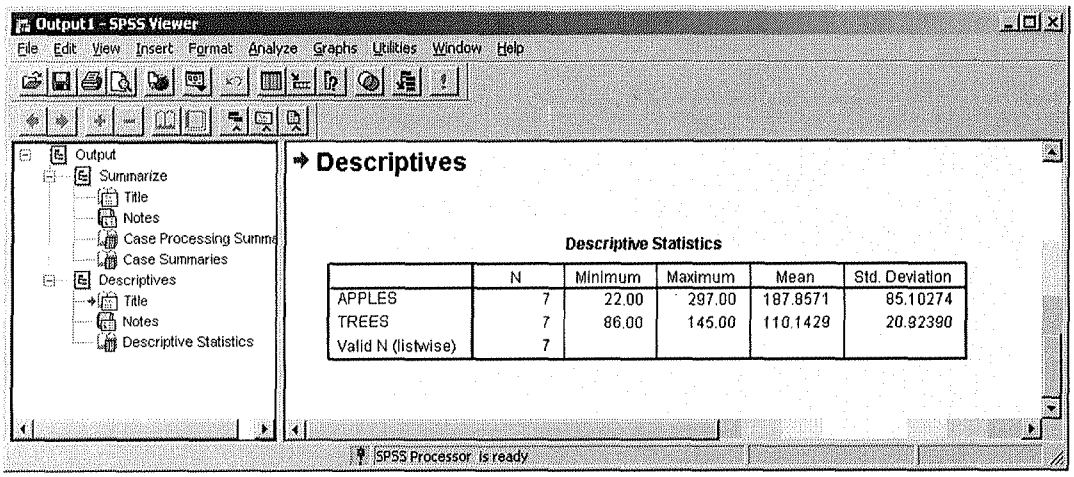


Figure 10



Exercise 5

2. Scatterplots:
From the main menu select **Graphs→Scatter ...** This will produce a further dialogue-box as shown in Figure 11. Select the **Simple** plot and click on the **Define** button to produce Figure 12. Select ***apples*** as the **Y-axis** and ***trees*** as the **X-axis** using the method in Exercise 3 and then click on **OK**. The **Output1 - SPSS Viewer** window now contains this simple scatterplot (as shown in Figure 13).

Figure 11

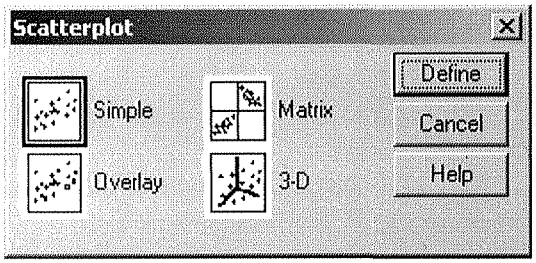


Figure 12

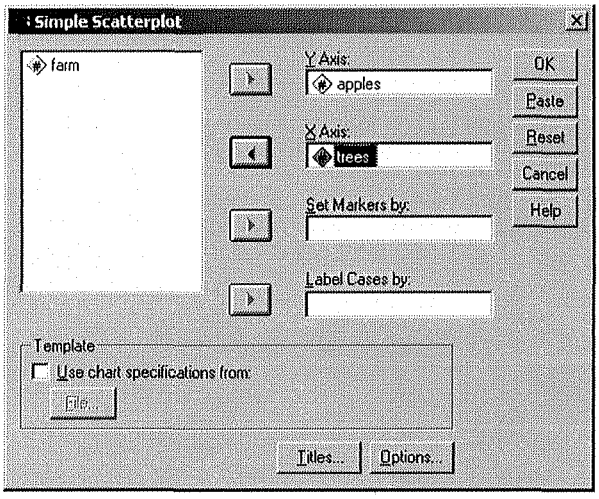
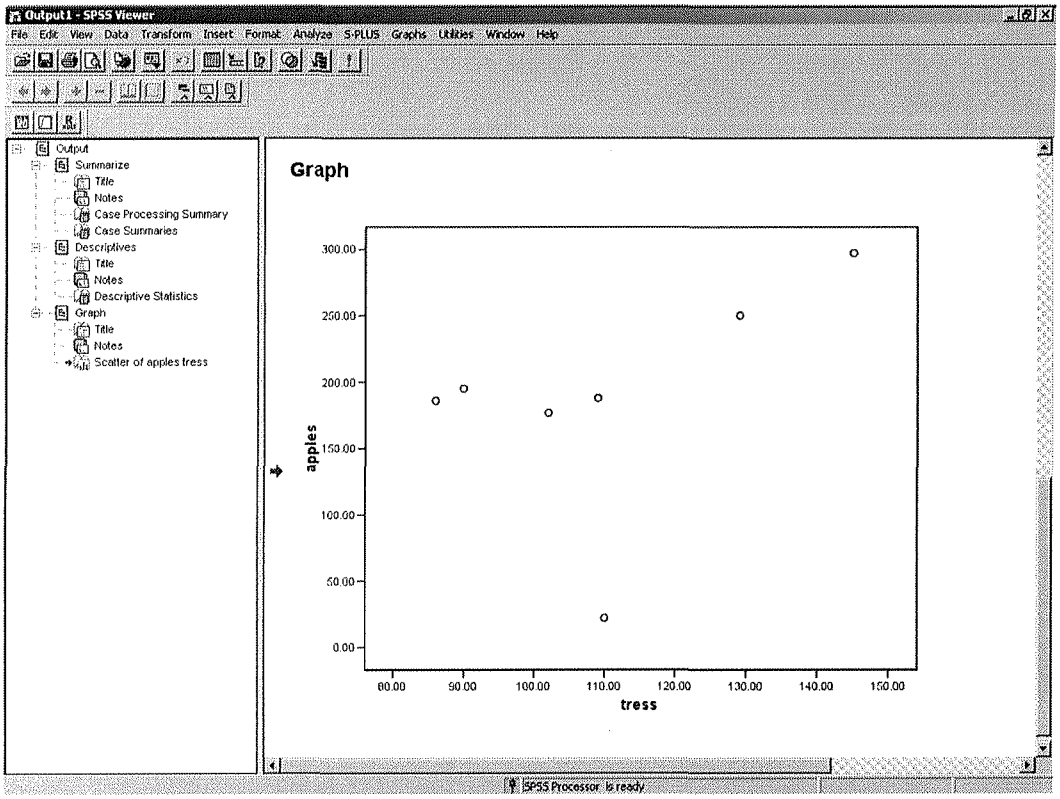


Figure 13



10. Saving Results

It is possible to print the contents of the **Output1 - SPSS Viewer** window directly but now that the facility for draft printing has been withdrawn in the public PC labs this exercise shows how to save your final results to be printed out at a later stage. Try the following exercise to save the contents of the **Output1 - SPSS Viewer** window so that you can print it out in your own department after this tutorial. [Another method of saving results would be to **Cut** and **Paste** the relevant sections into a Word document – you may find this easier when doing your own analysis]

Exercise 6

If necessary, move into the **Output1 - SPSS Viewer** window and click in it. Select **File→Save As** which will give the dialogue-box similar to that shown in Figure 14. Confirm that your folder **My Documents\SPSS_Tut** is selected and then supply a suitable **File name** such as ***apples***. [Note that the default file extension is given as ***.spo***] Click on **Save** and the complete contents of the output window will be saved and the default window name changed to **My Documents\SPSS_Tut*apples.spo***. It is possible to save the graphs separately in a special format that can be imported into a Word Processor, such as Word 2000, however, for most reports this quality is sufficient.

Exit from **SPSS 12** [**File→Exit**] but you do not need to save the contents of any other window.

Figure 14

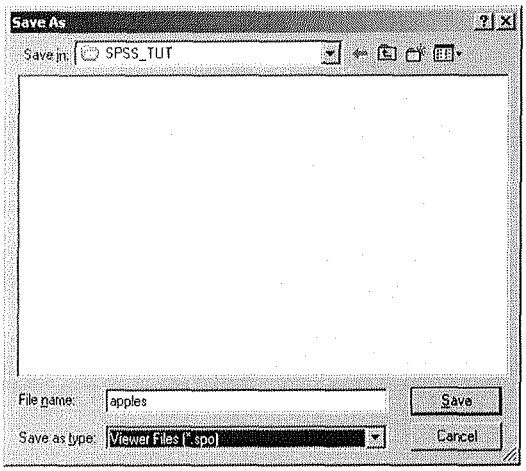
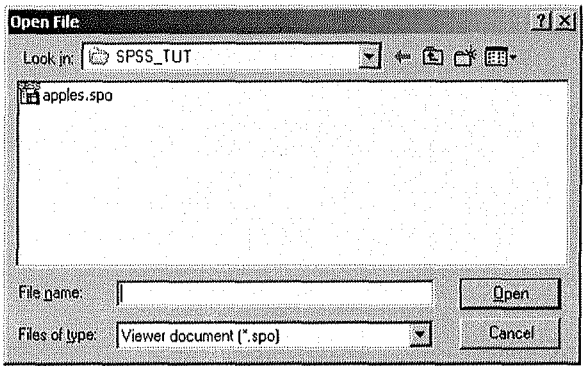


Figure 15



Exercise 7

To print the output file **My Documents\SPSS_Tut*apples.spo***, once saved, you can read it back into **SPSS 12** from the **File→Open→Output...** menu and choose the correct file as shown in Figure 15. Try it now before going on to the next set of data.

Do **not** attempt to print it in the tutorial PC lab.

Exit from **SPSS 12** [**File→Exit**] so that you have a clean session for the start of the next exercise.

From now on the resulting output from exercises and basic dialogue-boxes will not be shown unless further options are selected and additional screens appear. Ask the tutor if you are not sure but be adventurous at this stage. Make mistakes now on a small data set - do not wait until you have entered your data containing 126 variables and 3452 observations!

11. Reading Data From A Separate File

A data set that you wish to analyse may have been entered using a text editor, a database or into another spreadsheet facility. This example assumes that the data, with which you have been supplied, in **My Documents\SPSS_Tut*school.dat*** is in an ASCII (plain text) file created using an editor such as **Notepad**. There are many methods to read (or input) data into **SPSS 12** but for the purposes of this session, it is assumed that data values are separated by one or more spaces but that character values have no spaces between them. This is called **Freefield** data.

Other forms of data entry are not considered in this tutorial. For example, **Excel** spreadsheets can be opened directly into **SPSS 12** and you may prefer to explore this facility in your own time. For help on inputting complicated **formatted** data please come to the Statistical Computing Advisory Service where someone should be able to help you

The four columns of data represent:-

School	name of school
Quest	the number of the question asked – yes/no answers only
Number	the number of responses to that question
Correct	the number of "yes" responses to that question

The data file (without the column/variable names) is shown here:-

School	*Quest*	*Number*	*Correct*
Willink	1	10	7
Willink	2	15	8
Willink	3	12	11
Willink	4	10	9
Willink	5	20	14
Maiden	1	13	8
Maiden	2	12	9
Maiden	3	15	13
Maiden	4	27	22
Maiden	5	23	19
Stonely	1	5	4
Stonely	2	15	12
Stonely	3	7	7
Stonely	4	12	11
Crossley	1	30	25
Crossley	2	30	15
Crossley	3	30	28
Crossley	4	30	27
Crossley	5	30	29

Do **not** type this into the spreadsheet. Follow the next exercise to input the data.

Exercise 8

Begin by invoking **SPSS 12** again, and access the **File→Read Text data...** menu selecting the data set **My Documents\SPSS_Tut*school.dat***. You will need to change the option of **Files of type** to **Data (*.dat)** and choose the file before selecting **Open** to get Figure 16.

Follow the instructions listed here to read in this set of data.

- Confirm that the option **Does your text file match a predefined format?** is set to **No**.
- Click on **Next** button to get Figure 17.
- For the question **How are your variables arranged?** select **Fixed width**
- **Are variable names included at the top of your file?** select **No**
- Click on **Next** button to get Figure 18.
- Leave the options as they are set and click on **Next** button to get Figure 19.
- Scroll down the data shown and you will notice that the last letter of **Crossley** school goes past the end of field marker (↑). Move this marker by clicking on the upright line and dragging it one column to the right. Click on **Next** to get Figure 20.
- Highlight the first column **V1** (as demonstrated in Figure 21) by clicking in the grey box above the column. Under the option **Variable name** give the name ***school*** in the **Variable name** box and confirm **String** within the **Data format** box.
- Repeat this process with the next variable ***quest*** as a **Numeric** variable
- Again with variable ***number*** (also confirm that **Numeric** is in the **Data format** box) and then finally the last variable ***correct*** (again **Numeric**).
- Click on **Next** button to get Figure 22.
- Check that your screen is the same as Figure 22 and then click on **Finish** button.
- The **Variable View** window should contain the information shown in Figure 23 (check that ***school*** is a **String** type and all the others are **Numeric**) and the **Data View** window should contain the 19 observations of the data set.

Figure 16

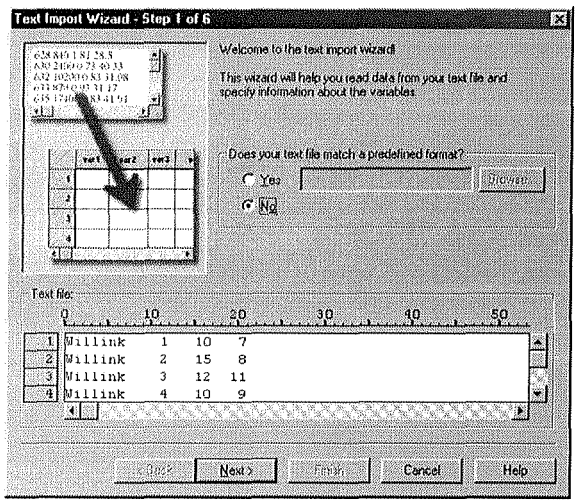


Figure 17

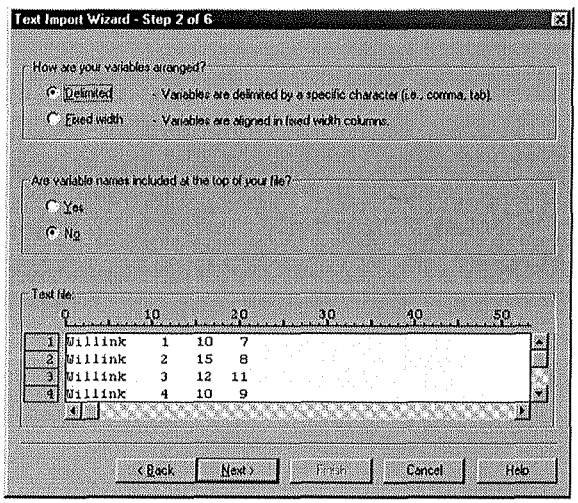


Figure 18

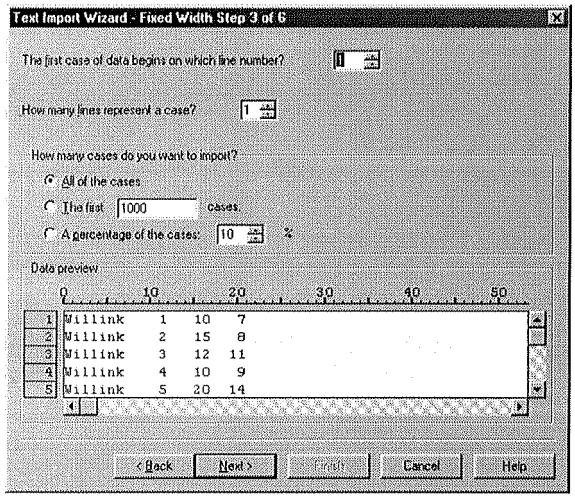


Figure 19

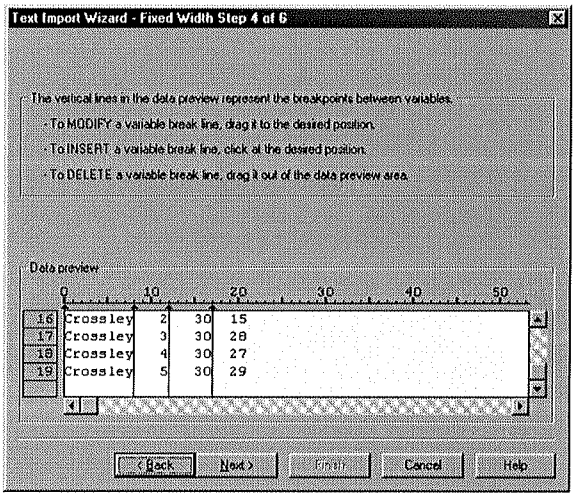


Figure 20

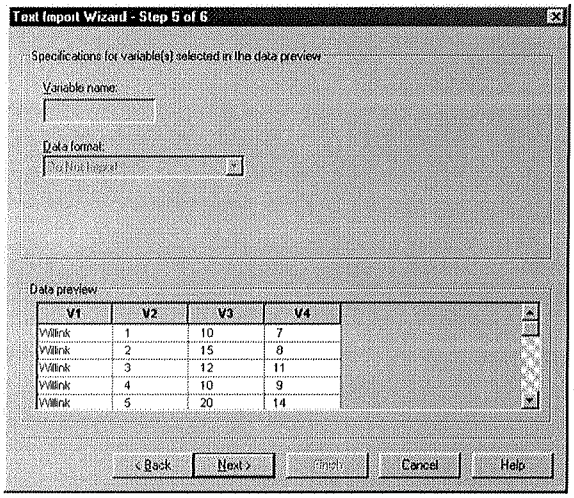


Figure 21

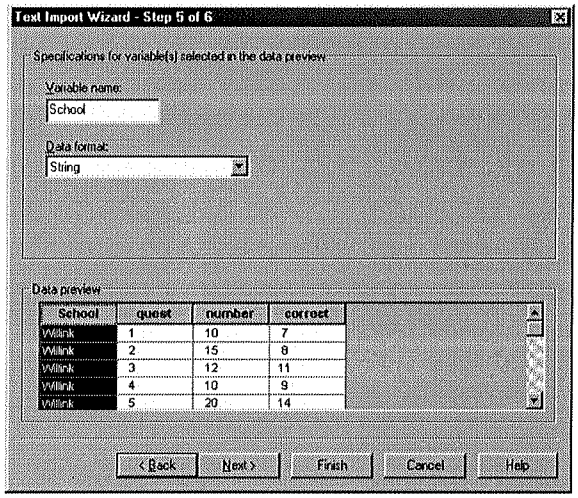


Figure 22

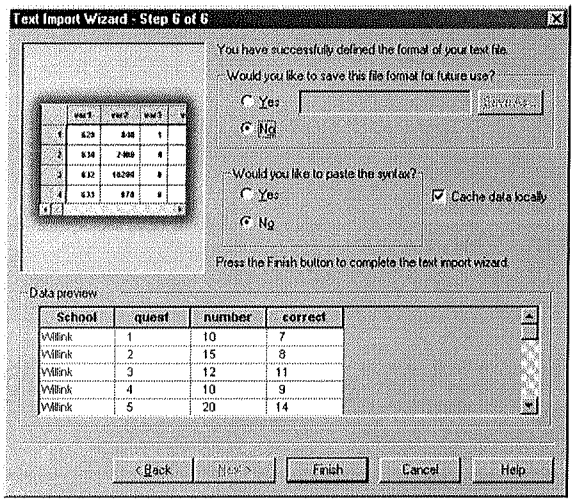
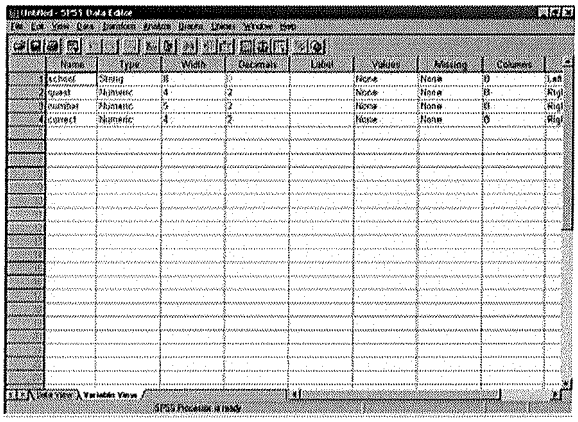


Figure 23



12. Transformations and Calculations on Data

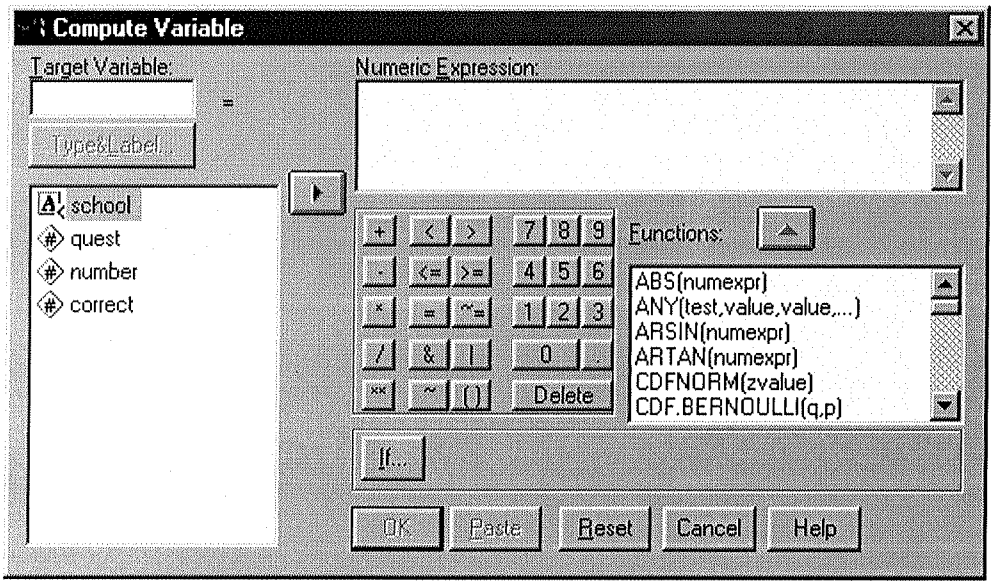
Using the data set just input, the percentage of "yes" answers for each question e.g. $\left(\frac{\text{number of correct answers}}{\text{total number of responses}}\right)*100$ might be required. Exercise 9 demonstrates this facility as an example of manipulation of data values. Many other calculations and transformations are possible but can be explored in your own time.

Exercise 9

Select **Transform→Compute** from the main menu to get the **Compute Variable** box shown in Figure 24. Complete the dialogue-box in the following way:

- Enter the name of the new variable ***percent*** in the **Target Variable** box
- Type in the expression ***(correct/number)*100*** in the **Numeric Expression** box.
- Click on **OK** and return to the **Untitled - SPSS Data Editor** window and check that one or two observations, of the new variable ***percent***, have been calculated correctly.
- Calculate the number of "incorrect" (or not "yes") answers to show you understand how to use this facility.

Figure 24



13. Creating Subgroups of Data

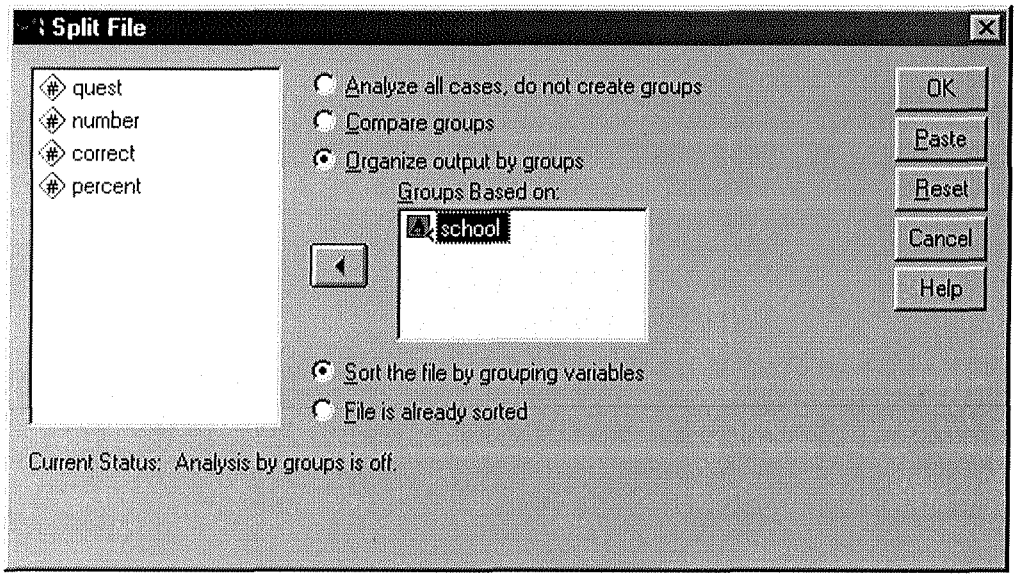
If you want to repeat an analysis for each subset of observations, it is possible to **split** the file into the groups required and then do the analysis automatically on each group. To demonstrate this facility try the following exercise to consider the mean of the percentage data for all schools together and then for each school individually.

Exercise 10

Calculate the overall means of the variables ***number, correct*** and ***percent*** (use Exercise 4 as a guide). Then split the file into the four different schools as follows:

- From the main menu select **Data→Split File** to get Figure 25.
- Select the option **Organise output by groups**.
- Select the variable ***school*** and move it into the **Groups Based on:** box. Your screen should now look like Figure 25.
- Click on **OK** as usual. Notice that the data in the **Untitled - SPSS Data Editor** window has been sorted by ***school*** and that it lists the `CROSSLEY` school first.
- Repeat the calculations with the same variables and see that four separate mean values (one for each ***school***) have been calculated for each variable.

Figure 25



Exercise 11

Be aware that the data file is considered to be split into the groups for any following analysis so complete the next exercise to show how to remove the grouping indicator.
Very simply, follow the previous exercise but this time select the option

Analyze all cases, do not create groups.

Notice that the box entitled **Groups Based on:** has become dimmed. Click on **OK**. You need to be aware of the **status** of the data set that you are analysing so that you always work with the data intended.

Exit from **SPSS 12** [**File→Exit**] but again you do not need to save the contents of any window.

14. Data Summary

If survey data are being analysed, almost the first thing that is required is to display the data in tabular form. This has the added bonus of being a way of checking categorical data. The following exercises demonstrate methods of tabulating data.

A file, **My Documents\SPSS_Tut*surv.dat,*** has been provided for you containing data in the following form:

Variable	**Options**	**Type**	**Suggested Variable Name**
Age		*Numeric*	***age***
Level of Education	N (None) O (O-level) A (A-Level) D (Degree) P (Postgraduate)	*String*	***educ***
Years in Work		*Numeric*	***work***
Work Place		*String*	***Place***
Marital Status	S (Single) M (Married) D (Divorced) W (Widowed)	*String*	***status***
Sex	M (Male) F (Female)	*String*	***Sex***
Number of Children		*Numeric*	***Child***
Smoker	Yes:No	*String*	***Smoke***
Drinks Alcohol	Yes:No	*String*	***Drink***

Exercise 12

Using the instructions in Exercise 8 read the data file **My Documents\SPSS_Tut*surv.dat*** into **SPSS 12**. Modify the variable break (Figure 19) between the last two columns (***smoke*** and ***drink***) as necessary and then remove the vertical line from within the ***drink*** column. The end of row automatically marks the end of the value.
When complete, check the **Untitled - SPSS Data Editor** window for any obvious errors and if necessary, ask the tutor for help. Check the Variable View window against Figure 26 with particular reference to the **Type** information. The value in the **Width** column will depend on whether you have moved a vertical line.
Before proceeding further save the data window (Exercise 1) for future use, as **My Documents\SPSS_Tut*mysurv.sav***.

[If you are having problems at this stage and are struggling with time, you have been supplied with a previously prepared **SPSS** file called **My Documents\SPSS_Tut*surv.sav***. Follow Exercise 2 to bring it into the **Untitled - SPSS Data Editor** window and continue with the next exercise.]

Figure 26

Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	
1	age	Numeric	3	2		None	None	8	Rigl
2	educ	String	3	0		None	None	3	Left
3	work	Numeric	5	2		None	None	8	Rigl
4	place	String	9	0		None	None	9	Left
5	status	String	4	0		None	None	4	Left
6	sex	String	4	0		None	None	4	Left
7	child	Numeric	6	2		None	None	8	Rigl
8	smoke	String	6	0		None	None	6	Left
9	drink	String	6	0		None	None	6	Left
10									

Exercise 13

This exercise demonstrates simple one-way frequency tables. Select **Analyze→Descriptive Statistics→Frequencies...** . Select variable ***sex*** to demonstrate Figure 27 then click on **OK**. Examine the output. Repeat with other variables as you wish, removing the previous selection by highlighting the variable name in the **Variable(s):** box and click on the  button. If you have more time select the **Statistics** and **Charts** buttons and try the options for more descriptive statistics and bar charts.

Exercise 14

Crosstabulations can be produce using a different menu selection. This time choose **Analyze→Descriptive Statistics→Crosstabs...** to produce Figure 28. Select a variable to appear in the row dimension and another to appear in the column dimension e.g. ***smoke*** and ***drink***. Click on **OK** and examine the output again. Can you interpret the results OK? Try adding the **Expected** values and the **Row percentages** to the **Cells** and calculate the **Chi-square** (look at **Statistics**).

Figure 27

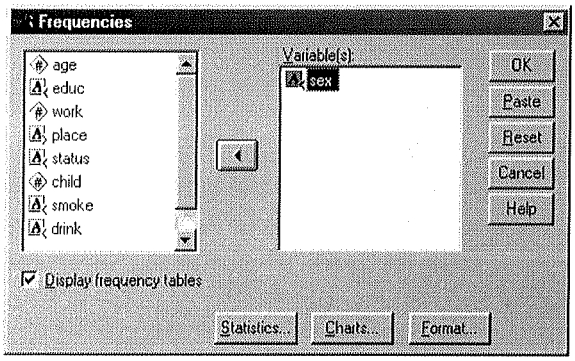
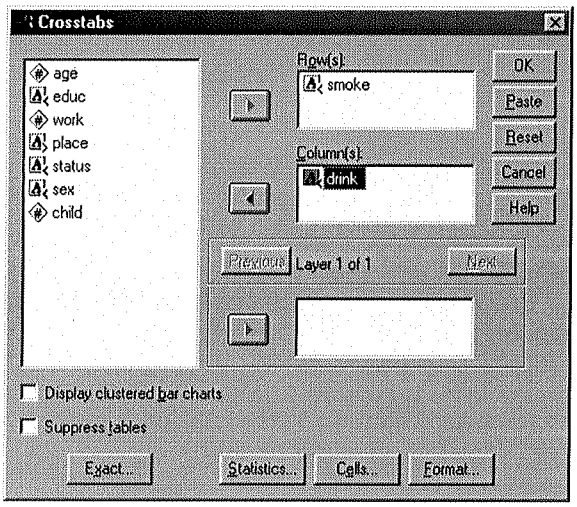


Figure 28



15. Statistical Analysis - Regression

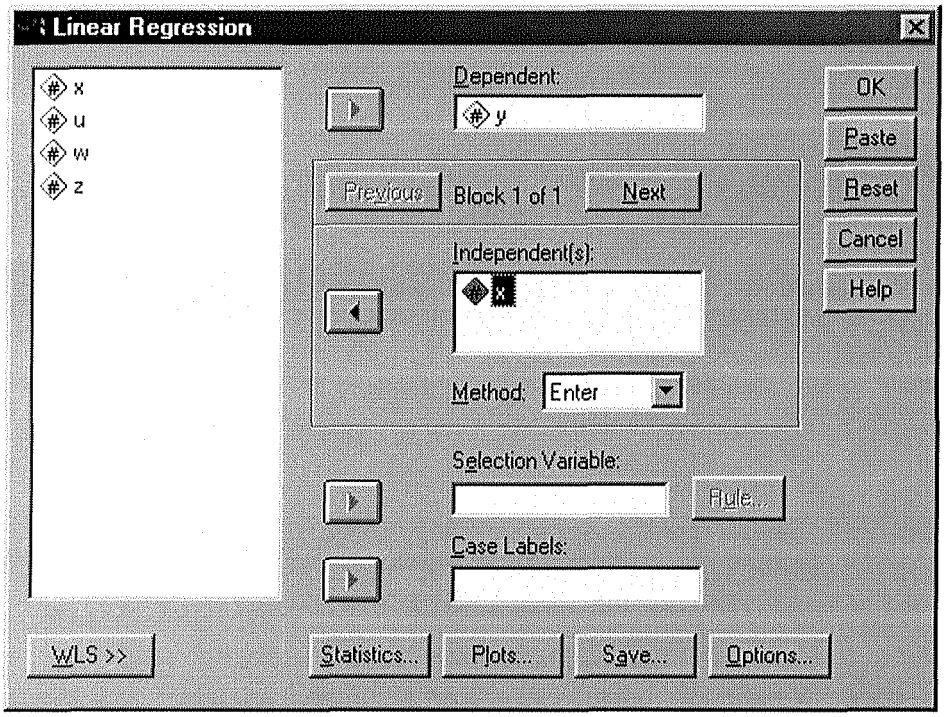
SPSS contains many facilities to perform statistical analyses. All should be used initially under the guidance of a statistician as there are many ways of producing the wrong analysis.

The following exercise uses **LINEAR REGRESSION** to perform simple regressions.

Exercise 15

The **SPSS 12** file **My Documents\SPSS_Tut*reganal.sav*** has been prepared for you so open up this data file into the **Untitled - SPSS Data Editor** window saving the **Output** window if you want to keep a copy of the previous exercises.
The new data set comprises 20 independent observations on variables ***U, W, X, Y*** and ***Z***.
Note: ***U*** represents the regulated ***U***nderfloor heating temperature (-10 to 10 degrees)
W represents the amount of ***W***ater vapour
X represents the time (in minutes) between readings
Y represents the ***Y***ield of the plant
Z represents an unknown chemical compound.
Using the plotting command (see Exercise 5) investigate the relationship between the dependent variable, ***Y*** , and one of the independent variables, ***X***. (To see how the yield of the plant varies with time).
Can you suggest a relationship between ***X*** and ***Y***? Variable ***Y*** looks as if it is increasing linearly with each observation, so a simple linear regression should give the best fit. From the main drop-down menu select **Analyze→Regression→Linear** and choose ***Y*** as the dependent and ***X*** as the independent variables (see Figure 29). Click on **OK**.
Look at the output but do not worry if you are unable to interpret this or any other statistical output. This tutorial is showing you the methods that can be used in SPSS when necessary and does not attempt to teach you statistical techniques.

Figure 29



The following exercise uses **CURVE ESTIMATION** to perform more complex regressions.

Exercise 16

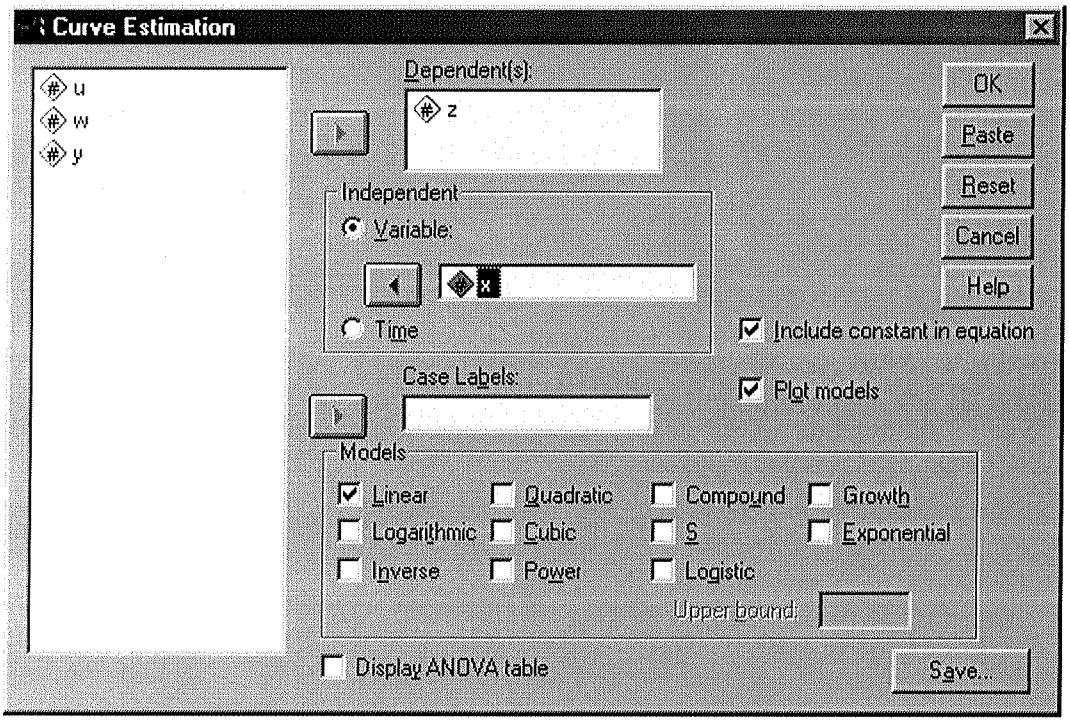
Repeat the plot between another pair of variables ***Z*** and ***X***. By eye, it does not look as if a linear regression would be the best fit between these variables ***Z*** and ***X*** but you can try it anyway. This time you will use a different method.
Select **Analyze→Regression→Curve Estimation...** and choose ***Z*** and ***X*** as appropriate (see Figure 30). Using all the default selections, fit the curve. Notice that a plot is automatically prepared with both the observed data and the fitted linear regression being plotted together. It does not appear to be a good fit, so repeat this part of the exercise selecting the **Quadratic** and then finally the **Cubic** options in the the **Models** selection box. This is also known as "polynomial regression".

Where is the observed data on the cubic plot?

.....

Repeat the previous exercise (Exercise 15) to show a straightforward multiple regression using "unrelated" explanatory variables. Select ***W*** as the dependent variable and ***U*** and ***X*** as the independent variables. It is not possible to show this on a two-dimensional graph.

Figure 30



Exercise 17

When regression analysis has been performed it is usual to want to save the predicted (or fitted) values and residuals and plot them against each other to check that they are randomly scattered about zero. **SPSS 12** allows you to both save and plot at the same time.
Select **Analyze→Regression→Linear...** and choose the simple regression of ***Y*** on ***X***. This time, however, click on the **Save** button and select **Standardised Predicted Values** and **Standardised Residuals**.

Click on **Continue** to return to the previous screen and then click on **Plots**.

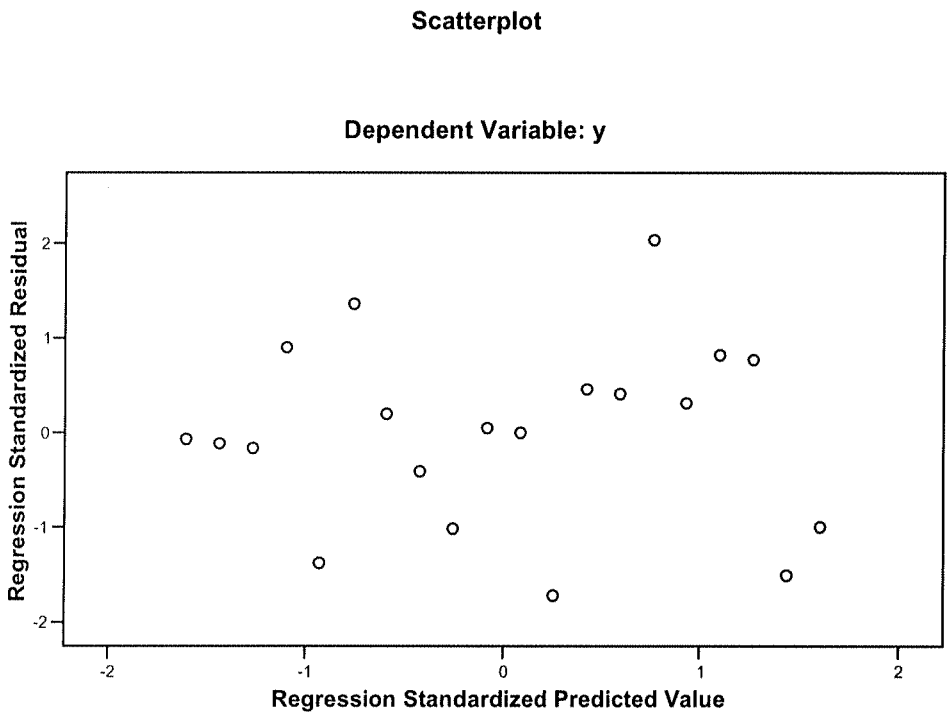
Choose ****ZPRED*** as the X-variable and ****ZRESID*** as the Y-variable. Click on **Continue** and then **OK**. Look at the **Output1 - SPSS Viewer** and **Reganal - SPSS Data Editor** windows to see the results. Did you achieve Figure 31?

What new columns are created?

.....

.....

Figure 31



16. Analysis of Variance

SPSS also has many commands to perform analysis of variance dependent on your data collection. This exercise demonstrates a simple analysis of variance of a balanced designed experiment.

Exercise 18

Another file has been created called **My Documents\SPSS_Tut*balaov.sav***. Open this into the **SPSS Data Editor** window (saving the previous results if required) and look at the data. The file contains information on the ***yield*** of tomatoes from an agricultural experiment which was set out in 3 blocks (***blk***) with 2 treatments (***side*** with 2 levels and ***strain*** with 4 levels). Analyse the data in the following way assuming that ***strain*** is randomised across all plots in each block - a factorial randomised block design.
Select **Analyze→General Linear Model→Univariate** and in the resulting window select ***yield*** as the dependent variable and then ***blk***, ***side*** and ***strain*** as the fixed factors to be defined. This time, rather than run the commands automatically (using **OK**), **Paste** the commands into the **Syntax1 - SPSS Syntax Editor** window, as shown in Figure 32, where you can see the default **Design** sub-command as follows:

DESIGN blk side strain blk*side blk*strain side*strain blk*side*strain .

Now run the commands shown in this window by clicking on **Run→Current** [or use **Ctrl-R**] and look at the results in the **Output1 – SPSS Viewer** window.

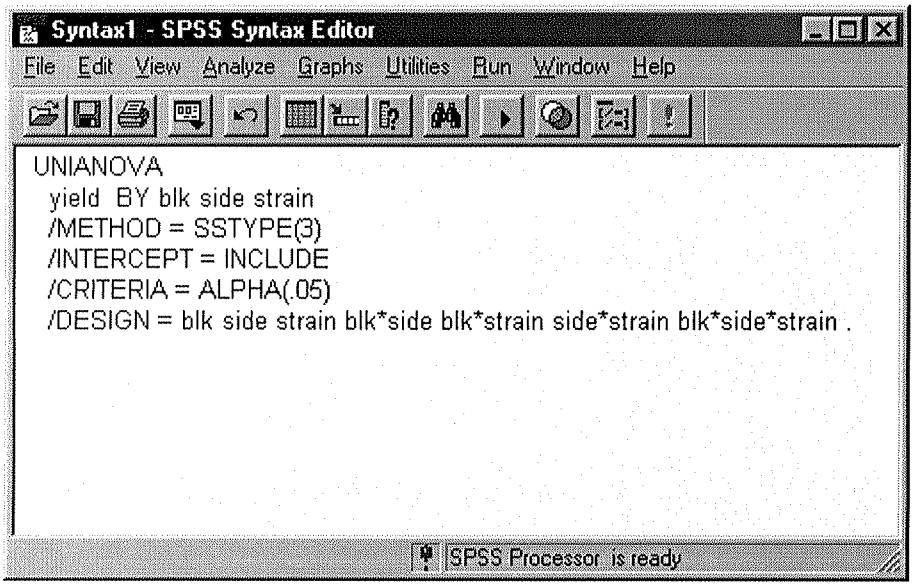
The experiment we are looking at was initially designed as a split-plot, in which each ***strain*** was randomised within each main plot (***side***). To analyse it as a split-plot, return to the **Syntax1 - SPSS Syntax Editor** window and edit the **Design** sub-command to the one below:

DESIGN blk side blk*side strain side*strain .

Re-run the syntax.

In order to display the means for ***side*** and ***strain***, return to the **Univariate** dialogue-box (via **Analyze→General Linear Model→Univariate** menu), select the **Options** button and choose the factors required, and continue as before defining the appropriate (second) design. This time, highlight and run just this **Selection** of the **Syntax1 – SPSS Syntax editor** window.

Figure 32

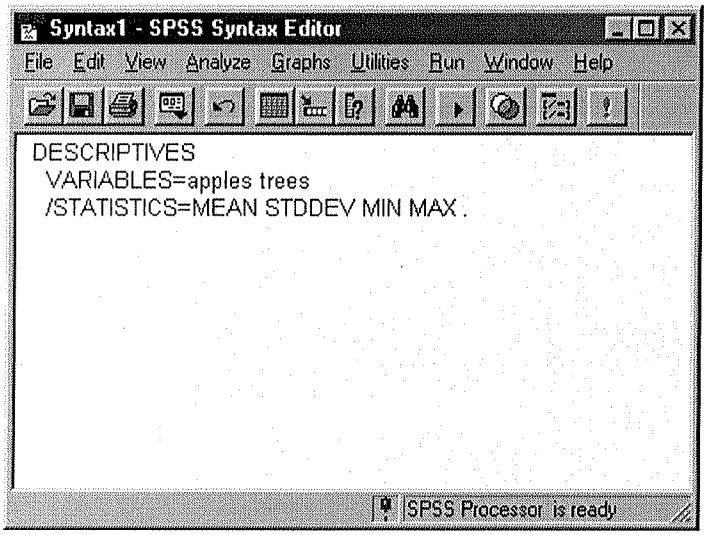


17. Saving Command Syntax

A copy of all of the syntax (**SPSS** command language) used to produce the analysis required is automatically saved in a file called `SPSS.JNL`. However, this file is usually overwritten every time **SPSS** is invoked.

It is also possible to direct certain parts of the syntax to a separate file for future reference, for example, you may want to repeat the exact analysis on a new set of data containing the same variables etc. As shown in the previous exercise there is a **Paste** button available. Having run a particular dialogue-box (e.g. **Descriptive Statistics**) and decided that this would be needed for a subsequent data set you can return to this box and select **Paste**. A new icon will appear on the taskbar which if you open it should contain similar information to that shown in Figure 33.

Figure 33



It is possible to save the contents of the **Syntax1 - SPSS Syntax Editor** window for use at a future date. If you want to save any of the command syntax that you have pasted into the window before exiting then move into the **Syntax1 - SPSS Syntax Editor** window and click on **File→Save As....** Give a suitable name to the file which is given the default extension of *.sps*. This can be reused in a later session opening the file straight into the **Syntax1 - SPSS Syntax Editor** window and using **Run→Current** [or use **Ctrl-R**].

18. Using the HELP menu

All statistical analysis packages include an extensive **Help** system. If you have time, make use of the drop-down menu and look at the **Statistics Coach** facility that is loaded in this PC lab. It may not be available in the public PC labs but is a simple introduction to basic analysis.

When finished use **File→Exit** saving any files you might like to refer to at a later date.

This is the end of the tutorial. Please logout of the machine as normal or you are welcome to try to repeat anything that you have not understood and ask questions of the tutor.

This tutorial should have given you some insight into the basic structure and use of SPSS 12. Any comments or criticisms about this document are welcome by the Statistical Advisory Computing Staff. Please remember we are available, whenever possible, from 9.30 to 12.30 in room G16, School of Applied Statistics, to help you construct the analysis of your data. We would be pleased to see you.