



INTRODUCTION TO GENSTAT 13 FOR WINDOWS



**University of
Reading**

January 2011

CONTENTS

1. GenStat basics	1
2. Data Input and Manipulation	1
2.1 Starting GenStat 13th Edition.....	1
2.2 Data input	3
2.3 Some basic data manipulation	11
2.4 Factor Columns.....	19
2.5 Understanding how GenStat works	25
3. Simple statistical inference	29
3.1 The use of boxplots.....	29
3.2 Comparisons of means	31
References.....	32
4. Simple linear regression	33
4.1 Setting up the data	33
4.2 Correlation and regression.....	35
4.3 A GenStat tutorial	37

1. GenStat basics

The aim of this introductory tutorial is for the user to become familiar with the basic operations of GenStat for Windows.

In this guide, we sometimes assume a user already has experience of Excel. We show how data entered into Excel can be analysed with GenStat and also how data from GenStat can be saved as an Excel file.

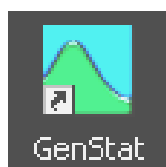
Users who are not familiar with Excel should omit these sections. Experience with Excel is not necessary for using GenStat.

2. Data Input and Manipulation

2.1 Starting GenStat 13th Edition

You start GenStat within Windows on a PC by clicking on the GenStat icon on the desktop or toolbar, or by selecting “GenStat”, from the Programs Menu. If no GenStat icon is available on the desktop, you can create one yourself by using Windows Explorer to select the subfolder Gen13Ed\bin (By default, GenStat is installed in the folder C:\Program Files\Gen13ed). Right click with the mouse on GenStat.exe and create a shortcut. This shortcut can now be dragged onto the desktop. You might rename the icon on the desktop (right click and click ‘rename’) as GenStat 13th Edition, to avoid confusion with previous versions.

If you are using a University of Reading networked machine on campus (for example in one of the computer laboratories or in the main library) you should instead click the start button located in the bottom left corner of the screen and then click ‘all programs’, and select the menu sequence **All Programs** → **GenStat** → **GenStat 13th Edition**. With this menu sequence highlighted right-click with the mouse and select ‘create shortcut’ from the list of options.



After starting GenStat, you see a standard Windows interface with a title bar, menu bar, tool bar, status bar and several windows, **Fig. 2.1a**. The Output window will contain the output from the operations we perform. The input log keeps a record of what has been done in an analysis. Many of the menus are standard for Windows applications. Only **Run**, **Data**, **Spread**, **Graphics** and **Stats** are GenStat-specific.

Fig. 2.1a GenStat Windows

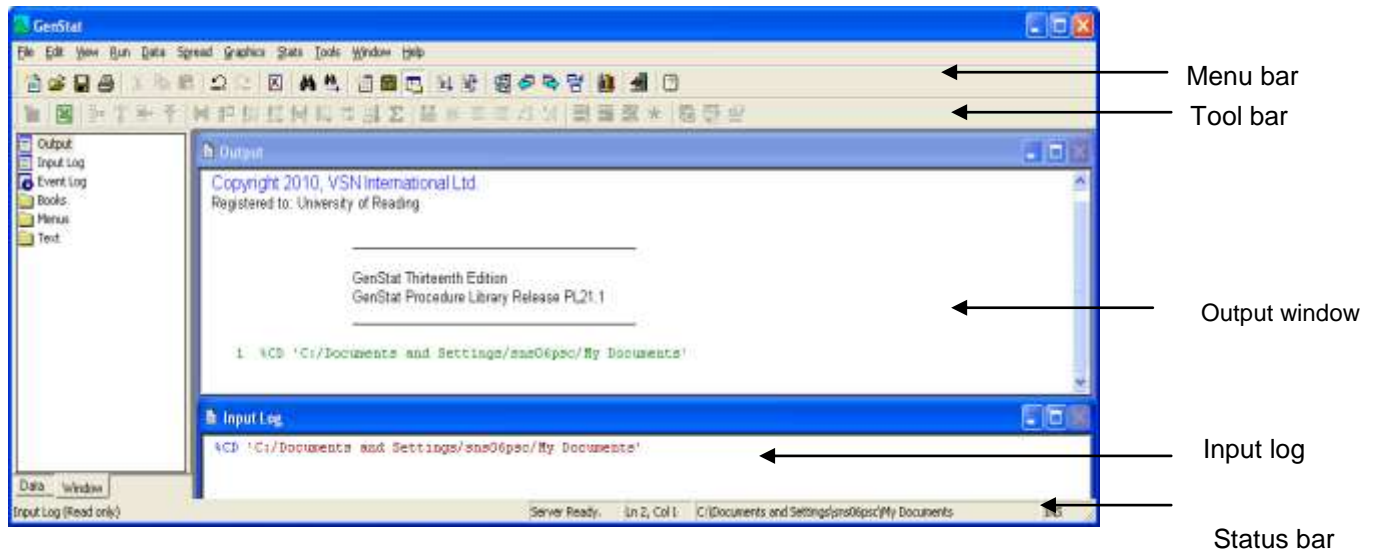
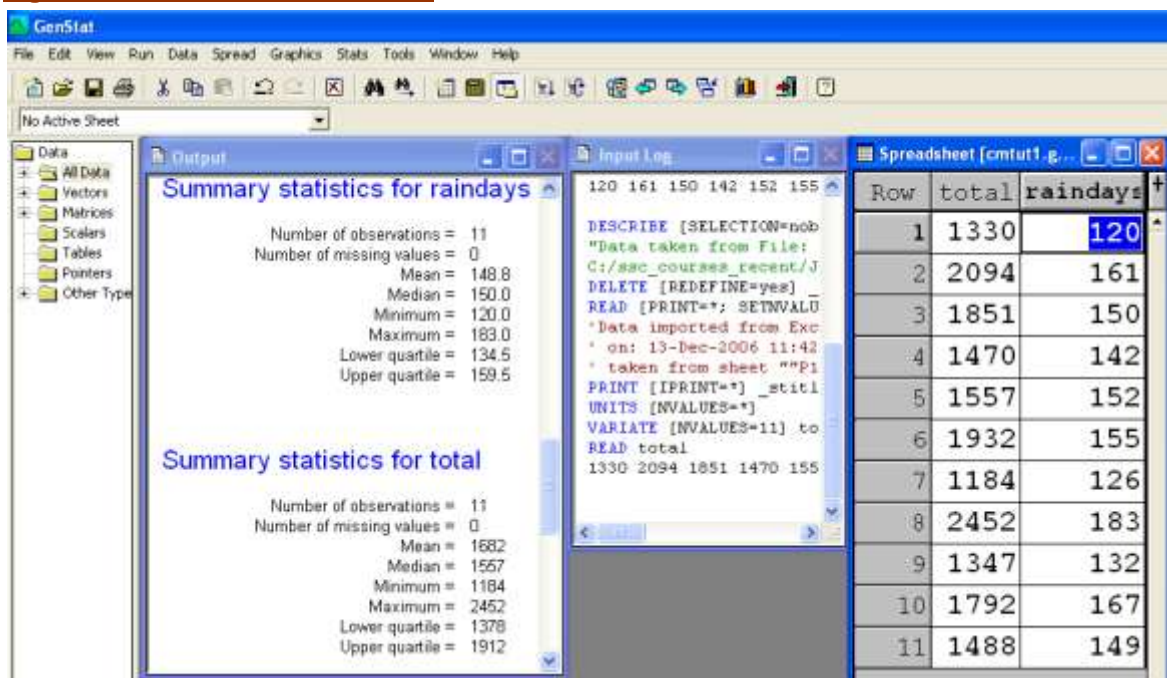


Fig. 2.1b shows an example of the interface after a spreadsheet has been opened and summary statistics have been requested.

Fig. 2.1b Windows ⇒ Tile Vertical



2.2 Data input

2.2.1 Data input using the Spread Menu

Close the startup window, if you haven't done it already.

- ⇒ We show two ways to enter data into GenStat. The first is within GenStat. Choose **Spread** ⇒ **New** ⇒ **Create**. (see *Fig. 2.2a*)
- ⇒ Choosing **Create** brings up a box allowing you to specify how many data columns you want, and how many rows of data there will be. Edit the box to make a GenStat spreadsheet with 11 rows and 2 columns as shown in *Fig. 2.2b*.

Fig. 2.2a Spread ⇒ New ⇒ Create

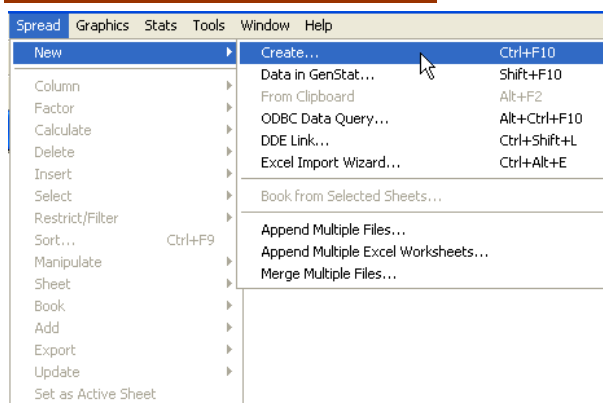


Fig. 2.2b Specify the size



Different types of spreadsheet can be made, but the default (i.e. what GenStat will select in the absence of further information) - **Vector** - is usually the type you will need.

- ⇒ Click **[OK]**, and an empty spreadsheet will appear. You can start to enter data by clicking in a cell in the spreadsheet. Type the number, and then press the **[Enter]** key. Enter the following numbers into the first column:

1330 2094 1851 1470 1557 1932 1184 2452 1347 1792 1488

- ⇒ Press the **[Enter]** key after the last number. The cursor will then move to the top of the next column. Enter these numbers into the second column:

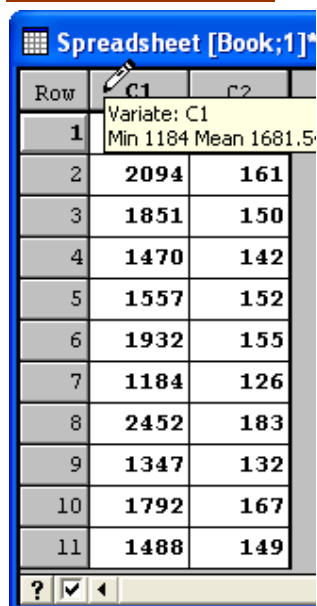
120 161 150 142 152 155 126 183 132 167 149

Make sure that you press the **[Enter]** key after typing the final number. The resulting sheet is shown in *Fig. 2.2c*. If you have made any mistakes, these can be easily corrected, using the arrow keys to move to the cell to amend and entering the correct value.

For each row, the value in the first column is the annual rainfall total, and the value in the second column is the number rain days. It is helpful to give the columns more meaningful names than the default C1, C2, etc.

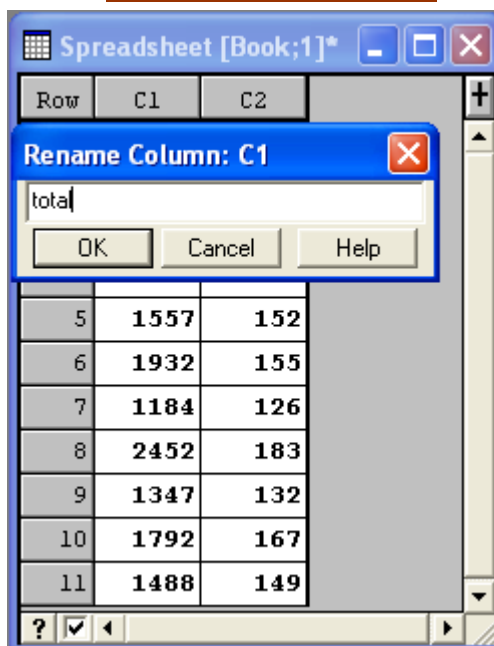
- ⇒ To give a name, position the cursor as shown *Fig. 2.2c*. It becomes a pencil, rather than a hand, and clicking on the mouse gives a popup screen where you can type the name for the column, as shown in *Fig. 2.2d*. Then press **[OK]**.

Fig. 2.2c Naming C1



Row	C1	C2
1		
2	2094	161
3	1851	150
4	1470	142
5	1557	152
6	1932	155
7	1184	126
8	2452	183
9	1347	132
10	1792	167
11	1488	149

Fig. 2.2d Giving the name



Row	C1	C2
5	1557	152
6	1932	155
7	1184	126
8	2452	183
9	1347	132
10	1792	167
11	1488	149

⇒ Once you have given column C1 the name *total*, repeat with C2 with the name *raindays*. These names now appear on the columns of the spreadsheet.

2.2.2 Organising the Windows

It is useful to decide how you wish to use the different windows in GenStat.

⇒ Use **Window** ⇒ **Tile Vertically** to give the layout with the three Windows namely the Output, the Input Log and the Spreadsheet. This is roughly as shown earlier in **Fig. 2.1b**.

These windows indicate one difference between most statistics packages, like GenStat, and spreadsheets, like Excel. With a spreadsheet you have effectively one type of window within which you can have your data and results.

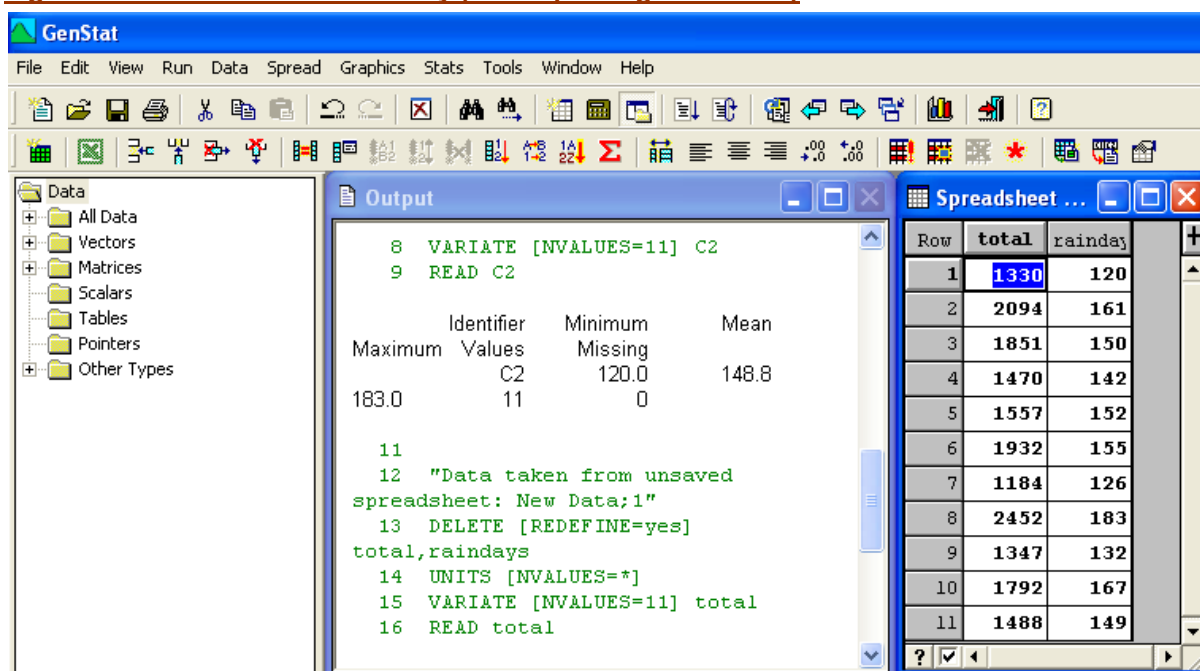
In GenStat you have one window for your data and this is called the spreadsheet. It does not include any results.

You have a separate window, called output, for the results.

You also have here a third window called the Input Log, as shown earlier in **Fig. 2.1b**. This keeps a record of what you have done.

⇒ Now minimize the Input Log and then use **Window** ⇒ **Tile Vertically** again (or press <Shift><F4>) to give the layout roughly as in **Fig. 2.2e**.

Fig. 2.2e Windows ⇒ Tile vertically (with Input Log minimised)



⇒ Now try maximizing the output window, and then reducing it to its half size again.

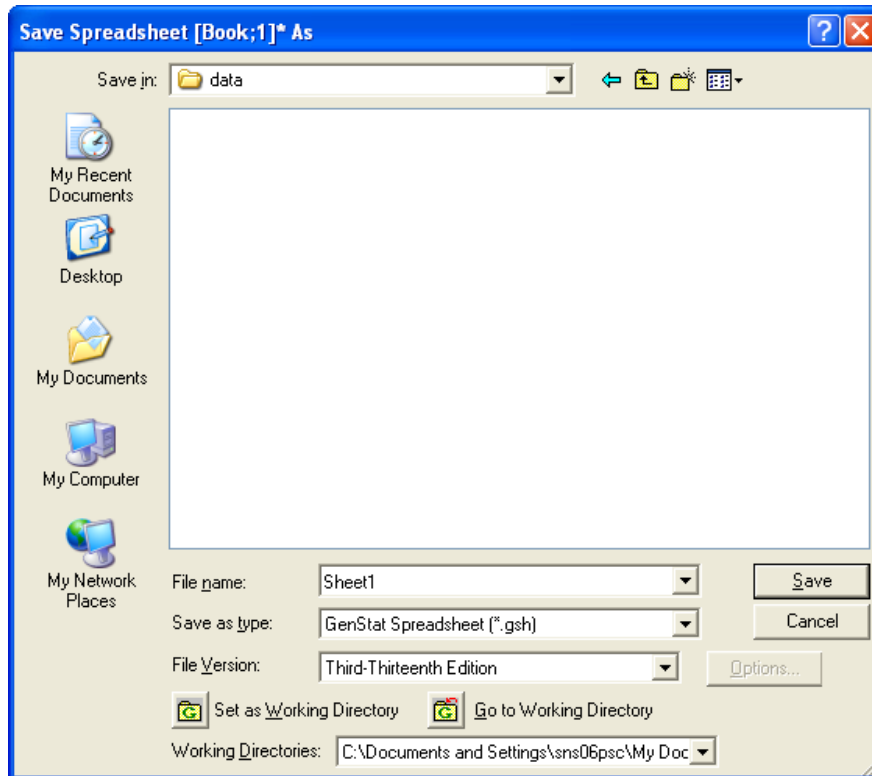
⇒ Then use **Window** ⇒ **Tile Horizontally**.

Which layout of the windows do you prefer?

2.2.3 Saving the file

- ⇒ With the spreadsheet window active, use **File** ⇒ **Save As** and save the file as **cmtut1.gsh**, as shown in **Fig. 2.2f**.

Fig. 2.2f File ⇒ Save As



- ⇒ Use **Run** ⇒ **Restart Session**, so you are ready to try the second way of entering data. It will warn you, as shown in **Fig. 2.2h**, but persevere by clicking on 'Yes'.

Fig.2.2g Run ⇒ Restart

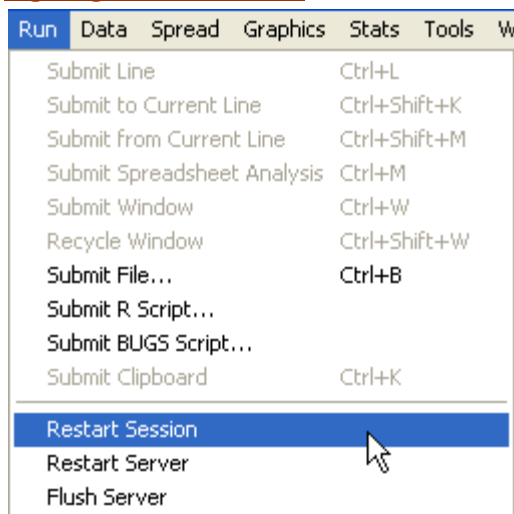
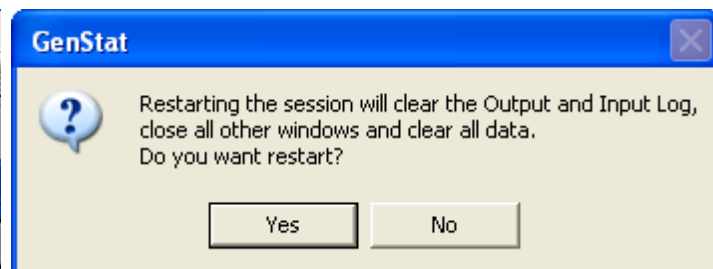


Fig. 2.2h Accepting the restart



If you are not experienced in computing, or if you are not familiar with Excel, then go to Section 2.3.

2.2.4 Data input from Excel worksheets

This section assumes you are familiar with Excel. If not, or if you are using different spreadsheet software, then omit this section and go to Section 2.3.

Most of your data is probably entered already, in a database or in a spreadsheet like Excel. Importing data into GenStat is easy.

⇒ Minimize GenStat and go into Excel.

We assume you are now in Excel.

⇒ Create a new Excel workbook and enter the same data as earlier, see **Fig. 2.2i**. In the cells above the data, you can enter the names for the columns : 'total' and 'raindays'.

Fig. 2.2i Data entry in Excel

	A	B
1	total	raindays
2	1330	120
3	2094	161
4	1851	150
5	1470	142
6	1557	152
7	1932	155
8	1184	126
9	2452	183
10	1347	132
11	1792	167
12	1488	149
13		

⇒ Save your Excel workbook and give it the name **cmtut1.xls**. You have now finished with Excel, so minimize Excel and go back to GenStat.

⇒ In GenStat, choose **File** ⇒ **Open** and select the Input file. Indicate that the file to import is of the '**Other Spreadsheet Files**' type as shown in **Fig. 2.2j**.

Fig. 2.2j Look for Excel file

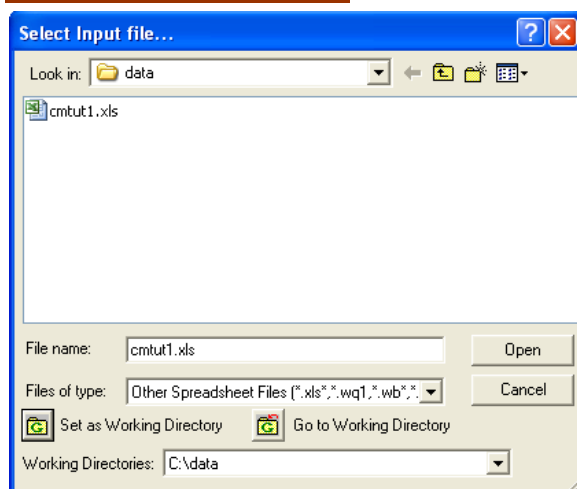
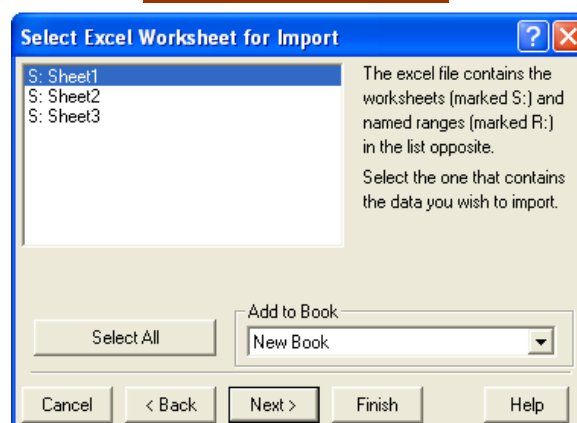


Fig. 2.2k Use first sheet



- ⇒ In the next window (**Fig. 2.2k**), you can select which worksheet of the workbook you want to import. In this case Sheet 1 is what you want, so just click **Next**, then **Finish**, to import the data into a GenStat spreadsheet.

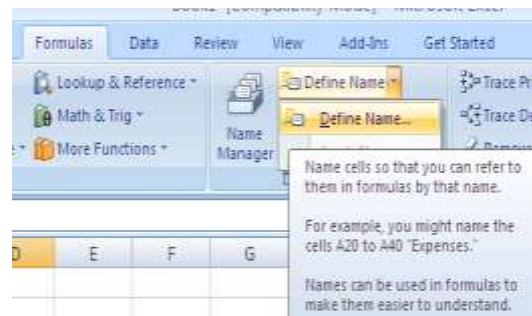
In this example the data were easy to import, because the Excel sheet only included what was to be imported. To import any set of data equally easily, into GenStat, from Excel, you can define a named range in Excel.

- ⇒ Go back into Excel and add a line or two of description as shown in **Fig. 2.2l**. Then, in Excel, highlight the range containing the data and the header row and choose **Formulas** ⇒ **Define Name**, see **Fig. 2.2m**.

Fig. 2.2l Excel file with description

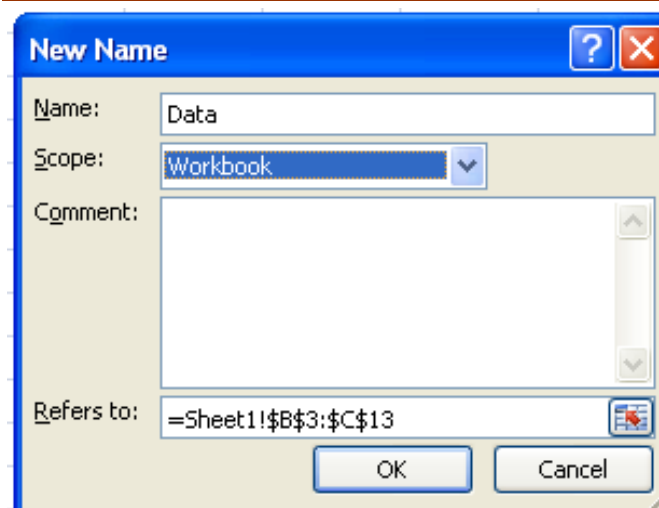
	A	B	C
1	Annual rainfall totals for Genstat tutorial		
2		Data in mm	Number of days with rainfall exceeding 0.85mm
3		1330	120
4		2094	161
5		1851	150
6		1470	142
7		1557	152
8		1932	155
9		1184	126
10		2452	183
11		1347	132
12		1792	167
13		1488	149

Fig. 2.2m Defining a name in Excel



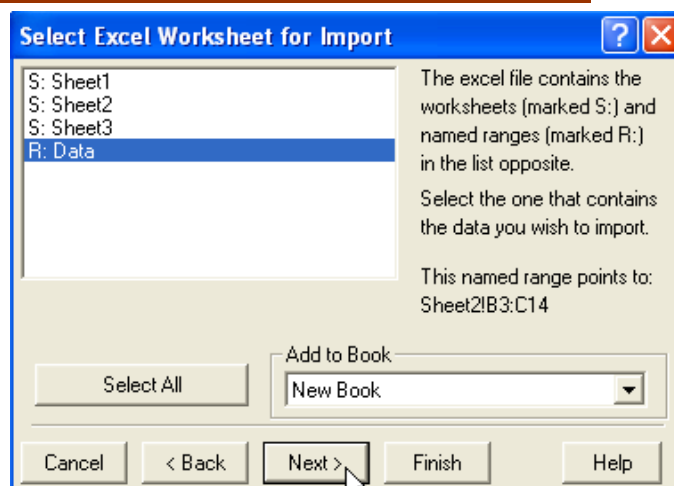
- ⇒ Give the range a name, for instance *Data*, as shown in **Fig. 2.2n**. Then save the Excel file and minimize Excel.

Fig. 2.2n Specifying the name of the range as “Data”



- ⇒ Go back to GenStat and restart the session by selecting **Run** ⇒ **Restart Session** and then clicking **Yes**, to clear all windows, dialogue boxes and the spreadsheet. When you now reopen the file *cmtut1.xls*, you are able to select the range *Data* as shown in **Fig. 2.2o**. The R:Data in **Fig. 2.2o** signifies that you are using a named range. Click **Next**, then **Finish**, to import the data into a GenStat spreadsheet.

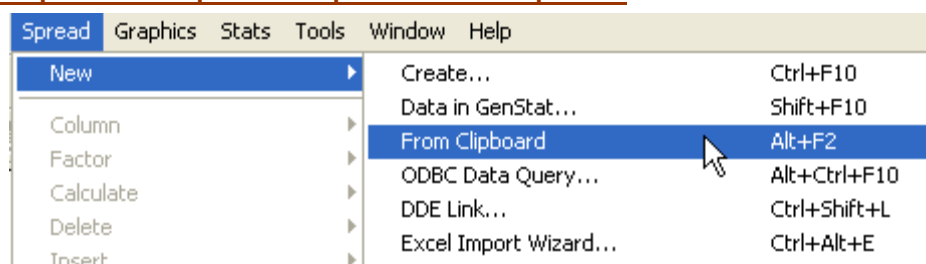
Fig. 2.2o Importing the named range into GenStat



An alternative way of transferring the data is to copy a range of cells from Excel and paste it into GenStat. This is not considered good practice in data management, but is a fast and easy way of data transfer for a quick provisional analyses.

- ⇒ To show this way, choose **Run** ⇒ **Restart Session** to clear all data out of GenStat. Go back into Excel. Highlight the range containing the data and column headers and choose **Edit** ⇒ **Copy**, or right click with the mouse in this range and click **Copy**. Now the data are loaded into the Windows clipboard. Go back to GenStat and choose **Spread** ⇒ **New** ⇒ **From Clipboard**, (see **Fig. 2.2p**), click **OK** then **Yes** and the data are pasted into a GenStat spreadsheet.

Fig.2.2p GenStat option to import from the clipboard



2.2.5 Advanced data input

It is also possible to import data from other file formats or to create links with other files. More information can be found in the full version of this guide.

2.2.6 Leaving GenStat

- ⇒ To end a GenStat session, choose **File** ⇒ **Exit**. You will be asked if you want to save any of the open windows or spreadsheets. Select **[Yes]** to save the spreadsheet, but **[No]** for the other windows, and exit GenStat.

As well as showing you how to enter data into GenStat, you have seen how easy it is to transfer data from another package, such as Excel. So, if you are already familiar with a spreadsheet or another statistical package, using GenStat does not have to stop you from using other software. You can use GenStat in addition. We will show examples from Excel spreadsheets at various points in this guide.

2.3 Some basic data manipulation

2.3.1 Summary statistics

- ⇒ Restart the session and reopen the file *cmtut1.gsh*. The data in the spreadsheet are passed into the GenStat server as soon as you click anywhere outside the spreadsheet or the spread menu.
- ⇒ Try doing this by clicking in the output window.

Some summary information about the two columns *total* and *raindays* will appear in the output window showing minimum, mean and maximum values, number of values and number of those that are missing. What are the values of these summaries for the two variates *total* and *raindays*?

- ⇒ For further statistical summaries use the **Stats** menu, as shown below. Choose **Stats** ⇒ **Summary Statistics** ⇒ **Summarize Contents of Variates**. Select the variates required in the resulting dialogue shown in **Fig. 2.3b**, and then click[Run].

Fig. 2.3a Choosing the dialogue

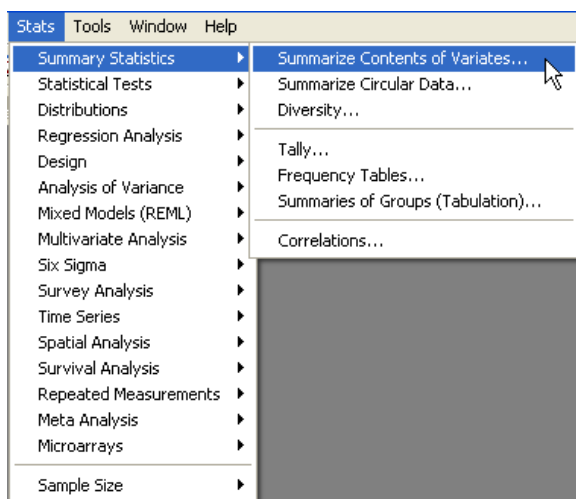
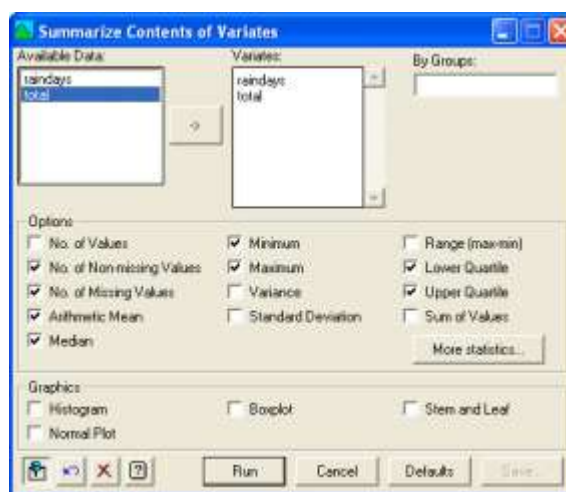


Fig 2.3b Selecting the columns





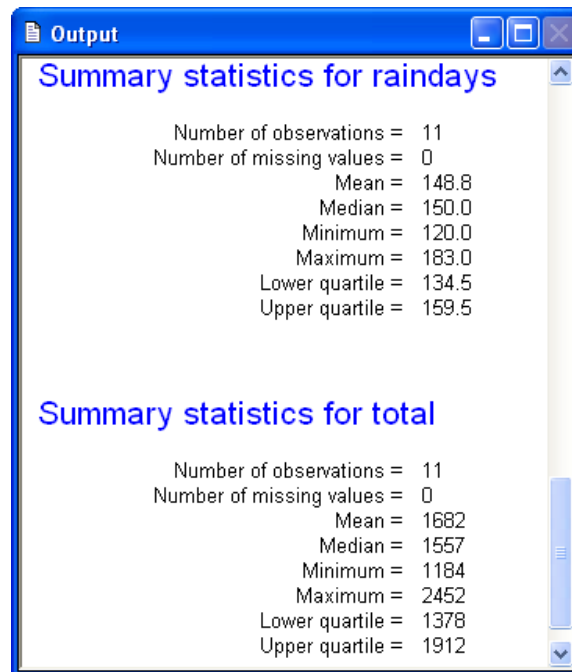
- ⇒ Select the Output Window. If you cannot see this window, try clicking the  or  buttons in the tool bar successively until it appears. Some of the results are shown in **Fig. 2.3c**. There are other statistics available within the same dialogue box.

Fig. 2.3c The results


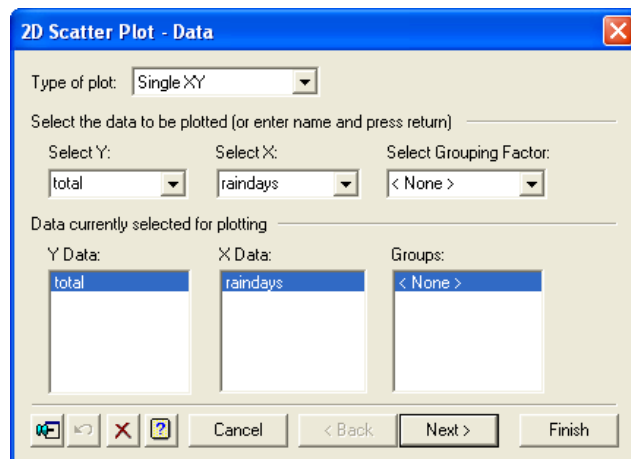
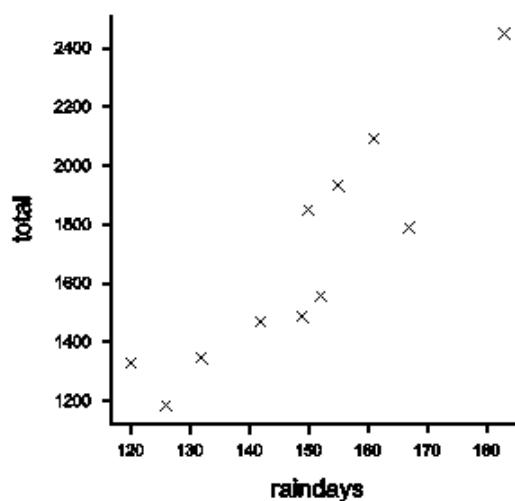
- ⇒ Find the **Summarize Contents of Variates** dialogue again. Click on the  button at the bottom right of the dialogue to clear all currently selected statistics. Reselect the variables and choose **Arithmetic Mean** and **Standard Deviation**. Click on **[More Statistics...]** and select **Standard Error of Mean**, and click **[OK]**. Click **[Run]**. You will notice the output now includes the standard error of the mean.
- ⇒ Use **Graphics** ⇒ **2D Scatter Plot** and complete the dialogue box as shown in **Fig. 2.3e**. Click **[Finish]** to produce the graph shown in **Fig. 2.3f**.





Fig.2.3d Choosing the scatter plot dialogue**Fig.2.3e Specifying the y and x**


The relationship between the total rainfall and the number of rain days is as follows:



Fig. 2.3f The graph is in its own window



⇒ Close the graph window with **File** ⇒ **Exit** (choosing **[No]** to the question about saving the graph).

Since Genstat version 8, there is a new button which toggles between  and . If the button is in the sideways position , clicking on **[Run]** will run your analysis and close the dialogue box. If the button is in the upward position , the dialogue box will be kept open so it can be used again. In this position, if you want to close the dialogue box, you could click on **[Cancel]**, or click on the cross at the top right hand corner of the dialogue box.

In the position , it is quite easy to get a large number of windows and dialogue boxes open at the same time, so it can be quite hard to find the one for which you are looking.

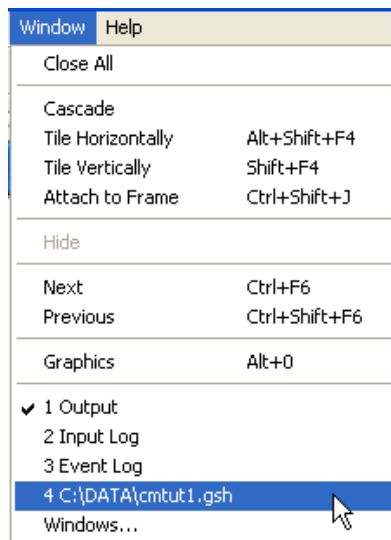
Clicking the  or  buttons in the tool bar can help find the one you want. Alternatively, to find a particular dialogue or menu box, just repeat the menu commands that opened it (e.g. **Graphics** ⇒ **2D Scatter Plot**) as this will bring back the box complete with anything that had been previously entered. It is a good idea to close a box by clicking **[Cancel]**, as soon as it is no longer required.

2.3.2 Calculating and formatting columns

It is easy to calculate new variates from those already entered. In this example, it would be interesting to find the mean rain per rainy day in each year. This is simplest to do within the spreadsheet.

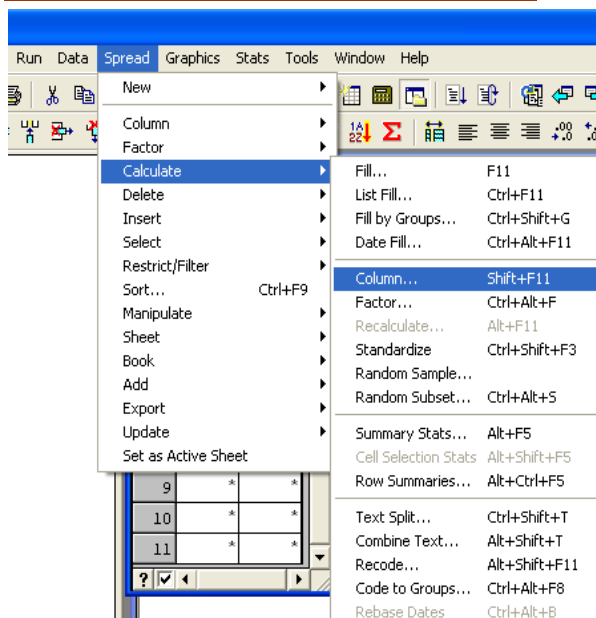
- ⇒ First, the spreadsheet needs to be selected. Do this, either by clicking somewhere in it (if you can see it), or use the toolbar arrow buttons or the **Window** menu, as shown in **Fig. 2.3g**.

Fig.2.3g Selecting the spreadsheet

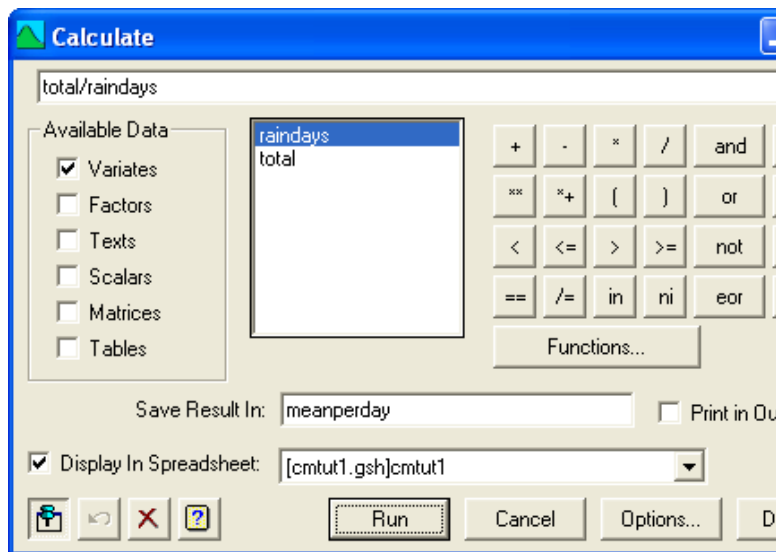


- ⇒ To calculate a new column, choose **Spread** ⇒ **Calculate** ⇒ **Column** as shown in **Fig. 2.3h**.

Fig. 2.3h Choosing the calculate dialogue



- ⇒ Complete the box as shown in **Fig. 2.3i**. The calculation can either be typed into the top box, or you can use the mouse to click on the operator buttons and double click on the variates as required.

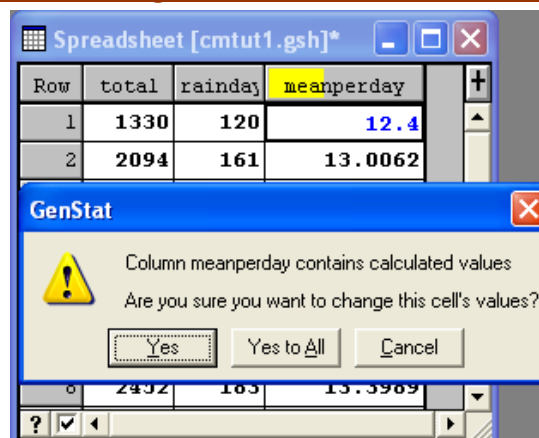
Fig. 2.3i Giving the formula**Fig. 2.3j The results**

Row	total	rainday	meanperday
1	1330	120	11.0833
2	2094	161	13.0062
3	1851	150	12.34
4	1470	142	10.3521
5	1557	152	10.2434
6	1932	155	12.4645
7	1184	126	9.39683
8	2452	183	13.3989
9	1347	132	10.2045
10	1792	167	10.7305
11	1488	149	9.98658

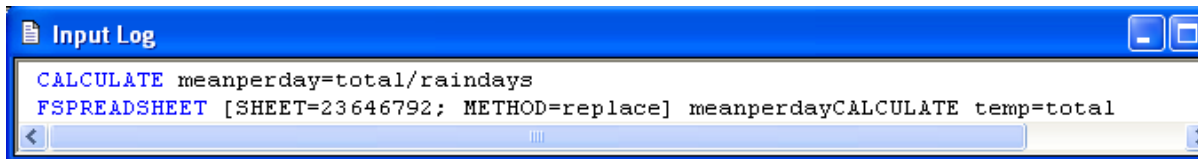
Type the name of the new column into the bottom box labelled **Save Result In**, *Fig. 2.3i*. Then click **[Run]**. Click **[Cancel]** after this to remove the dialogue box.

There is now a new variate, called *meanperday*, added to the spreadsheet, as shown in *Fig. 2.3j*, which holds the 11 values of the mean rain per rain day. The name is part shaded (in yellow on a colour screen) to indicate that the column *meanperday* is a calculated column.

⇒ To illustrate the difference between an ordinary and a calculated column, try to change a value in the *meanperday* column. GenStat gives a warning, see *Fig. 2.3k*.

Fig. 2.3k Showing that the column was calculated

Thus GenStat's spreadsheet is a little like an ordinary spreadsheet in that it records the calculation, rather than just doing the transformation. If you change a value in the original column, the derived values do not, however, change automatically. You could then use **Spread** ⇒ **Calculate** ⇒ **Recalculate**, to update the derived values.

Fig.2.3l Commands to keep a record of your work


```

CALCULATE meanperday=total/raindays
FSPREADSHEET [SHEET=23646792; METHOD=replace] meanperdayCALCULATE temp=total

```

You may have noticed that commands have been appearing in the **Input Log** as you work. This is a record of what you have done, written in the GenStat command language. You can re-run any of these commands with the **Run** menu, or copy them into a new window to make a program.

To understand for instance what has happened within GenStat when you did the last calculations, we have shown the input log in the figure above. There you see, in the first line, that the spreadsheet generated a command that was executed by the GenStat Server. The next line shows that the results were then passed back to the spreadsheet. (Ignore the large number there, 23646792, which may well be different when **you** run the commands. It is an internal reference number so GenStat knows which spreadsheet contains the new column.)

Calculations will normally be done in a spreadsheet as above. Once you become experienced in using GenStat, you could alternatively do calculations only in the GenStat server, using the **Data** ⇒ **Calculations** menu, rather than the **Spread** ⇒ **Calculate** ⇒ **Column** route that you used above. The result is the same to the GenStat Server, but you would not automatically see the calculated column in a spreadsheet.

In the spreadsheet, each value of *meanperday* is displayed with the same number of significant digits. This may lead to a variable number of decimal places across the different values in the column.

- ⇒ You can change this with **Spread** ⇒ **Column** ⇒ **Attributes/Format**, **Fig. 2.3m**. Make sure that you select *meanperday* in the column box. A faster way is to right-click in the *meanperday* column and to choose the **Column Attributes** option. The same Column Attributes Dialogue Window will appear.
- ⇒ Type '1' in the **Decimals** box **Fig. 2.3n**, and check that **Fixed** is now the *Numeric Format*. You will now observe that *meanperday* appears in the spreadsheet with just one decimal place. You may also wish to enter a concise explanation of the contents of the column in the **Description** box. Click **[OK]** to effect the change.

Fig. 2.3m Choosing the Format dialogue

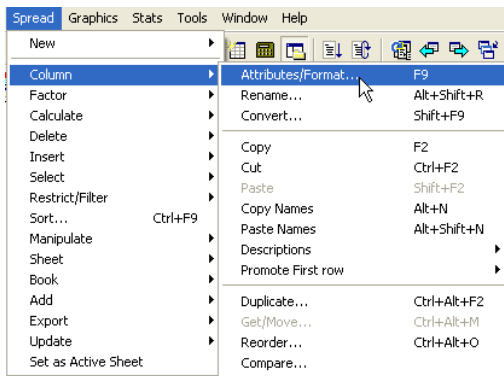
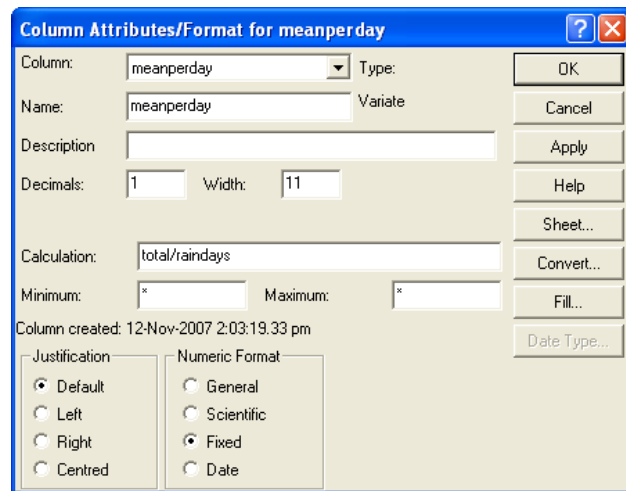


Fig. 2.3n Give the column one decimal place



Now assume that these data values in your spreadsheet came from 11 years in order. It would be useful to have this information entered too.

- ⇒ Make your spreadsheet active and click in the first column (*total*) of the spreadsheet. Choose **Spread** ⇒ **Insert** ⇒ **Column before Current Column**. This gives a dialogue box called **Create a new column** as shown in **Fig. 2.3o**.
- ⇒ Type *year* in the name box and click on **[OK]**. A new column will appear in the spreadsheet filled with missing values (denoted by ***) as in **Fig. 2.3p**. You could now type in the numbers 1 to 11, or the real years, if they are known, but there is a quicker way to fill in regular sequences.

Fig. 2.3o Spread ⇒ Insert ⇒ Column

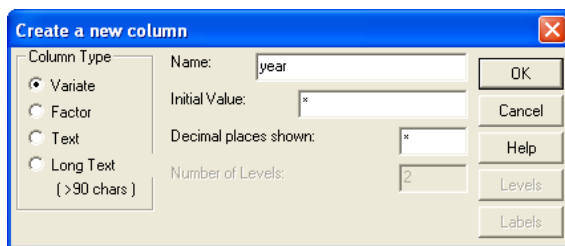
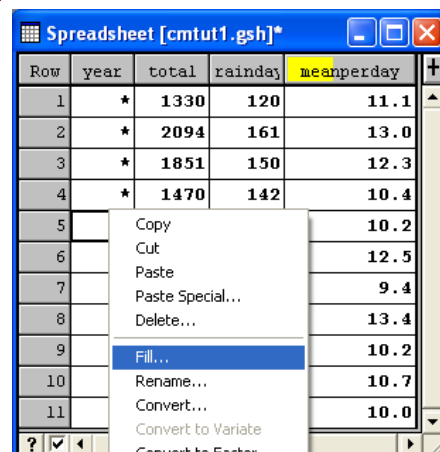
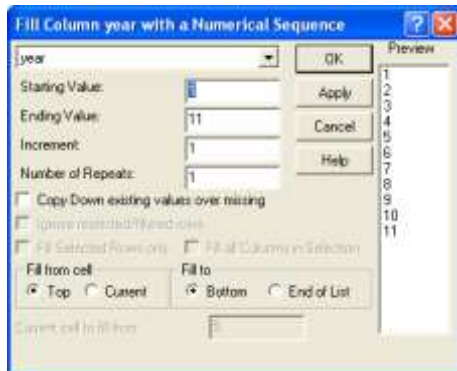
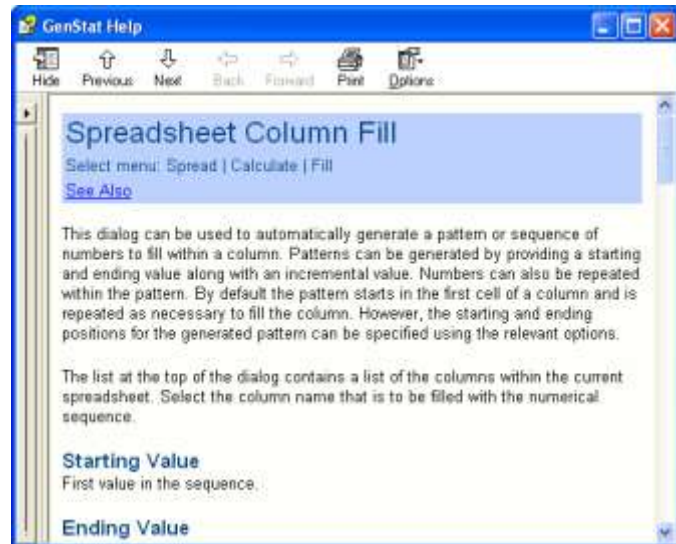


Fig. 2.3p Right-click and choose Fill



- ⇒ Right click in the Spreadsheet and choose **Fill** from the popup menu as shown in **Fig. 2.3p** or choose **Spread** ⇒ **Calculate** ⇒ **Fill**. In the **Fill** dialogue, shown in **Fig. 2.3q**, make sure that **year** is in the top box. Clicking **[OK]** will fill **year** with the numbers 1 to 11. **Fill** can also be used to make patterned sequences.
- ⇒ Details of the use of this, or any other dialogue, can be found by clicking the **[Help]** button in the dialogue box. An example is given in **Fig. 2.3r**.

Fig. 2.3q Complete dialogue**Fig.2.3r Try the Help button**

- ⇒ Try plotting the mean per day against the year as a line graph. Use **Graphics** ⇒ **2D Line Plot** ⇒ **SingleXY** type with *meanperday* as the **Y** and *year* as the **X**. Click **[Finish]**.

Now investigate the graph. What is the year with the lowest mean? Is there any obvious pattern?

- ⇒ Remember to close the graph with **File** ⇒ **Exit** or by clicking on the cross at the top right of the graph.

2.4 Factor Columns

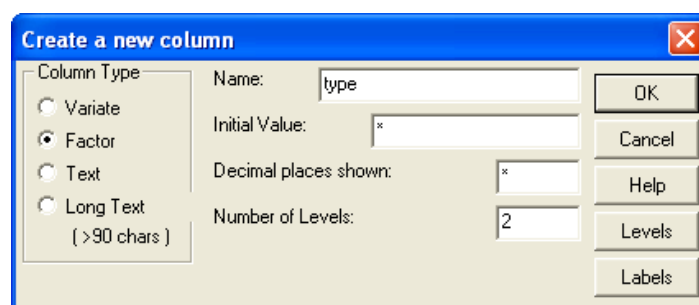
2.4.1 Introducing factors

So far, all the information entered into GenStat has been numerical. It is possible to include textual information as well. One structure that accepts this kind of information is a **factor**. This is a special column used to indicate groups in the data.

Four years in the data set used in the previous sections were El Nino years, the second, third, eighth and tenth. The remaining years were ordinary years. So we will make a factor with two groups or **levels**, and here, one is labelled E and the other O.

- ⇒ Click in the first column of the spreadsheet (*year*) and choose **Spread** ⇒ **Insert** ⇒ **Column after Current Column**. Give the name *type* into the **Name** box, and click to select **Factor** under **Column Type** in **Fig. 2.4a**. The dialogue will change to let you specify the number of factor levels.

Fig. 2.4a Spread ⇒ Insert ⇒ Column



- ⇒ Specify that the factor has 2 levels and then click on the **[Labels]** button. The dialogue shown in **Fig. 2.4b** appears. Type 'E' and press the **[Enter]** key. The next level (2) will become selected. Type 'O', press **[Enter]** and then click **[OK]** to make the changes take effect.

Fig.2.4b Add labels to the factor

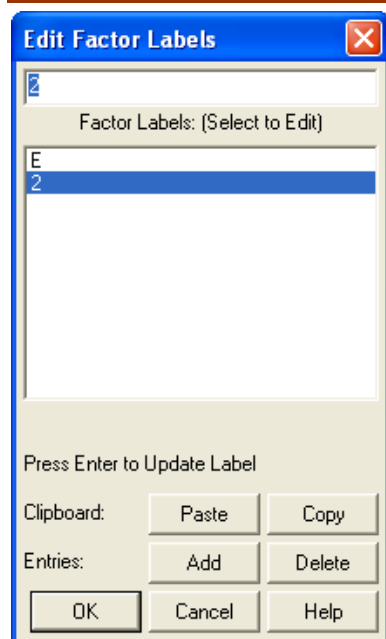
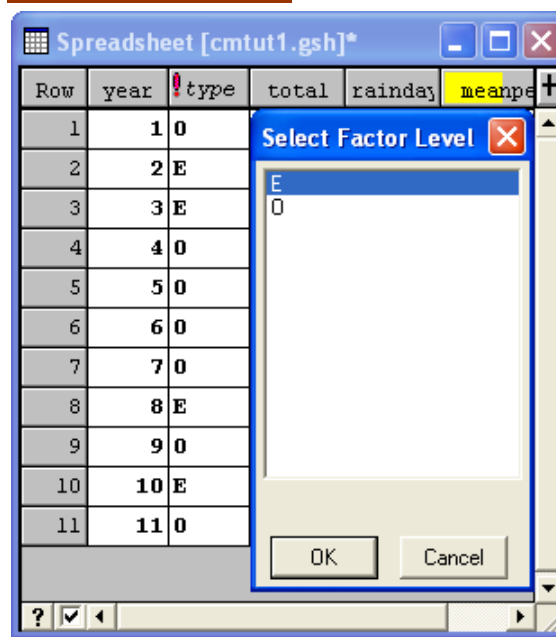


Fig. 2.4c Entry of data



⇒ Click **[OK]** in the **Create a new column** dialogue, as shown in **Fig. 2.4a** to make the new column, which contains empty cells.

⇒ Now type the following values into the new column, as shown in **Fig. 2.4c**.

O E E O O O O E O E O

If you make a mistake by typing lower case 'e' instead of an upper case 'E', GenStat will turn it into an upper case 'E'; if you type the wrong letter, GenStat will give you a message and ask you to retype your entry. Double clicking gives a pop-up menu, as shown above (**Fig 2.4c**), which lists the allowable levels.

⇒ The *factor* column can be used to label a graph. Choose **Graphics ⇒ 2D Scatter Plot....** Fill in the boxes as in **Fig. 2.4d**, including the *Select Grouping Factor:* box.

⇒ Click **[Next]** and in the dialogue corresponding to *Lines and Symbols*, request a **X Cross** as the symbol for *Plot1 - Group1* and a **O Circle** for *Plot1 - Group2*. Click **[Finish]** to produce the graph shown in **Fig. 2.4e**.

Fig. 2.4d Graphics ⇒ Point Plot

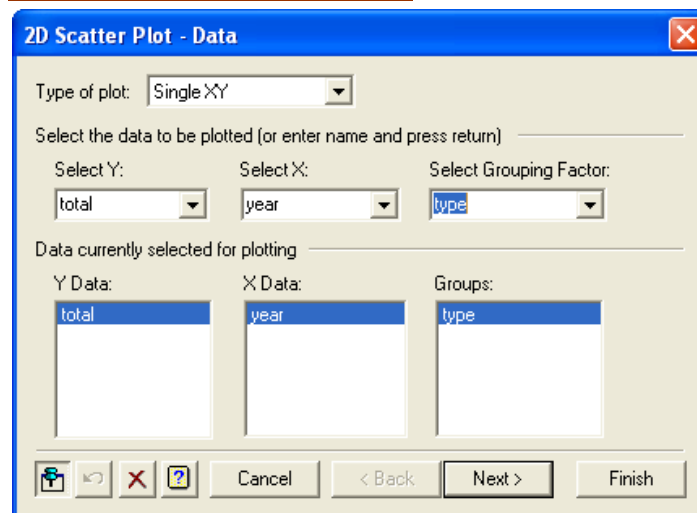
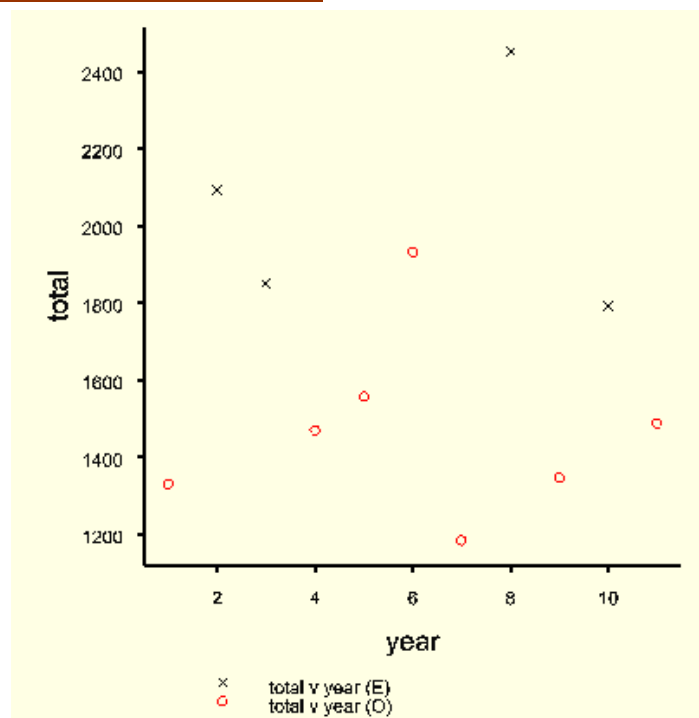
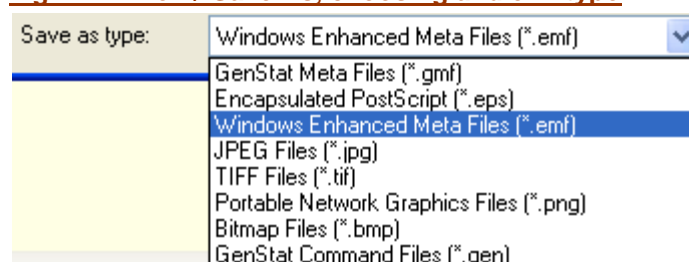


Fig. 2.4e Resulting graph

By using the **Edit** ⇒ **Edit Graph** once you have the graph, and then by double clicking on the graph, you can choose to edit the Axes key or graph symbols. If you wish, you could try producing a different layout to that above. Edit facilities can be used to modify the layout of the graph until it is ready for reporting or publishing.

Graphs can be saved in different formats by choosing **File** ⇒ **Save as** see **Fig. 2.4f**. They can be read into WORD using the menu sequence **Insert** ⇒ **Picture** ⇒ **From File...** in WORD.

You leave the GenStat Graphics Window by choosing **File** ⇒ **Exit** from the menu bar.

Fig. 2.4f File ⇒ Save As, choosing and emf type

Back in the spreadsheet, the column called *type* can be modified to display longer labels.

⇒ Select the *type* column in the spreadsheet. Right click and choose **Column Attributes**. Click the **[Labels]** button, and edit the labels (to be *El Nino* and *Other*), making sure that you press **[Enter]** after typing each new label. Click **[OK]** twice when you have finished, and the labels in the variety column should now be modified.

Alternatively, the full labels could have been entered when the factor was first created. You would still have been able to enter the values into the column by typing E or O only, the first letter of the labels.

Earlier, you used **Stats** ⇒ **Summary Statistics** ⇒ **Summarise Contents of Variates** to give some summaries of the data. Now, with the data in two groups, it is useful to give the summaries for each group individually. The dialogue used earlier can be used for this, but a more general alternative is:

⇒ **Stats** ⇒ **Summary Statistics** ⇒ **Summaries of Groups (Tabulation)**, *Fig. 2.4g*.

Fig. 2.4g Stats ⇒ Summary ⇒ Tabulation

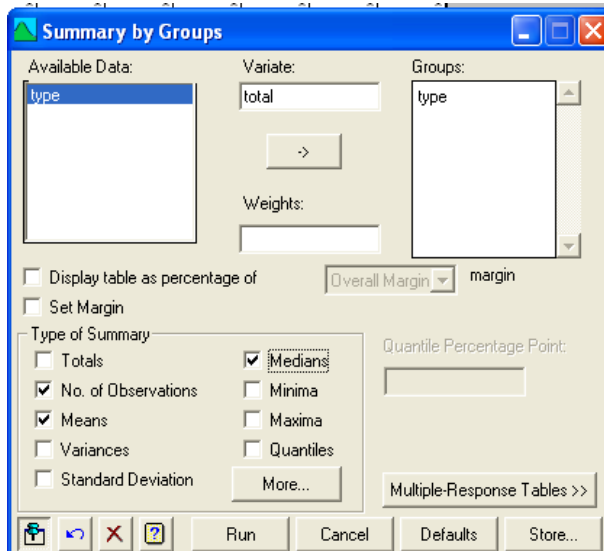


Fig. 2.4h Results

type	Nobservd	Mean	Median
El Nino	4	2047	1972
Other	7	1473	1470

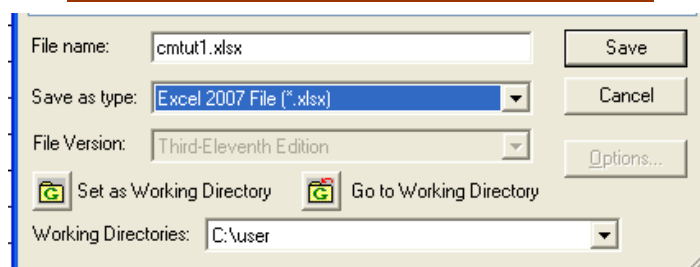
⇒ Complete the dialogue as shown and press **[Run]**. The results are shown in *Fig. 2.4h* in the Output Window.

2.4.2 Saving data

Before continuing, save the spreadsheet.

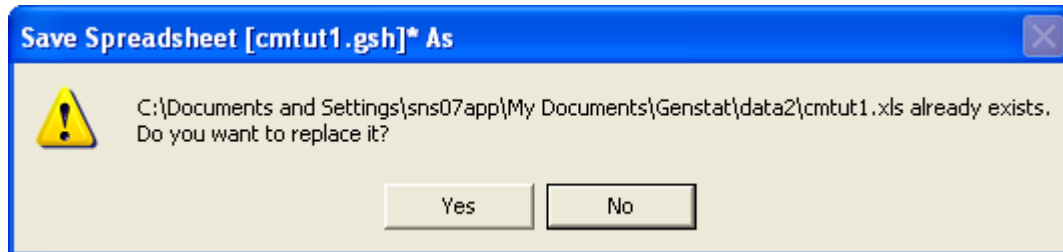
⇒ Choose **File** ⇒ **Save as**. By default, a Window appears suggesting you save the data as a **GenStat spreadsheet** (*.gsh). A wide range of other file formats is also available.

Fig. 2.4i File ⇒ Save As, then change to Excel



In Section 2.2 we showed how data could be imported from an Excel worksheet or could be entered directly in GenStat using the Spread menu. In Section 2.2.4 we had imported the file *cmtut1.xls* from Excel and have modified it. If you change the format in the figure above, and specify an Excel file then, when you try to save, you get a warning message.

Fig. 2.4j GenStat warning that the original file will be replaced



2.4.3 Deleting data

In this section we will delete the column, called *meanperday*, that has been generated. We also show the difference between deleting a whole column and deleting its contents.

- ⇒ First select the column *meanperday*. Then click in the name field (or press **<ALT><Ctrl>C**, or use **Spread** ⇒ **Select** ⇒ **Current Column**). Clicking again will deselect the column. Practice selecting and deselecting columns. Finishing with the *meanperday* column selected.
- ⇒ Once selected, you might think that the **<Delete>** key should delete the column. Press the **<Delete>** key. It deletes (as expected) just the data. The column remains! Use **Edit** ⇒ **Undo Del Cells** to get the column back.
- ⇒ What you need to do is to delete the whole column. The column should still be selected. Use **Spread** ⇒ **Delete** ⇒ **Current Column**. You can also select one or more rows and delete them in the same way.

2.4.4 Available variables

- ⇒ You can check which variables are currently available to the GenStat server by referring to the Data window pane on the left hand side of the GenStat screen. This is like a Windows explorer with folders and subfolders whose contents are shown by clicking on the nodes, as shown in **Fig. 2.4k**.
- ⇒ If the Data window pane is not visible, you can make it show again using the menu selection **Data** ⇒ **Display** or pressing the **F5** key; see **Fig. 2.4l**.

Fig.2.4k Data display window pane

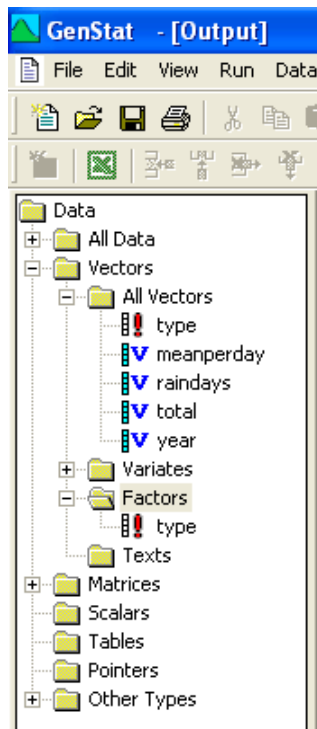
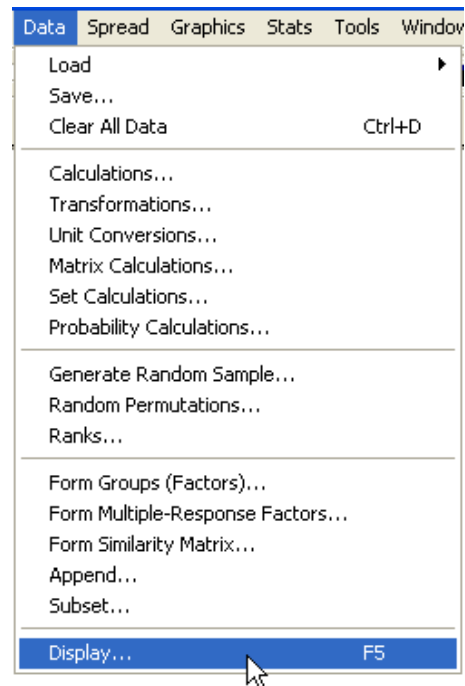


Fig. 2.4l Show the Data display window pane



This lists the names of the structures and their **types**. All structures used so far are **variates** (meanperday, raindays, total and year) and **factors** (type), but later on you will use other types of columns too. This is also a useful dialogue box from which you can delete columns when they are no longer needed.

- ⇒ To print a variable in the Output window, right click on the relevant node in the **Data** window pane, select **Display** and click **OK**.

2.5 Understanding how GenStat works

2.5.1 A first introduction to the GenStat command language

Although in Chapter 2.1, we mentioned that GenStat is basically a standard Windows application, the truth is a bit more complex. Before the Windows version you could use GenStat as long as you knew the "language". You simply typed commands, which you submitted to GenStat.

GenStat 11th Edition is indeed a Windows application, but the menus are based on an underlying command language. You can still use GenStat by typing commands in a new Text Window as we show now. At the same time, we show how GenStat may be used as a calculator.

- ⇒ Restart GenStat. Use **File** ⇒ **New** and choose the Text Window, **Fig 2.5a**. This gives you an **Input Window**. In this window, type **Print 3+4** as shown in **Fig. 2.5b**.

Fig.2.5a File ⇒ New

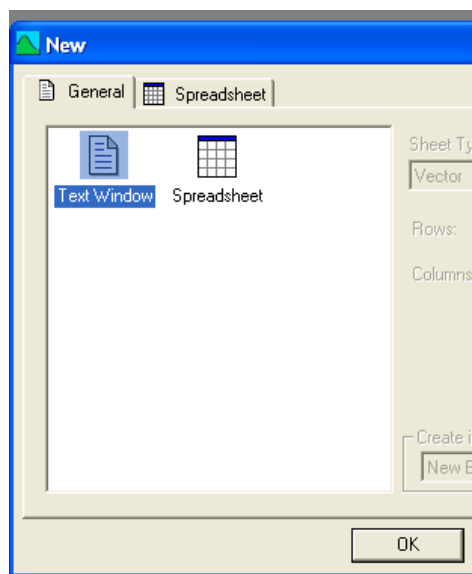
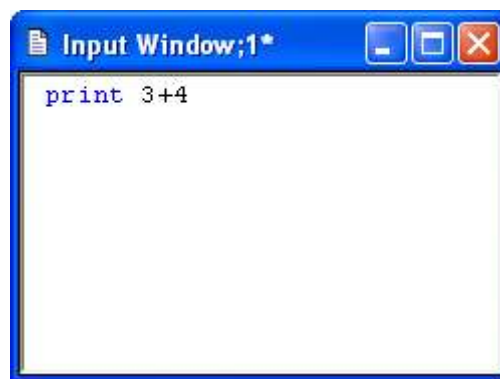


Fig. 2.5b Complete the text window



- ⇒ Now select the **Run** menu as shown in **Fig. 2.5c**. You can choose either **Submit Line** (or the shortcut key combination Control+L if the cursor is still on the line you typed) or **Submit Window**.

Fig.2.5c The Run menu

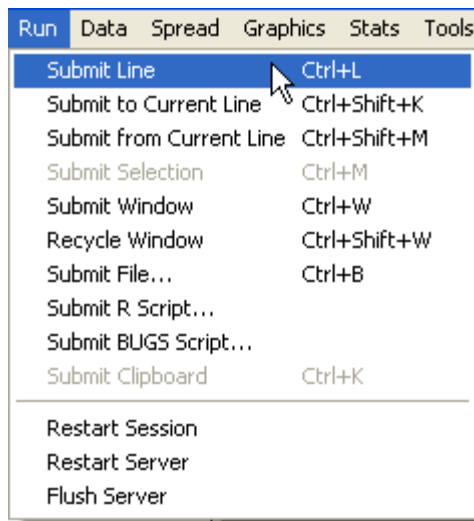
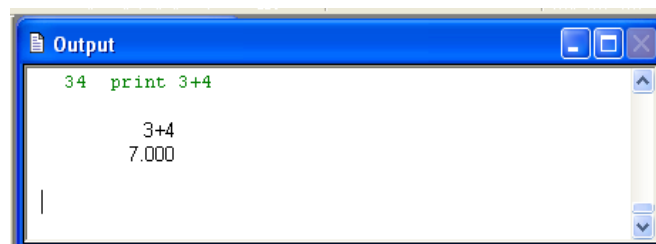


Fig. 2.5d Results



You have now submitted your "program" of commands to the GenStat server. The results are put in the **Output Window**, see **Fig. 2.5d**.

You can go to the output window in various ways, e.g. by using the Windows menu. There you see that GenStat normally "echoes" the command and shows you that $3+4=7$.

An alternative to typing the command is to use the Data menu **Data** ⇒ **Calculations**, see **Fig. 2.5e**. This gives the dialogue shown in **Fig. 2.5f**.

Fig. 2.5e Choose the dialogue

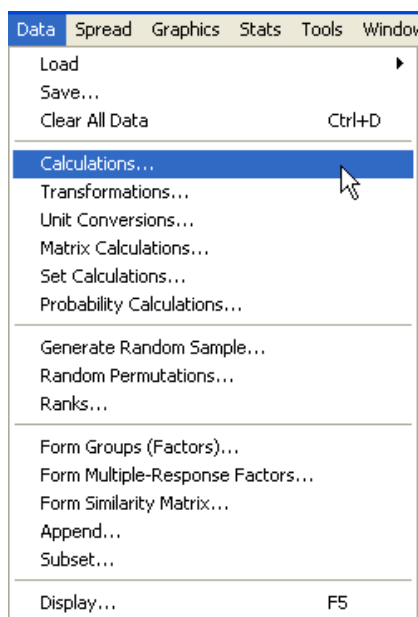
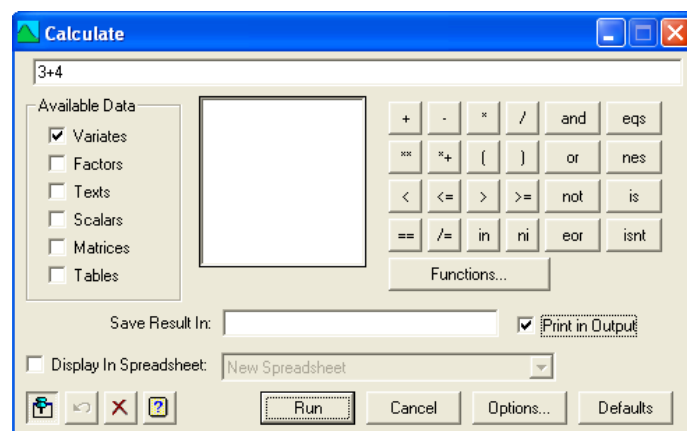


Fig. 2.5f Another way of calculating



Then type **3 + 4** as the function, click on **Print in Output** and then on **[Run]**. If you look in the Output window (**Fig. 2.5g**), you see that $3 + 4$ still equals 7!

Fig. 2.5g 3 + 4 is still 7!

```
print 3+4
3+4
7.000

PRINT 3+4
3+4
7.000
```

Fig.2.5h The Input log

- ⇒ The **Input Log Window** is also useful. It keeps a record of all the commands you have submitted. Access it by **Window** ⇒ **Input Log**. You see that the use of the **Calculation** menu has resulted in GenStat preparing the commands `PRINT 3+4` for you and has submitted them to the GenStat server.

So, that is how GenStat works. You prepare commands, which are submitted to the GenStat server. The Windows version has simply given you a variety of ways to prepare the commands for GenStat. GenStat obeys the commands and puts the results in the **Output Window**. It keeps a record in the **Input Log Window**.

If the commands produce graphs, then GenStat puts the graphs in a **Graphics Window**. If you make a mistake in the command, it prints an error message in the **Event Log Window** (and in the **Output Window**).

The example above ($3 + 4 = 7$) indicates that GenStat may be used as a simple calculator. This is worth a little practice. It is useful to have a scientific calculator. Also it is sometimes useful to transform data.

- ⇒ For example, if you want to calculate the difference between 4.35 and 2.37 expressed as a percentage of 4.35, open the calculator with **Data** ⇒ **Calculations**, check that **Print in Output**, is still ticked and type the following calculation in the top box:

100 * (4.35 - 2.37) / 4.35

- ⇒ Click **[Run]**. This will give the following in the output window:

```
100*(4.35- 2.37) / 4.35
45.52
```

i.e. the difference is 45.52% of 4.35.

It is important that the brackets () are included where appropriate to make sure that the calculation you are trying to do has only one meaning.

The symbols +, -, *, / are used for the operations of addition, subtraction, multiplication and division respectively and ** is used for powers. There are also various mathematical functions available. One is for calculating the square root of a number. The function is **SQRT()**, where the number whose square root is required is given in the parenthesis, for example **SQRT(12.37)**. **Fig. 2.5i** gives an overview of how to perform some calculations by using the Input Window. More information can be found in the GenStat Help file under '*Calculations and manipulations*', accessed via **Help** on the tool bar, viewing **Contents and Index**, and then clicking on GenStat Command Language.

⇒ Try more calculations to see how all this works, using both an **Input window** and the **Data** ⇒ **Calculations** dialogue box. Some examples are given below.

Fig. 2.5i Some basic calculations using the Input Window

<u>Symbol</u>	<u>Operation</u>	<u>Example</u>	<u>Result</u>
+	addition	PRINT 3+4	7.000
-	subtraction	PRINT 3-4	- 1.000
*	product	PRINT 3*4	12.00
/	division	PRINT 3/4	0.7500
**	exponentiation	PRINT 3**4	81.00
<u>Function</u>	<u>Operation</u>	<u>Example</u>	<u>Result</u>
SQRT(x)	Square root	PRINT SQRT(4)	2.00
EXP(x)	Exponential function	PRINT EXP(1)	2.718
LOG(x)	natural logarithm of x, for x > 0	PRINT LOG(2.718)	0.9999
LOG10(x)	logarithm to base 10 of x, for x > 0.	PRINT LOG10(10)	1.000
ROUND(x)	rounds the values of x to the nearest integer.	PRINT ROUND(1.234)	1.000
Other examples			
PRINT 1/2			0.5000
PRINT 100*(4.35 -2.37)/4.35			45.52

2.5.2 Server sessions

⇒ After the above calculations, the Input and Output Windows look a mess. All the data can be cleared out of the GenStat server with **Data** ⇒ **Clear All Data** or **Run** ⇒ **Restart Session**. Less drastically, you can 'clean up' the output window by clicking the 'Clear Output' button (



) in the toolbar.

3. Simple statistical inference

In the analysis so far, we have just considered descriptive statistics. Thus we have summarised the data numerically and drawn graphs. In this example, we introduce ideas of statistical inference. We take an example from *Mead, Curnow and Hasted (2003)*, pages 33-34 and 38-39. This compares wheat yields for 6 farmers where there was a new system of giving agroclimatic advice, compared to 10 farmers, who used the standard system. The yields, in tons per hectare, were as follows:

new: **2.5 2.1 2.4 2.0 2.6 2.2**
 standard: **2.2 1.9 1.8 2.1 2.1 1.7 2.3 2.0 1.7 2.2**

3.1 The use of boxplots

- ⇒ Use **Spread** ⇒ **New** ⇒ **Create** as shown earlier in Chapter 2.2.1 to create a blank spreadsheet and set it to have 2 columns of 16 rows. Enter the data as shown in **Fig.3.1a** and give the first column the name *yield* and the second column the name *system*. Make this latter column in to a factor and add labels to its numerical values (see chapter 2.4.1).
- ⇒ Save the spreadsheet, giving it the name **cmtut2.gsh** (see Chapter 2.3.4 if you need instructions on saving).

Fig. 3.1a Dataset

Row	yield	system
1	2.5	1
2	2.1	1
3	2.4	1
4	2	1
5	2.6	1
6	2.2	1
7	2.2	2
8	1.9	2
9	1.8	2
10	2.1	2
11	2.1	2
12	1.7	2
13	2.3	2
14	2	2
15	1.7	2
16	2.2	2

Fig. 3.1b Graphics ⇒ Boxplot

Boxplot - Data

How are the data organized?

List of variates Single variate with groups

Select a single data set to be plotted (or enter name and press **Enter**)

Select Data:

yield

yield

↓

Data currently selected for plotting

Data: yield

Groups: system

Labels:

Buttons: **Finish**, **Cancel**, **< Back**, **Next >**

- ⇒ One way to present the data is to use a boxplot. Use **Graphics** ⇒ **Boxplot**, complete the dialogue as shown in **Fig. 3.1b** and click **[Finish]**. This gives the display shown in **Fig. 3.1c**.

Fig. 3.1c Resulting graph

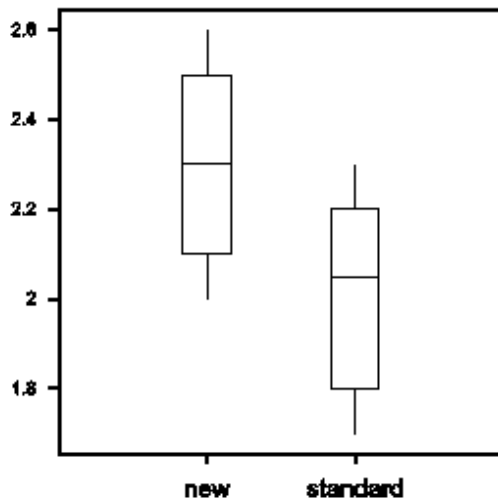


Fig. 3.1d Changed

Row	yield	system
1	2.5	New
2	2.1	New
3	2.4	New
4	2	New
5	2.6	New
6	2.2	New
7	2.2	Standard
8	1.9	Standard
9	1.8	Standard
10	2.1	Standard
11	2.1	Standard
12	1.7	Standard
13	2.3	Standard
14	2.9	Standard
15	1.7	Standard
16	2.2	Standard

One use of boxplots is to show outliers.

- ⇒ Return to the current spreadsheet, insert a value of **2.9** instead of 2 for the 8th value in the Standard group (row 14) and re-do your boxplot. The general shape of the graph is the same, but the odd value is indicated as deserving close scrutiny.
- ⇒ There are two ways of displaying the boxplot. Use **Graphics** ⇒ **Boxplot** and click **[Next]**. You can now choose between two types: Box and Whisker, **Fig. 3.1e** and Schematic, **Fig. 3.1f**. Try both, as shown below. The advantage of a schematic boxplot is that you can easily discover outliers.

Fig. 3.1e Box and Whisker plot

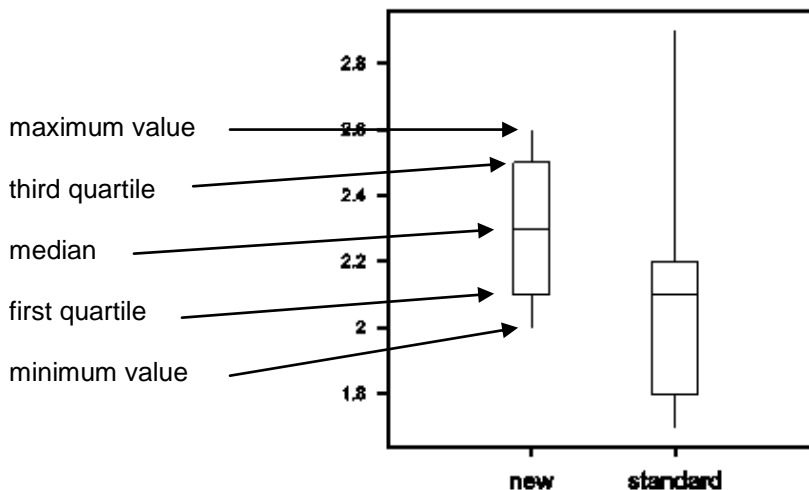
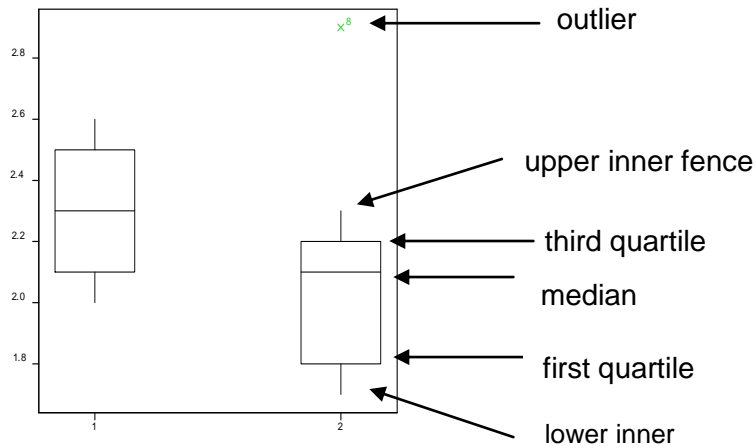


Fig. 3.1f Schematic boxplot



In a Box and Whisker boxplot, the ends of the whiskers mark the minimum and maximum values of the data set, in a schematic boxplot they mark the 'upper and lower inner fences'. The upper inner fence is defined as the upper quartile plus 1.5 times the interquartile range, or the maximum value if that is smaller. The lower fence is defined similarly.

Extreme values between 1.5 and 3 times the interquartile range (plus the upper or minus the lower quartile) are by default marked as green crosses. More extreme values (more than 3 times the above mentioned range) are marked as red crosses.

⇒ If you made this change, then set the edited value back to 2.0 in the spreadsheet before continuing.

3.2 Comparisons of means

⇒ Simple comparisons of the means of two different samples can be made with **Stats** ⇒ **Statistical Tests** ⇒ **One and two sample t-tests**. Complete the dialogues as shown in **Fig. 3.2a** and **3.2b**. The **[Options]** button in **Fig. 3.2a** gives the dialogue in **Fig. 3.2b**, but keep this unchanged for this example. Click **[OK]**, then **[Run]**.

Fig. 3.2a Stats ⇒ Tests ⇒ One and Two Sample

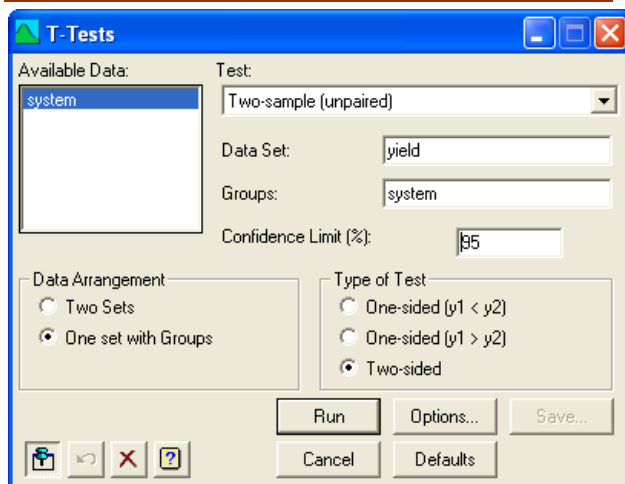
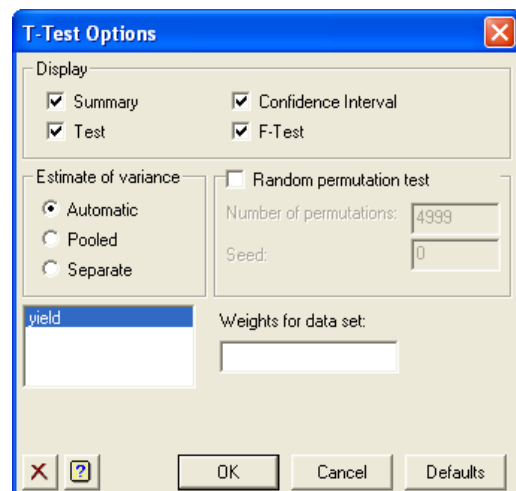
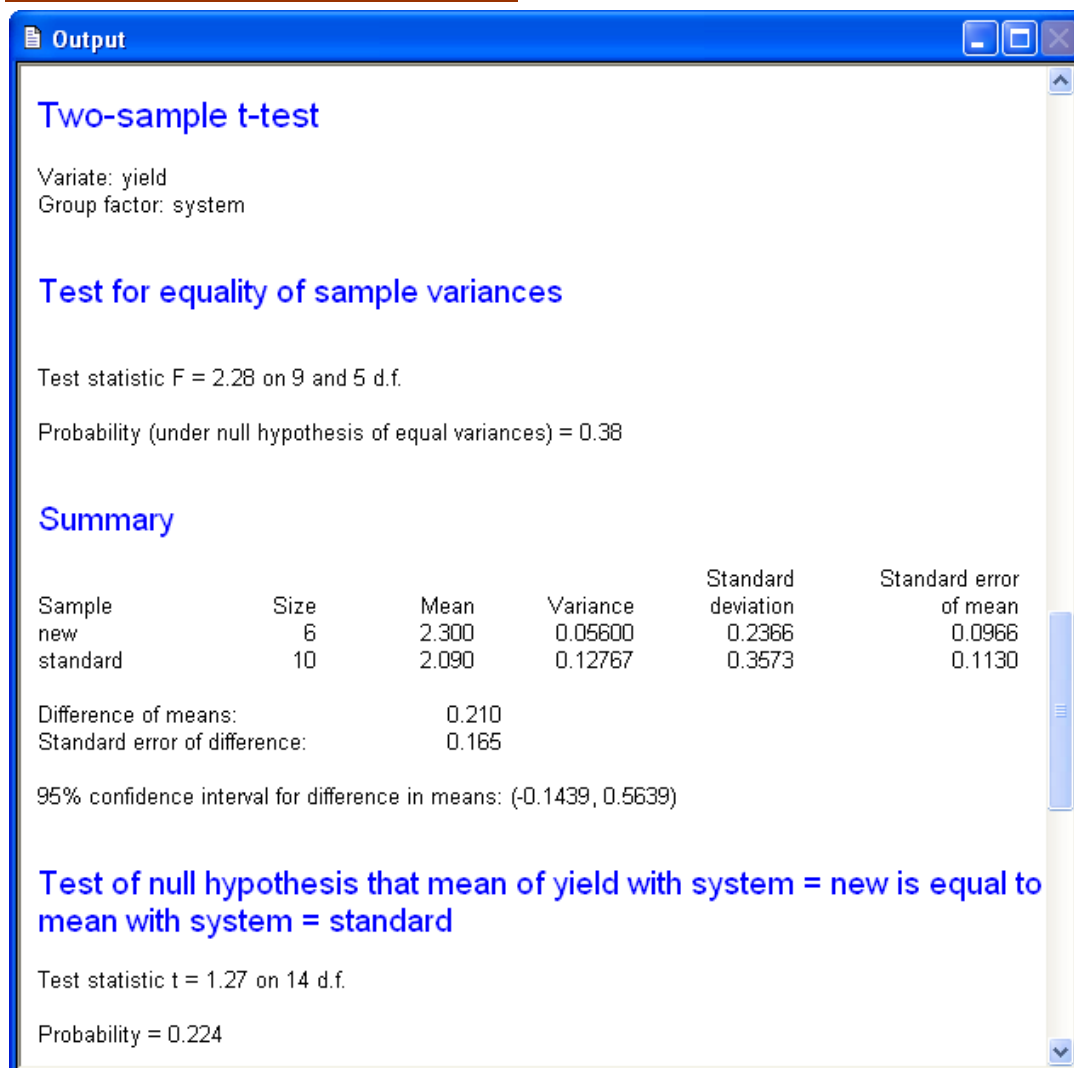


Fig. 3.2b Options sub-dialogue



The output window shows the results, see **Fig. 3.2c**.

Fig. 3.2c Output from the t-test dialogue

References

Mead, R., Curnow, R. N. and Hasted, A. M. (2003) *Statistical Methods in Agriculture and Experimental Biology*, 3rd edn. Boca Raton: Chapman & Hall/CRC Press.

4. Simple linear regression

4.1 Setting up the data

We now introduce some key elements of data analysis, by means of simple linear regression. This example is taken from pages 178 - 181 of *Mead, Curnow and Hasted (2003)*.

⇒ Use **Run** ⇒ **Restart Session** to start a new job. Accept the option **[Yes]** to proceed.

Fig. 4.1a Restart again

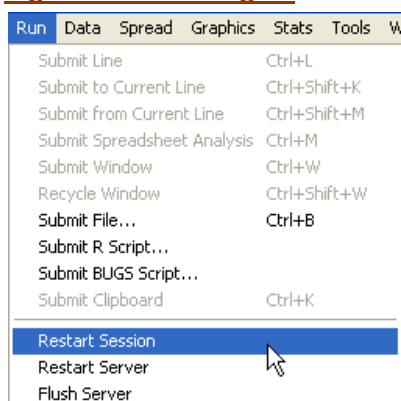
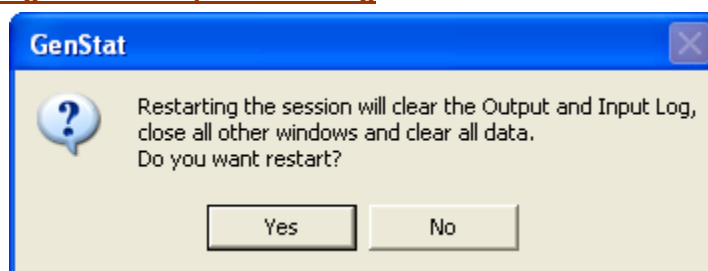


Fig. 4.1b Accept the warning



⇒ Use **Spread** ⇒ **New** ⇒ **Create** and make a spreadsheet with 2 columns and 17 rows as shown in **Fig. 4.1d**.

Fig. 4.1c Create a new sheet

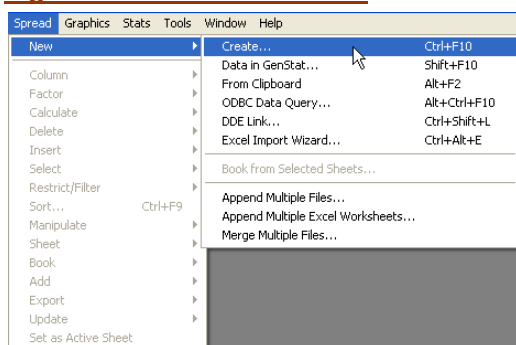
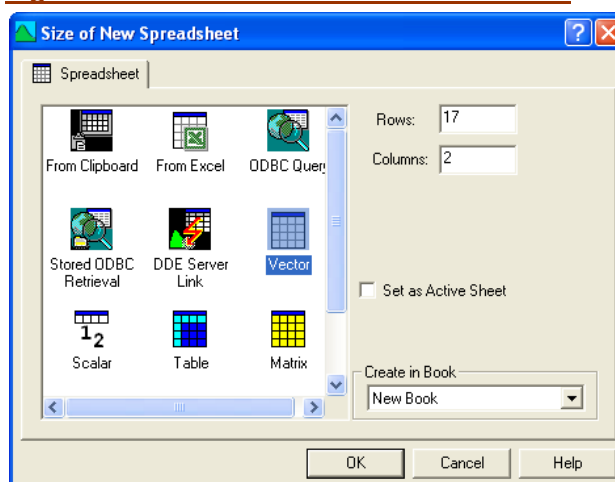
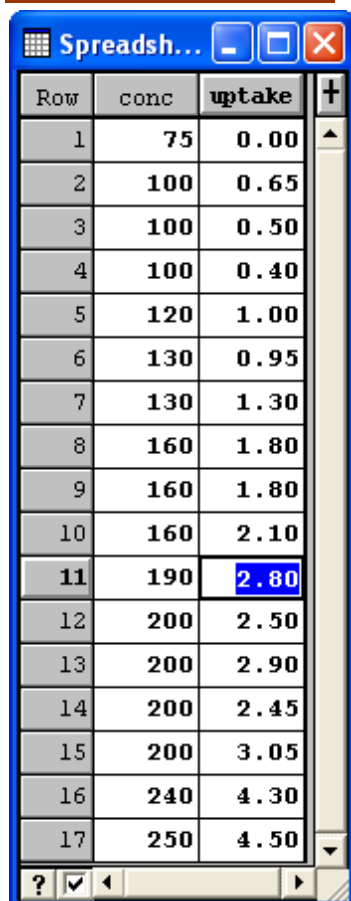


Fig. 4.1d Sheet with 17 rows and 2 columns



- ⇒ Enter the data and name the two columns as shown in **Fig. 4.1e** (see page 3 for instructions on naming columns, if necessary). Save the data giving the file the name **cmtut5.gsh**. Click outside the spreadsheet to transfer the data to the GenStat server. This gives some default summary statistics for each of the two columns.

Fig. 4.1e Regression data



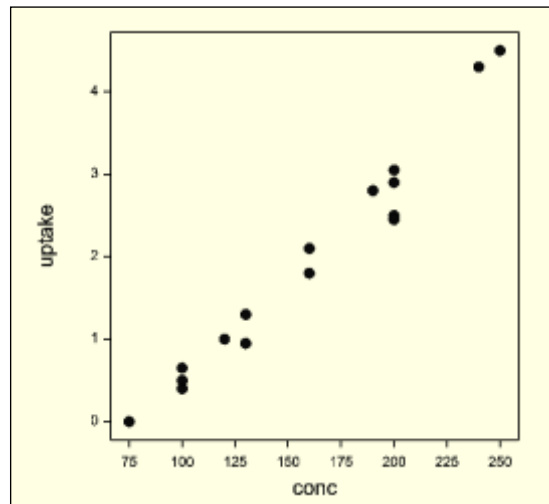
Row	conc	uptake
1	75	0.00
2	100	0.65
3	100	0.50
4	100	0.40
5	120	1.00
6	130	0.95
7	130	1.30
8	160	1.80
9	160	1.80
10	160	2.10
11	190	2.80
12	200	2.50
13	200	2.90
14	200	2.45
15	200	3.05
16	240	4.30
17	250	4.50

- ⇒ Choose **Graphics** ⇒ **2D Scatter Plot...** and complete the dialogue as shown in **Fig. 4.1f** to give the scatterplot.

Fig. 4.1f Graphics ⇒ Scatter plot



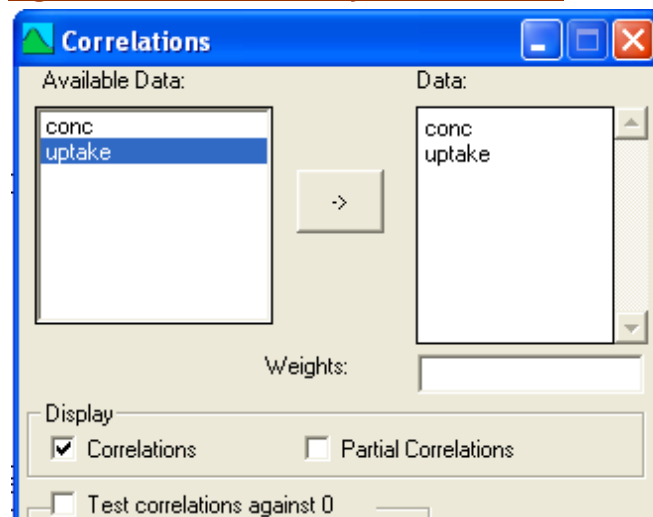
Fig. 4.1g Results



4.2 Correlation and regression

- ⇒ Choose **Stats** ⇒ **Summary Statistics** ⇒ **Correlations** and complete the dialogue as shown in **Fig. 4.2a** to give the correlation between *uptake* and *conc*. You should find a value of 0.9836.

Fig. 4.2a Stats ⇒ Summary ⇒ Correlations



- ⇒ Choose **Stats** ⇒ **Regression Analysis** ⇒ **Linear models**, **Fig. 4.2b**, and complete the dialogue as shown in **Fig. 4.2c**.

Fig. 4.2b Regression dialogue

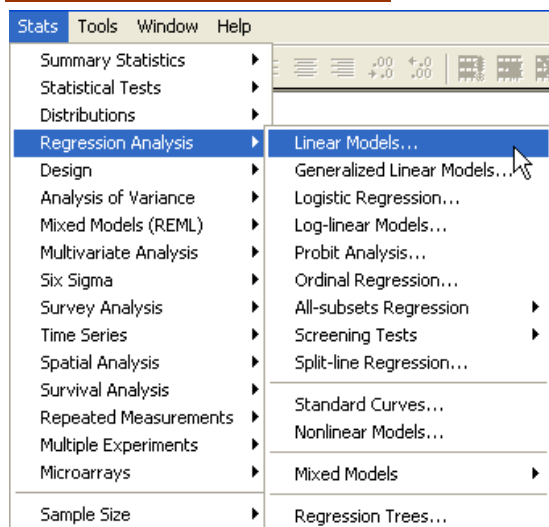
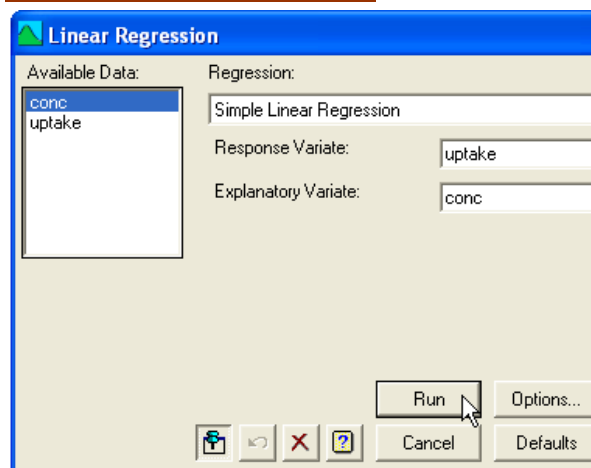


Fig. 4.2c Specify the y and x



⇒ Choose **Options...** and tick the box labelled 'confidence intervals'.

⇒ Press **Ok** and then **Run**.

Look at the table of regression coefficients shown at the bottom of the *Output Window* (Fig. 4.2d). The estimate of **-2.043** for the **Constant** is the *intercept* of the line. The estimate of **0.02494** for **conc** is the *slope* of the line fitted to the data when **uptake** is plotted against **conc**.

Thus these results show that the fitted equation is:

$$\text{uptake} = -2.043 + 0.02494 * \text{conc}$$

Fig. 4.2d Table of regression coefficients

Estimates of parameters

Parameter	estimate	s.e.	t(15)	t pr.	lower 95%	upper 95%
Constant	-2.043	0.198	-10.32	<.001	-2.465	-1.621
conc	0.02494	0.00118	21.11	<.001	0.02243	0.02746

⇒ Return to the regression dialogue to give a plot of the fitted line. Click on **[Further Output]** then **[Fitted Model]** see Fig. 4.2e, to give the plot shown in Fig. 4.2f.

⇒ Return to the dialogue, click on **[Further Output]** again and select **[Model Checking]**. Accept all the defaults by pressing **[Run]**. In the Graphics Window, 4 plots will be shown as in Fig. 4.2g.

A teaching session on Regression Analysis, which includes procedures for checking model assumptions, will help you understand this output.

The key point is that the examples in chapters 3 and 4 have shown it is easy to “do statistics” once you have become familiar with the use of dialogues in GenStat.

This allows training courses to concentrate on statistical concepts since the computing part has become easy.

Fig. 4.2e Further output sub-dialogue

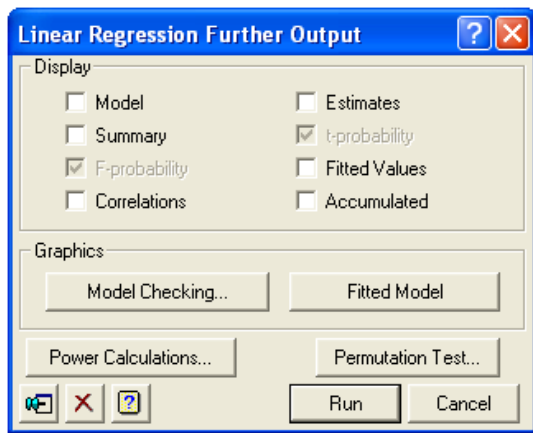


Fig. 4.2f Fitted model

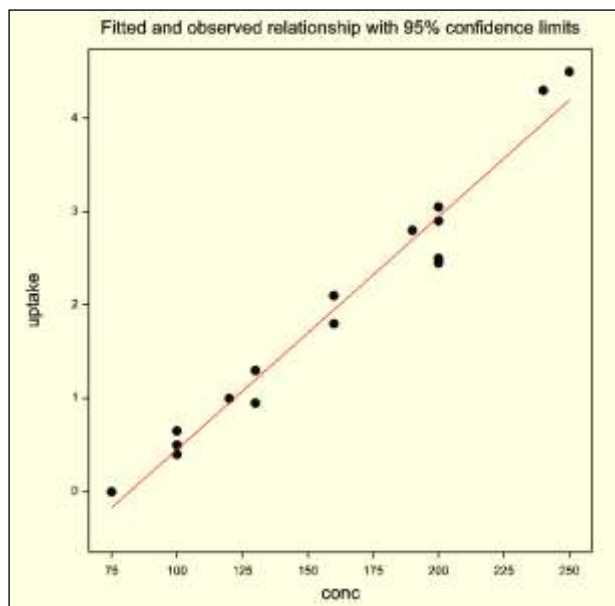
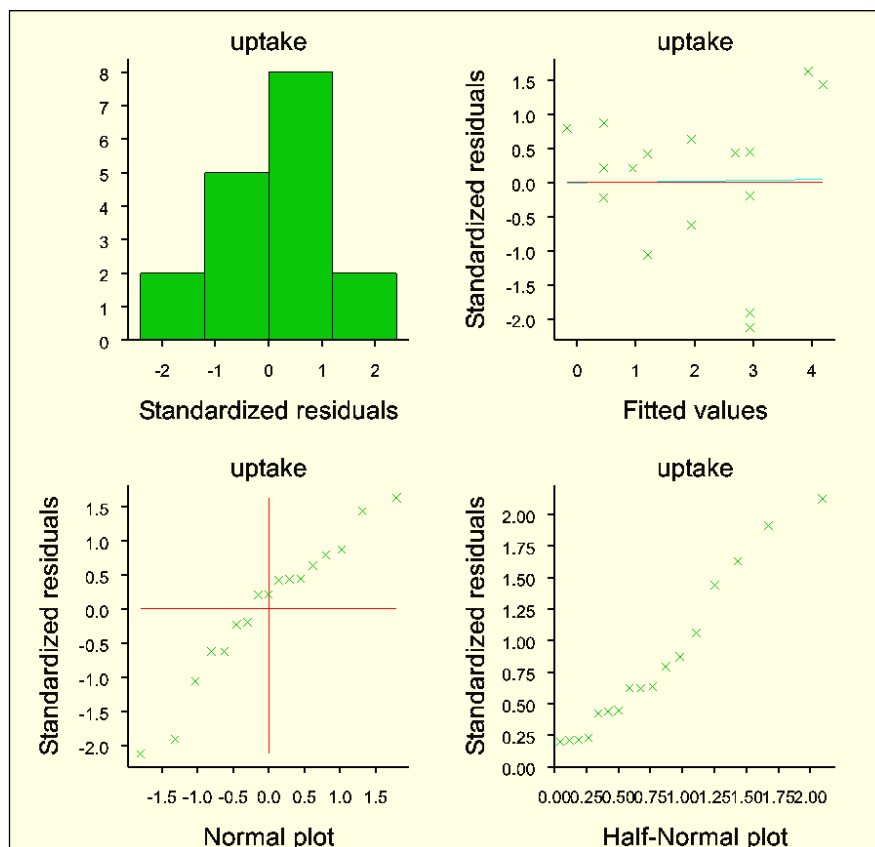


Fig. 4.2g Model checking



4.3 A GenStat tutorial

GenStat includes its own tutorials as part of the software. Use **Help** ⇒ **Tutorials** and try the one called **Linear Regression**. If you find it helpful then try some of the other tutorials.