INTRODUCTION TO GENSTAT FOR WINDOWS

The University of Reading

Statistical Services Centre
1. GenStat basics
2. Data Input and Manipulation
   2.1 Starting GenStat 7th Edition
   2.2 Data input
   2.3 Some basic data manipulation
   2.4 Factor Columns
   2.5 Understanding how GenStat works
3. Simple statistical inference
   3.1 The use of boxplots
   3.2 Comparisons of means
   3.3 References
4. Simple regression
   4.1 Setting up the data
   4.2 Correlation and regression
   4.3 A GenStat tutorial
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 7th Edition**

You start GenStat within Windows on a PC by clicking on the GenStat icon on the desktop or toolbar or by selecting “GenStat executable”, from the Programs Menu. If no GenStat icon is available on the desktop, you can create one yourself.

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.

---

1 By default, GenStat is installed in the folder C:\Program Files\Gen7Ed. Use Windows Explorer and go to the subfolder Gen7Ed\bin. Right click with the mouse on Genwind7.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 7th Edition, to avoid confusion with previous versions.
Fig. 2.1a GenStat Windows

Fig. 2.1b shows an example of the interface after a spreadsheet has been opened.

Fig. 2.1b Windows ⇒ Tile Vertical
2.2 Data input

2.2.1 Data input using the Spread Menu

⇒ We show two ways to enter data into GenStat. The first is within GenStat. Choose Spread ⇒ New ⇒ Create.

Fig. 2.2a Spread ⇒ New ⇒ Create

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.

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].
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, see 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

⇒ Use File ⇒ Save As and save the file as cmtut1.gsh, as shown in 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’.

![Run ⇒ Restart]

![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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>total</td>
<td>raindays</td>
</tr>
<tr>
<td>2</td>
<td>1330</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>2094</td>
<td>161</td>
</tr>
<tr>
<td>4</td>
<td>1851</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>1470</td>
<td>142</td>
</tr>
<tr>
<td>6</td>
<td>1557</td>
<td>152</td>
</tr>
<tr>
<td>7</td>
<td>1932</td>
<td>155</td>
</tr>
<tr>
<td>8</td>
<td>1184</td>
<td>126</td>
</tr>
<tr>
<td>9</td>
<td>2452</td>
<td>183</td>
</tr>
<tr>
<td>10</td>
<td>1347</td>
<td>132</td>
</tr>
<tr>
<td>11</td>
<td>1792</td>
<td>167</td>
</tr>
<tr>
<td>12</td>
<td>1486</td>
<td>149</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⇒ Save your Excel workbook and give it the name cm tut1.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

The file contains worksheets (marked S) named ranges (marked R) in the list opposite. Select the one that contains the data you wish to import.
In the next window, you can select which worksheet of the workbook you want to import. In this case just click 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 Insert ⇒ Name ⇒ Define, see Fig. 2.2m.

Give the range a name, for instance Data, as shown in Fig. 2.2n. Then save the Excel file and minimize Excel.
Go back to GenStat and restart the session by selecting Run ⇒ Restart Session and then clicking [Restart], to clear all windows, dialogue boxes and the spreadsheet. When you now reopen the file cmntut1.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.

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, as will be seen in Chapter 18, 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 and the data are entered into a GenStat spreadsheet.

⇒
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 Chapter 18.3 (page Error! Bookmark not defined.) where we show how to establish an ODBC link.

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 spreadsheet 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 for these two variates?

⇒ 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 [OK].

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. 2.2c. There are other statistics available within the same dialogue box.
Find the **Summarize Contents of Variates** dialogue again. Click on the [Clear] button to clear all currently selected statistics. Reselect the variables and choose **Arithmetic Mean**, **Standard Deviation** and **Standard Error of Mean**, and click [OK].

Use **Graphics ⇒ Point Plot** and complete the dialogue box as shown in *Fig. 2.3e*. Click [Finish] to produce the graph shown in *Fig. 2.3f*. 
The relationship between the total rainfall and the number of rain days is as follows:

![Graph showing the relationship between total rainfall and number of rain days](image)

⇒ Close the graph window with **File ⇒ Exit** (choosing **[No]** to the question about saving the graph), and then close the point plot and summary dialogue boxes by choosing **[Cancel]**.

Many dialogue boxes in GenStat do not close when you click **[OK]**. They only close if you click on **[Cancel]**. This is so you can easily repeat an operation, or get more output from the current analysis without having to go back through the menus. It is quite easy to get a large number of windows and dialogue boxes open at once, so it can be quite hard to find the one for which you are looking. Clicking the **[Reload]** or **[Close]** 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 ⇒ Point Plot**) as this will bring back the box complete with anything that had been entered. It is a good idea to close a box by clicking **[Cancel]** as soon as it is no longer needed.
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 rainday 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

Fig. 2.3h  Choosing the calculate dialogue

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

⇒ 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.
Type the name of the new column into the bottom box labelled **Save Result In**, Fig. 2.3i. Then click [OK]. 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**.

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. 2.3f Commands keep a record of your work

In the spreadsheet, each value of `meanperday` is displayed with the same number of significant digits which may lead to a variable number of decimal places.

⇒ 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 may also wish to enter a concise explanation of the contents of the column in the Description box. Now, whenever `meanperday` is printed in the output, it will be displayed with 1 decimal place by default. Click [OK] to effect the change.
Assume that these data values came from 11 years in order. It would be useful to have this information entered too.

⇒ 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 *), **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.
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*.

Try plotting the mean per day against the year as a line graph. Use **Graphics** ⇒ **Line Plot** ⇒ **SingleXY type** with *meanperday* as the *Y* and *year* as the *X*.

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** at the top 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 this data set were El Nino years, the second, third, eighth and tenth. The remainder 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. Type 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](image)

⇒ 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.

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

![Fig. 2.4b Add labels to the factor](image)

![Fig. 2.4c Entry of data](image)
Now type the following values into the new column, as shown in Fig. 2.4c.

O E E 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 ⇒ Point Plot. Fill in the boxes as in Fig. 2.4d, and click [Finish] to produce the graph shown in Fig. 2.4e. If you first click [Next], you can add titles to the graph and the axes.
Fig. 2.4e Resulting graph

By using the Edit ⇒ Edit Graph once you have the graph, or right clicking in the graph, you can choose to edit the Axes key or graph symbols. If you wish, you could try producing the layout 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. 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 two [OK] buttons 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:

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

Fig. 2.4g  Stats ⇒ Summary ⇒ Tabulation

Fig. 2.4h  Results

Ξ  Complete the dialogue as shown and press [OK]. The results are shown in Fig. 2.3h in the Output Window.

2.4.2 Saving data

Before continuing, save the spreadsheet.

Ξ  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. We had imported the file \textit{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.

\textbf{Fig. 2.4j GenStat warning that the original file will be replaced}

![GenStat warning](image)

\subsection*{2.4.3 Deleting data}

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

\begin{itemize}
  \item First select the column \textit{meanperday}. Then click in the name field (or press $<$ALT$>$<$C$>, or use \textit{Spread $\Rightarrow$ Select $\Rightarrow$ Current Column}). Clicking again will deselect the column. Practice selecting and deselecting columns. Finishing with the \textit{meanperday} column selected.
  \item 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 \textit{Edit $\Rightarrow$ Undo} to get the column back.
  \item What you need to do is to delete the whole column. The column should still be selected. Use \textit{Spread $\Rightarrow$ Delete $\Rightarrow$ Current Column}. You can also select one, or more, rows and delete them in the same way.
\end{itemize}
2.4.4 Available variables

⇒ You can check which variables are currently available to the GenStat server using Data ⇒ Display or pressing the F5 key, see Fig. 2.4k and Fig. 2.4l.

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.

⇒ Click [Close] to close the Display dialogue box.
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 7th 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 the Input 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

![Image of New window]

 ⇒ Now select the Run menu as shown in Fig. 2.5c. You can choose either Submit Line (if the cursor is still on the line you typed) or Submit Window. Choose one of these.
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**.

Then type $3 + 4$ as the function, click on **Print in Output** and then on [OK]. If you look in the Output window, you see that $3 + 4$ still equals 7!
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 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 Fault 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:

\[
100 \times \frac{(4.35 - 2.37)}{4.35}
\]

Click [OK]. This will give the following in the output window:

\[
\frac{100 \times (4.35 - 2.37)}{4.35} \approx 45.52
\]

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 \texttt{SQRT()}, where the number whose square root is required is given in the parenthesis, for example \texttt{SQRT(12.37)}. \textit{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 ‘List of functions for expressions’.
⇒ 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

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Operation</th>
<th>Example</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>addition</td>
<td>PRINT 3+4</td>
<td>7.000</td>
</tr>
<tr>
<td>-</td>
<td>subtraction</td>
<td>PRINT 3-4</td>
<td>-1.000</td>
</tr>
<tr>
<td>*</td>
<td>product</td>
<td>PRINT 3*4</td>
<td>12.00</td>
</tr>
<tr>
<td>/</td>
<td>division</td>
<td>PRINT 3/4</td>
<td>0.7500</td>
</tr>
<tr>
<td>**</td>
<td>exponentiation</td>
<td>PRINT 3**4</td>
<td>81.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Operation</th>
<th>Example</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQRT(x)</td>
<td>Square root</td>
<td>PRINT SQRT(4)</td>
<td>2.00</td>
</tr>
<tr>
<td>EXP(x)</td>
<td>Exponential function</td>
<td>PRINT EXP(1)</td>
<td>2.718</td>
</tr>
<tr>
<td>LOG(x)</td>
<td>natural logarithm of x, for x &gt; 0</td>
<td>PRINT LOG(2.718)</td>
<td>0.9999</td>
</tr>
<tr>
<td>LOG10(x)</td>
<td>logarithm to base 10 of x, for x &gt; 0.</td>
<td>PRINT LOG10(10)</td>
<td>1.000</td>
</tr>
<tr>
<td>ROUND(x)</td>
<td>rounds the values of x to the nearest integer.</td>
<td>PRINT ROUND(1.2345678)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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 (X) 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 simple 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 standard information. 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, 2.1, 1.7, 2.2

3.1 The use of boxplots

⇒ Because these columns are of different lengths, they are entered into two separate spreadsheets. For the first set, use Spread ⇒ New ⇒ Create as shown earlier in Chapter 2.2.1. Set it to have 1 column of 6 rows, enter the data as shown in Fig. 3.1a and give the column the name new.

⇒ Save the spreadsheet, giving it the name cmtut2.gsh (see Chapter 2.3.4) if you need instructions on saving. Then use Spread ⇒ New ⇒ Create again. Change the number of rows to 10 and enter the second set of data into this other spreadsheet, naming the column as standard, see Fig. 3.1b. Save the spreadsheet, giving it the name cmtut3.gsh.

Fig. 3.1a 1st sample  Fig. 3.1b 2nd sample  Fig. 3.1c Graphics ⇒ Boxplot

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

⇒ Go back to the spreadsheet and insert a value of 2.9 instead of 2.0 for the 8th value in the Standard group. 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. 3.1f and Schematic, Fig. 3.1g. Try both, as shown below. The advantage of a schematic boxplot is that you can easily discover outliers.
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. Use the Options button in Fig. 3.2a to give the dialogue in Fig. 3.2b.

The output window shows the results, see Fig. 3.2c.
### 3.2.1 Some more data manipulation: appending spreadsheets

In the 2-sample example that was used for the t-test, the data were put into separate spreadsheets. Data often need reorganising before analysis and here this step is illustrated by joining the data together for the two sets. Fig. 3.2f shows what we are aiming for.
What we wish to do is to append the data from the two columns and add a further column, that specifies from which set each observation has come.

⇒ If the spreadsheets are no longer in GenStat then they will have to be opened. They were saved earlier with the names \texttt{cmtut2.gsh} and \texttt{cmtut3.gsh}, see \textit{Fig. 3.2d} and \textit{3.2e}.

⇒ Click in the shorter spreadsheet \texttt{cmtut2.gsh}, so it is the active window. Use \texttt{Spread} \Rightarrow \texttt{Manipulate} \Rightarrow \texttt{Append} and complete the dialogue as shown in \textit{Fig. 3.2g}. This appends \texttt{cmtut3.gsh} to the data in \texttt{cmtut2.gsh} and adds the information for a factor that distinguishes between the two groups. Press [OK].

\textit{Fig. 3.2g} \texttt{Spread} \Rightarrow \texttt{Manipulate} \Rightarrow \texttt{Append} (with cmtut2.gsh as the active window)

The layout of the data shown in \textit{Fig. 3.2f} is more common and is used in most of the remainder of this guide.

⇒ Rename the column \texttt{new} in the long spreadsheet to \textit{yield}.

⇒ Use \texttt{File} \Rightarrow \texttt{Save As} to save the spreadsheet, giving it the name \texttt{cmtut4.gsh}.

3.3 References

4. Simple regression

4.1 Setting up the data

We now introduce some key elements of data analysis, by means of simple 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

⇒ 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

⇒ 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 summary statistics for each of the two columns.

⇒ Choose **Stats ⇒ Summary Statistics ⇒ Summarize Contents of Variates** and specify some summary statistics as described in Section 2.3.1.

---

**Fig. 4.1e Regression data**

<table>
<thead>
<tr>
<th>Row</th>
<th>Conc</th>
<th>Uptake</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>0.65</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>130</td>
<td>0.95</td>
</tr>
<tr>
<td>7</td>
<td>130</td>
<td>1.30</td>
</tr>
<tr>
<td>8</td>
<td>160</td>
<td>1.80</td>
</tr>
<tr>
<td>9</td>
<td>160</td>
<td>1.80</td>
</tr>
<tr>
<td>10</td>
<td>160</td>
<td>2.10</td>
</tr>
<tr>
<td>11</td>
<td>190</td>
<td>2.80</td>
</tr>
<tr>
<td>12</td>
<td>200</td>
<td>2.50</td>
</tr>
<tr>
<td>13</td>
<td>200</td>
<td>2.90</td>
</tr>
<tr>
<td>14</td>
<td>200</td>
<td>2.45</td>
</tr>
<tr>
<td>15</td>
<td>200</td>
<td>3.05</td>
</tr>
<tr>
<td>16</td>
<td>240</td>
<td>4.30</td>
</tr>
<tr>
<td>17</td>
<td>250</td>
<td>4.50</td>
</tr>
</tbody>
</table>

---

**Fig. 4.1f Produce some descriptive statistics**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary Statistics</strong></td>
<td><strong>Summarize Contents of Variates...</strong></td>
<td><strong>Statistical Tests</strong></td>
<td><strong>Stem and Leaf...</strong></td>
<td><strong>Design</strong></td>
<td><strong>Frequency Tables...</strong></td>
<td><strong>Summarize Groups (Tabulation)...</strong></td>
</tr>
<tr>
<td><strong>Distributions</strong></td>
<td><strong>Italy...</strong></td>
<td><strong>Analysis of Variance...</strong></td>
<td><strong>Summarize Circular Data...</strong></td>
<td><strong>Mixed Models (REML)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td><strong>Multivariate Analysis</strong></td>
<td></td>
<td><strong>Correlations...</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td></td>
<td><strong>Six Sigma</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analysis of Variance...</strong></td>
<td></td>
<td><strong>Survey Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mixed Models (REML)</strong></td>
<td></td>
<td><strong>Time Series</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Spatial Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Survival Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Repeated Measurements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Multiple Experiments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⇒ Choose **Graphics ⇒ Point Plot** and complete the dialogue as shown in **Fig. 4.1g** to give the scatterplot.
4.2 Correlation and regression

Choose \textbf{Stats} \Rightarrow \textbf{Summary Statistics} \Rightarrow \textbf{Correlations} and complete the dialogue as shown in \textbf{Fig. 4.2a} to give the correlation between \textit{uptake} and \textit{conc}. You should find a value of 0.984.

Choose \textbf{Stats} \Rightarrow \textbf{Regression Analysis} \Rightarrow \textbf{Linear models}, \textbf{Fig. 4.2b}, and complete the dialogue as shown in \textbf{Fig. 4.2c}. 

\textbf{Fig. 4.1g} Graphics \Rightarrow \textbf{Point plot}

\textbf{2D Scatter Plot - Data}

Type of plot: \textbf{Single \(XY\)}

Select the data to be plotted (enter name and position here).

Select \(Y\): \textbf{uptake}

Select \(X\): \textbf{conc}

\textbf{Data currently selected for plotting:}

\textbf{Y Data:} \textbf{uptake}

\textbf{X Data:} \textbf{conc}

\textbf{Fig. 4.1h} Results

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{scatter_plot.png}
\caption{Scatter plot of uptake vs. conc.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{correlation_dialogue.png}
\caption{Correlation dialogue in Genstat.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{regression_dialogue.png}
\caption{Regression analysis dialogue in Genstat.}
\end{figure}
The results are in the output window. They show the fitted equation is:

\[
\text{uptake} = -2.043 + 0.02494 \times \text{conc}
\]

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

Repeat the steps to give [Further Output] again and select [Model Checking]. Accept all the defaults by pressing [OK]. In the Graphics Window, 4 plots will be shown as in Fig. 4.2f.
This example should 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. The computing has become easy.

### 4.3 A GenStat tutorial

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