Genstat®

Spreadsheet

www.vsni.co.uk

VSNi

A Guide to the Genstat® Spreadsheet

by David Baird and Darren Murray

Genstat is developed by VSN International Ltd, in collaboration with practising statisticians at Rothamsted and other organisations in Britain, Australia, New Zealand and The Netherlands.

Published by: VSN International, 2 Amberside, Wood Lane,
Hemel Hempstead, Hertfordshire HP2 4TP, UK
E-mail: info@genstat.co.uk
Website: http://www.genstat.co.uk/

First published 2009, for GenStat *for Windows* 12th Edition
This edition published 2015, for Genstat *for Windows* 18th Edition

Contents

Introduction

In this Guide we describe Genstat's extensive spreadsheet facilities for data entry, import, export and manipulation. Initially we show how you can load data in an Excel file into a Genstat spreadsheet. We then show how you can use the Genstat spreadsheet itself for data entry and verification. In an analysis you may sometimes want to work with subsets of your data, and we describe how these can conveniently be created using the spreadsheet. You may need to calculate or recode data from columns in the spreadsheet and we show you various ways of doing this. The data may not always be stored in a convenient form or may require rearranging before analysis. We demonstrate some of the relevant spreadsheet data manipulation techniques, such as appending, merging, stacking and unstacking data. Within Genstat you can have multiple spreadsheets contained together within a book where each spreadsheet is stored on a separate page. We show how you can store and manipulate spreadsheets in a book. We also describe how you can use Genstat's ODBC facilities to read and write data to databases. Finally we show you some server commands that allow you to use data in spreadsheets in your Genstat programs.

1 Reading Excel files

You can read data from external files into Genstat using the File menu on the Genstat menu bar. The menu covers a wide range of formats, including spreadsheets (like Excel) and databases. When Genstat reads a file in one of these formats, it automatically puts the data into a Genstat spreadsheet. Excel probably represents the most common "foreign" format. So we start this Guide by showing you how to use the Excel wizard to load an Excel file.

When reading data from a foreign file, Genstat expects the data to be in a rectangular column format. In a spreadsheet, such as Microsoft Excel, the data need to be arranged in a group of columns forming a rectangle where the columns are of the same length. If the rectangular area contains empty rows or columns then, by default, these will be removed when the data is opened in Genstat. You can specify column names for your data by entering a label for the name in the first row of the column within the rectangular block. A spreadsheet column name must start with a letter (A-Z, a-z or %) and can only contain letters, numbers or the symbols % and _. When data are read into Genstat, a check is made to see if a column name meets these conditions and modifies any names that include invalid characters. For example, if the first character of the column name is a number, then Genstat will create a new name by prefixing the label for the column with a %. When no column names are provided, Genstat will generate default column names using the notation C1, C2 etc... You can specify missing data values by either by leaving the cells blank or by entering an asterisk (*).

When the data columns are read into Genstat, any numerical columns will be imported as variates and any column containing labels (excluding the column name) will be imported as a text data structure. Within a Genstat spreadsheet a text column is marked by a green 'T' next to its column name and the contents are right justified by default. A column of numbers or text can also be read into Genstat as a factor. You can specify a column to be a factor by appending an exclamation mark (!) onto the column name (e.g. crop!).

	A	B	C
1	counts	crop!	
2	18	pea	
3	117	pea	
4	21	cereal	
5	7	pea	
6	176	cereal	
7	85	cereal	
8	244	cereal	
9	4	pea	
10	55	cereal	
11	8	pea	
12			

Figure 1.1

Figure 1.1 shows an example of a block of data within the `Genstat Data` worksheet of the Excel file `Bacteria.xls`, which has been arranged for input into Genstat. The data values are a set of counts from an experiment: the numbers of one particular type of bacteria found in small samples of soil growing two different types of crop. The second column contains categorical data and has had the symbol '!' appended to the column to specify the column is to be a factor.

We can import the data into Genstat using the File menu. The data files used in this Guide are stored in a directory (or folder) called `Data` that is alongside the folder that contains the Genstat executable program.

In this example we want to open the Excel file containing the data shown in Figure 1.1. To open the file we select the Open line in the File menu on the menu bar. This opens the Select Input file menu (Figure 1.2). This has all the standard controls provided by Windows in an "open-file" menu, and we can move to the `Data` folder in the usual way. Once we have found the folder, it is advantageous to click on Set as in the Working Directory drop-down list. The Select Input file menu will then automatically open in this folder when we use it in the future. We now select Other Spreadsheet Files from the drop-down list in the bottom-right corner so that we can see what Excel files are available.

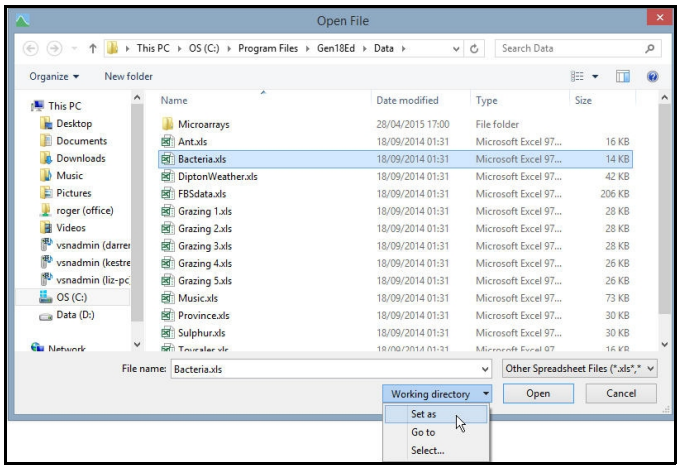


Figure 1.2

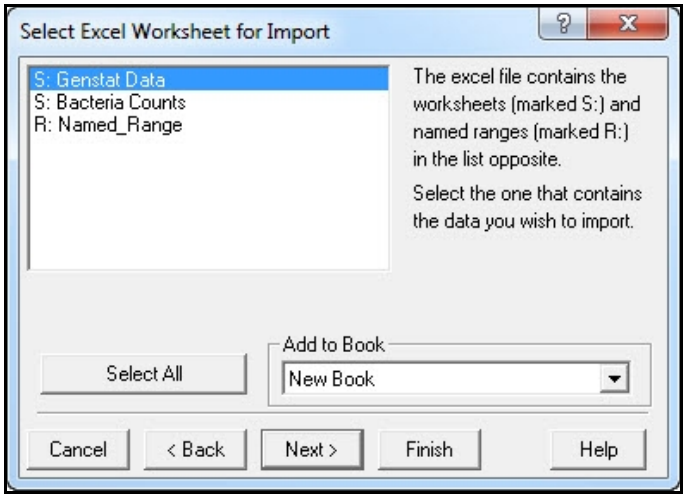


Figure 1.3

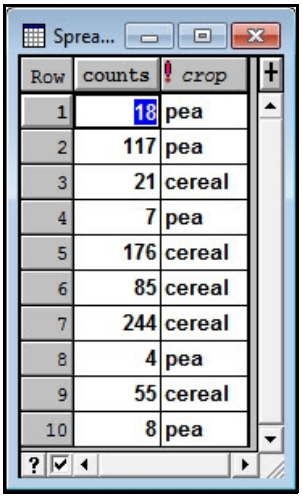


Figure 1.4

Selecting the file `Bacteria.xls` and clicking on Open, or double-clicking on the filename, gives the menu shown in Figure 1.3. This is the initial menu of a wizard for the input of data from an Excel file. It lists all the available worksheets and named ranges within the Excel file, with worksheet names prefixed by 'S:' and named ranges by 'R:'. In this example, we have selected the worksheet Genstat Data. We have no other books or spreadsheets open within Genstat, so the Add to Book drop-down list is left as New Book. We will explain how to form books of spreadsheets in Chapter 7. Until then, we will keep our spreadsheets separately.

Subsequent menus allow you to select ranges and columns, and set various other options controlling how the data are transferred to Genstat. In this case we want to take all the data on the page, and will leave the other options with their default settings. (The subsequent menus will be shown later though; see Figures 1.7, 1.8 and 1.9.) So we click on Finish to open the two columns of data into a Genstat spreadsheet, as shown in Figure 1.4.

If you click on the Output window, the data in the spreadsheet are automatically

transferred to Genstat's central data pool. The Output window displays a brief summary of the data that have been transferred, as shown below.

```
Data imported from Excel file: C:\Program Files\Gen18Ed\Data\Bacteria.xls
 on: 13-May-2015 14:04:39
 taken from sheet "Genstat Data", cells A2:B11

  Identifier     Minimum        Mean     Maximum      Values     Missing
      counts       4.000       73.50       244.0          10           0

  Identifier      Values     Missing      Levels
        crop          10           0           2
```

In fact, whenever you change from the spreadsheet window to another window, Genstat will update the central pool with any changes that you have made in the spreadsheet. You can verify that the data have arrived, by looking in the Data View pane (Figure 1.5).

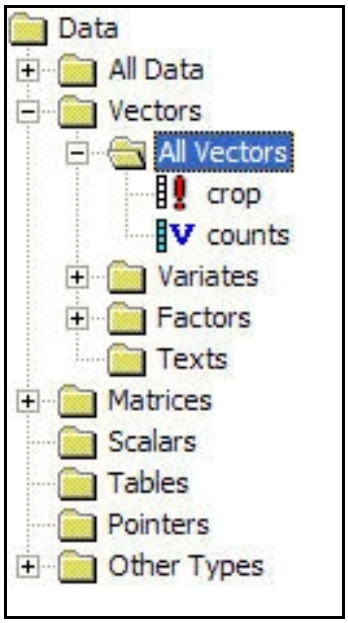


Figure 1.5

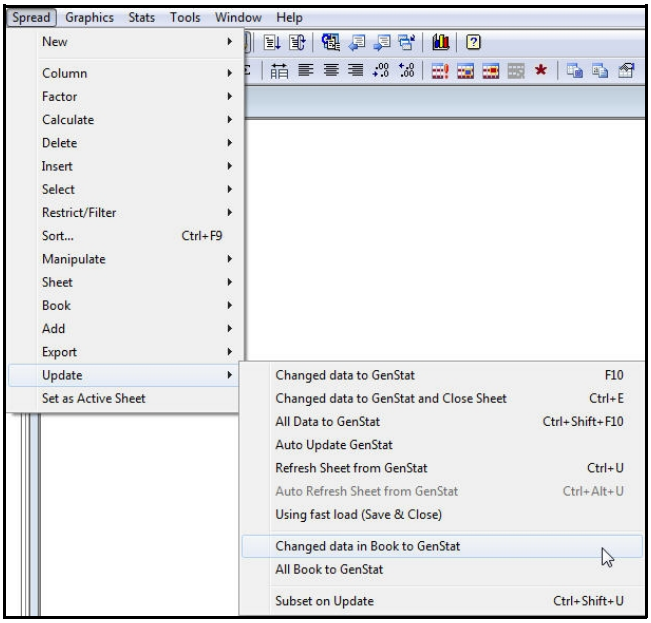


Figure 1.6

You can transfer the data explicitly, using an option from the Update submenu of the Spread menu (Figure 1.6). For example, selecting the Change data to Genstat and Close Sheet item updates Genstat and then closes the spreadsheet. The standard method of updating the pool uses the Genstat `READ` command. The Using fast load (Save & Close) item provides a more efficient alternative, using the `SPLOAD` command, for large spreadsheets in Genstat's own gsh format.

We shall now close the Genstat spreadsheet, and input some data from the other Excel worksheet.

Data are not always stored in a singular rectangular format within a

Named_Range fx counts2

	A	B	C	D	E
1	**Counts of bacteria and crop type**				
2					
3	**counts1**	**crop1**		**counts2**	**crop2**
4	18	pea		32	cereal
5	117	pea		45	pea
6	21	cereal		65	cereal
7	7	pea		76	pea
8	176	cereal		87	cereal
9	85	cereal		7	cereal
10	244	cereal		311	pea
11	4	pea		275	pea
12	55	cereal		78	pea
13	8	pea		4	cereal

Figure 1.7

spreadsheet, but may have multiple blocks of data entered on a single worksheet. Figure 1.7 shows an example of this in the worksheet Bacteria Counts from the file Bacteria.xls. In this worksheet there is a title in row 1 of column A, and two rectangular sets of data records. In this example we just want to open the second rectangle of data (`counts2` and `crop2`) within a spreadsheet.

We again use the File menu, and select the file `Bacteria.xls` as shown in Figure 1.2. So we again reach the initial menu of the Excel wizard, as shown earlier in Figure 1.3.

There are two ways of reading a rectangular range of data from Excel into Genstat. If we select the worksheet Bacteria Counts in Figure 1.3 and click on the Next button (instead of Finish), the second menu in the wizard allows the range to be specified explicitly. You check the Specified Range radio button (instead of the default All cells), and enter the range D3:E13 into the adjacent field as shown in Figure 1.8.

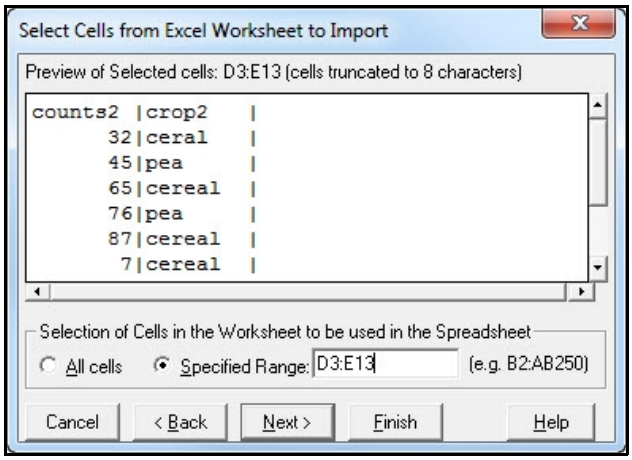


Figure 1.8

Alternatively, you can create a named range for the rectangular block of data within Excel and select this from the worksheet list in Figure 1.3. To create a named range in Excel, you first select the desired rectangle either with the mouse or by using the shift and cursor keys. Once the rectangle has been selected, you can name the range by clicking in the Name Box and typing its name. In Figure 1.7 we have selected the range D3 to E13 and entered its name as Named_Range in the Name Box. If you select Named_Range as the worksheet or range in Figure 1.3 and again click Next, you will see that the range D3 to E13 is set up automatically in the second menu of the wizard, just as in Figure 1.8.

The third menu in the wizard (Figure 1.9) allows you to choose which of the columns in the worksheet or range to read. By default they are all read.

The final part of the wizard, shown in Figure 1.10, is a menu with tabs controlling more advanced aspects. This time we have not put an exclamation mark at the end of the column name to specify that the column `crop2` is to be a factor.

So, instead, we select the Factors tab, and check the Suggest columns with only a few unique values to be Factors box. If this option is set, Genstat will check all the columns for repeated values or labels and, if any are detected, you will be prompted with a menu offering you the choice to convert them.

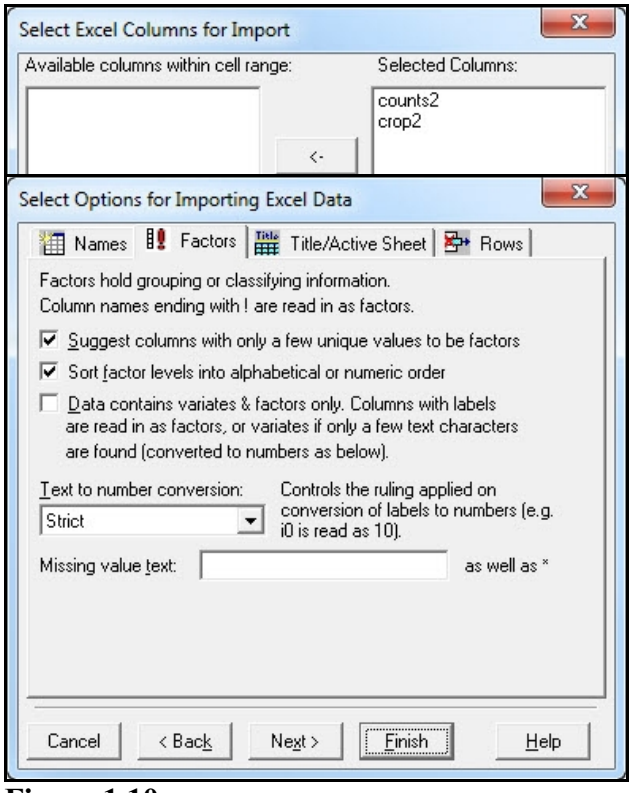


Figure 1.10

On clicking Finish Genstat detects that the column crop2 has repeated labels and displays the menu shown in Figure 1.11. This menu displays all the columns that have repeating values and the current data type for each column is indicated by a prefix to the name (T specifies a text, F a factor and V a variate). To change the type of crop2 from a text to a factor we double-click on the name crop2 in the list (alternatively you can click on the button labelled Factor). This changes the prefix from T to F specifying the column will be a factor. Clicking on OK forms another Genstat spreadsheet and, if we now click on the Output window, the data in the spreadsheet are transferred to Genstat's central data pool as shown below.

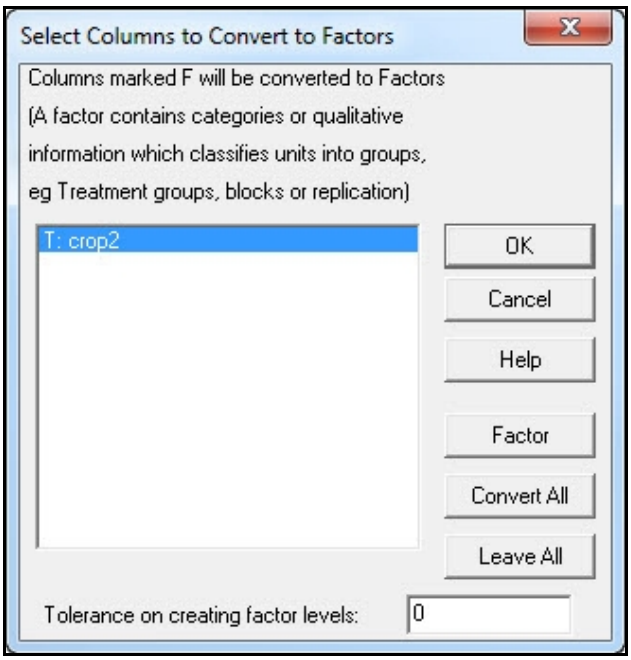


Figure 1.11

```
Data imported from Excel file: C:\Program Files\Gen18Ed\Data\Bacteria.xls
on: 13-May-2015 14:09:26
 taken from sheet ""Bacteria Counts"", cells D3:E13

Identifier    Minimum       Mean    Maximum     Values    Missing
  counts2      4.000      98.00      311.0         10          0

Identifier    Minimum       Mean    Maximum     Values    Missing
  crop2                                             10          0
```

An alternative way to input data is to use the facilities within the Spread menu. In this example we will copy the columns count1 and crop1 from the file Bacteria.xls (see Figure 1.7) via the clipboard into a Genstat spreadsheet. As with the layout within a spreadsheet, Genstat expects the data on the clipboard to be in a rectangular format with columns of equal lengths. In Excel we select the data, including the column names (data range A3:B13), and then select Copy from the Edit menu. Note that when you are using Excel, if you do any other operation on the spreadsheet before going to Genstat, Excel clears the data from the clipboard. The data are available to Genstat only while the dotted lines are moving

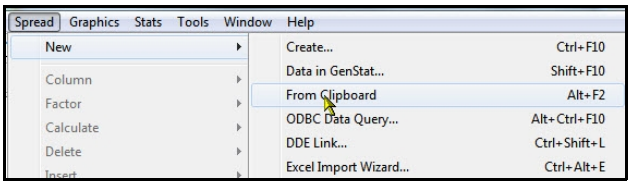


Figure 1.12

around the selected cells in Excel. Now, in Genstat, we create a spreadsheet of the data, by selecting the From Clipboard item from the New option on the Spread menu as shown in Figure 1.12. The New Spreadsheet from Clipboard menu (Figure 1.13) is then produced to control the process. We leave the Suggest columns to be factors box checked (and again ask to open a New Book). Genstat displays the factor conversion menu again. This time it will show `crop1` as the column with repeated values rather than crop2, as in Figure 1.11. Leaving `crop1` as a text and clicking OK produces the spreadsheet shown in Figure 1.14.

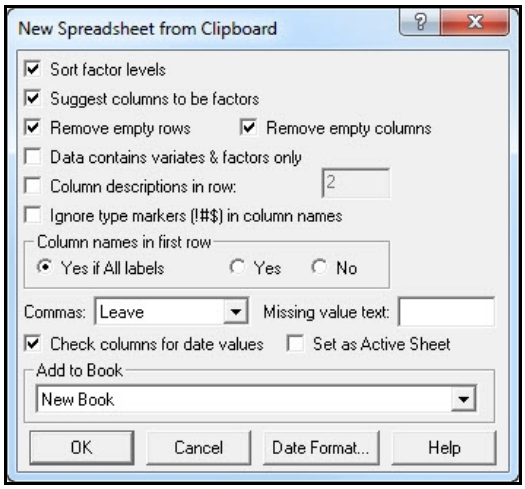


Figure 1.13

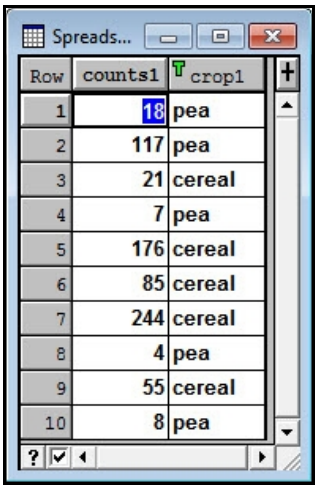


Figure 1.14

If we change our mind and want to reconsider converting `crop1` to a factor, it is not too late as this can be done very easily using the Spread menu. Put the cursor into one of the cells in the `crop1` column, and click on the Convert to sub-option of the Factor option of the Spread menu, as shown in Figure 1.15.

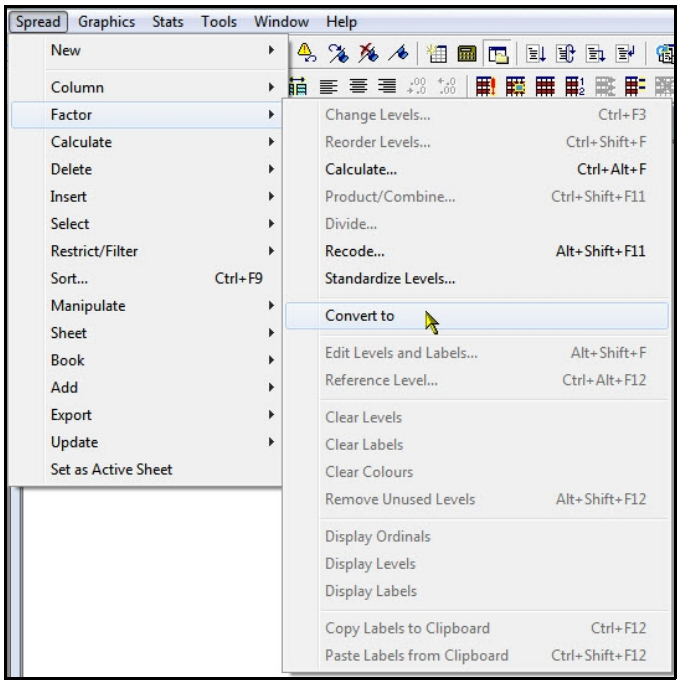


Figure 1.15

1.1 Exercise

The file `Traffic.xls` is an Excel data file with one worksheet called counts storing one set of data in the area B3:D43. Using the Open option of the File menu, load the data into a Genstat spreadsheet, converting `day` and `month` to factors.

1.2 Reading data from multiple sheets or files

Sometimes data on separate categories or trials will have been entered on separate pages with an Excel file or in separate files. Assuming that these sheets have columns which match in type (i.e. matching columns are both text) and position or name, then these sections of data can be combined into a single spreadsheet with a factor created to indicate the sheet or file that they have come from. If all the data are in one Excel file, then use the Append Multiple Excel Worksheets option on the Spread | New menu. This will open the File Open dialog, and use this to select the Excel file with the data you want to read in.

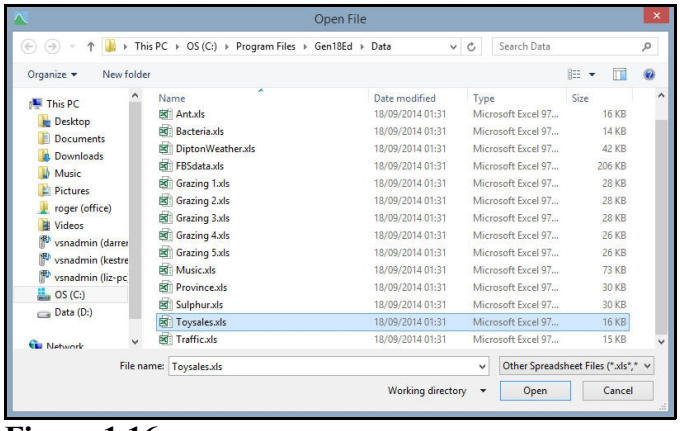


Figure 1.16

In this example we have chosen the file `Toysales.xls` in the Genstat `Data` directory. These data are described more fully in Section 3.5.

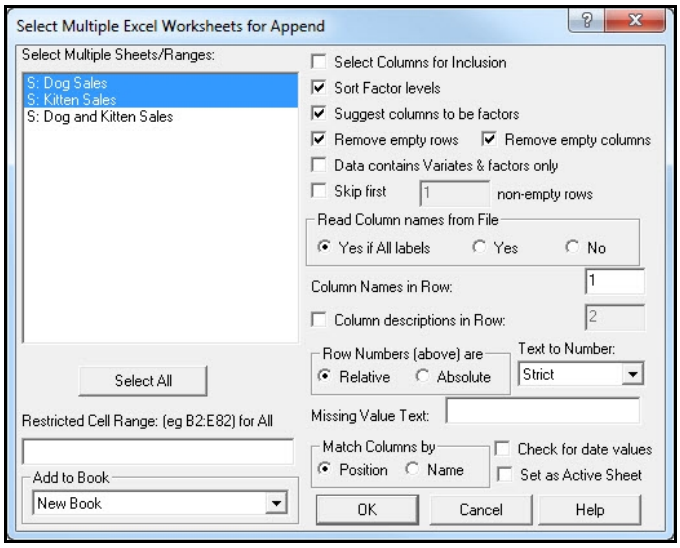


Figure 1.17

After you have selected the file, you will get the dialog in Figure 1.18, which allows you to specify how the data is read from each page. The same options are used for reading in each page selected in the file. You need to select the sheets to be appended, either by clicking on each one individually while holding down the control or shift key, or else by using the Select All button if all sheets are to be appended. Here we just select the first two sheets Dog Sales and Kitten Sales as the third sheet holds the combined information in a different format. The most important information to provide to this dialog is how the columns are to be matched between the two sheets. The Match Columns by box has two options: Position or Name. If the columns are matched by position, then column 1 from the second sheet will be appended to column 1 from the first sheet, and the second sheets column names will be ignored. If they are appended by name, then the column names between the two sheets must match, but they need not be in the same order. If there is not a matching column with the same name in one of the sheets then missing values will be inserted. Of course if the columns have the same names and are in the same order, then either of these options will give the same results. Once you click OK you get a dialog which detects that some columns look like factors as they have a few unique values that repeat (as below). If you just click OK, then this will close without making any changes, or you could double-click an entry to make it a factor. The resulting spreadsheet is shown to the right.

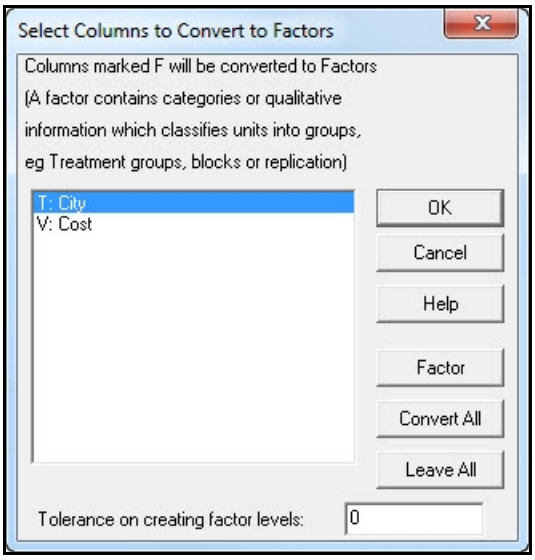


Figure 1.18

The resulting spreadsheet contains the factor column Worksheet that gives the name of the worksheet that each row came from (Figure 1.19).

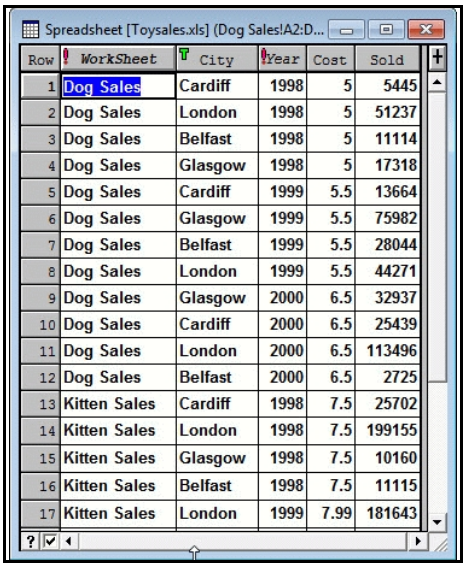


Figure 1.19

1.3 Appending data from multiple files

If the data you wish to append are in multiple files or on pages in Genstat Book (GWB) file, then these can be appended using the Append Multiple files option in the Spread | New menu (Figure 1.20). This opens the dialog in Figure 1.21, which allows you to select the files and specify how the columns are matched. If you specify a file type that has multiple pages (a Genstat Book or Excel spreadsheet), then you will be prompted on which pages are to be appended, and if it is a foreign file format you will get the usual options dialogs for each file that is appended coming up.

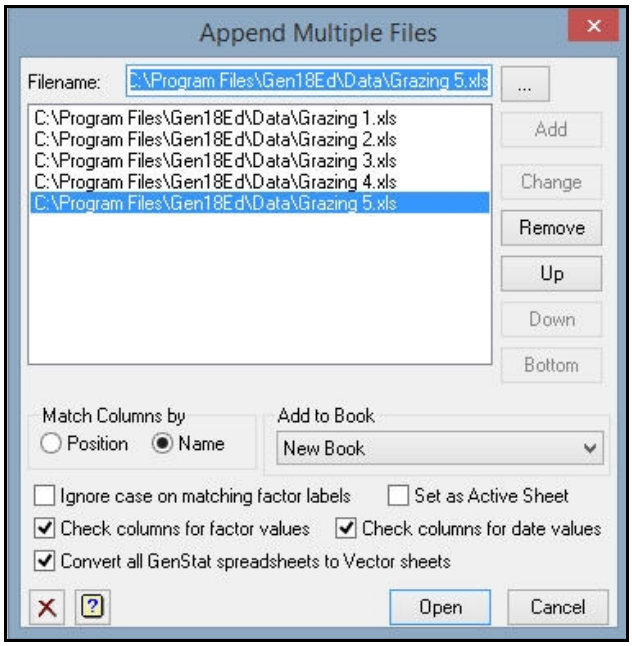


Figure 1.20

In this example, there are 5 Excel files, `Grazing 1.xls` - `Grazing 5.xls` in the Genstat `Data` directory that contain the initial and final live-weights of calves that were put in 12 paddocks to graze for two weeks. The files give the results from 5 grazing periods over the summer. The paddocks contain 3 types of herbage (DE, TE and TH). The columns in the files all have the same names and order, so it doesn't matter whether we chose to match columns by positions or names. Using the browse button (which has the three dots on it), this opens the File Open dialog. Use the mouse with the control or shift key to select the five files as shown in the dialog above. When you click the Open button, these files will be added to the filename list. You can use the browse button multiple times if the files are in different directories. The order of the files defines the order that the data appears in the resulting spreadsheet. If the order of the filenames is not what you want in the spreadsheet, you can reorder the files in the filename list using the Up, Down and Bottom buttons which move the filename which has currently been selected. Now clicking the Open button will open each of the selected files. As they are Excel files, you will be prompted for the options for reading in an Excel file fives times, once for each file.

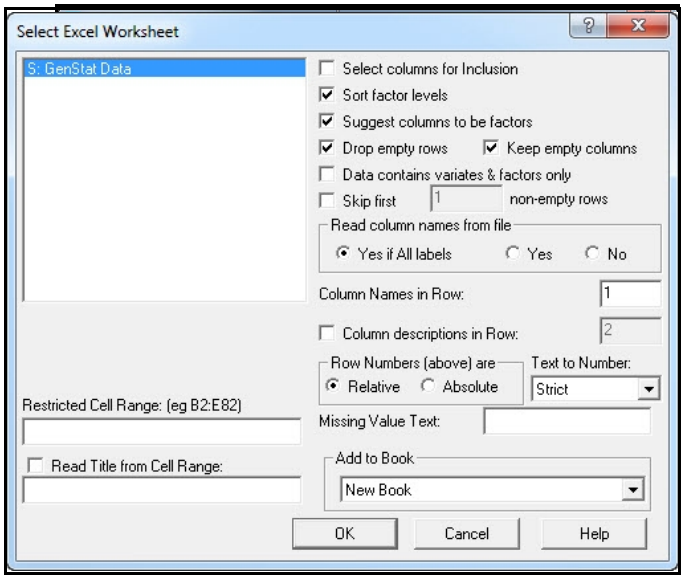


Figure 1.22

The Excel options dialog for the first file `Grazing 1.xls` is shown in Figure 1.22, and as the data in the Excel files we do not need to change any of the options to read in the data. Four identical dialogs will come up, and the OK button can just be clicked for each one.

Once the final Excel options dialog has been closed the resulting spreadsheet will be opened, as shown in Figure 1.23. This contains the 4 columns from the Excel files plus a factor column File which gives the filename (minus any extension) that the rows of data came from. If any of the files contain multiple sheets, then the File label also contains a second part which gives the page name of the sheets.

Spreadsheet [Grazing 11.GWB]*

Row	File	Plot	Treatment	Initial_Lwt	Final_Lwt
1	Grazing 1	1	TE	122.5	146.5
2	Grazing 1	1	TE	118.5	131.5
3	Grazing 1	1	TE	118	142
4	Grazing 1	1	TE	120.5	148.5
5	Grazing 1	1	TE	144.5	173.5
6	Grazing 1	1	TE	121.5	153
7	Grazing 1	1	TE	126	152
8	Grazing 1	1	TE	122	155.5
9	Grazing 1	2	TH	117	151
10	Grazing 1	2	TH	123	153
11	Grazing 1	2	TH	118	143
12	Grazing 1	2	TH	124	155
13	Grazing 1	2	TH	112	138
14	Grazing 1	2	TH	140	177.5
15	Grazing 1	2	TH	130.5	153.5
16	Grazing 1	2	TH	108	128
17	Grazing 1	3	DE	132	162
18	Grazing 1	3	DE	122.5	151
19	Grazing 1	3	DE	116	142.5
20	Grazing 1	3	DE	121	148

Figure 1.23

1.4 Reading and working with dates

In Excel, dates are stored as the number of days since 1 January 1900. There is an error in the Excel calculation of the number of days, as they include the day 29 February 1900. This did not exist, as only centuries divisible by 400 have a leap year (e.g the year 2000 was a leap year, but 1900 was not). Time within a day is stored as a fraction of a day. So 6am, 12am and 6pm are +0.25, +0.5 and +0.75 respectively. Genstat stores dates as the number of days since 1 March 1600. When reading a date column in from Excel, Genstat flags the date as having a base date of 1 January 1900. To mark a column in Excel as a date, the column name should have an suffix of `:D` and a time column should have a suffix of `:T`. In the later Excel 2007 .XLSX format, dates can be recognised by the majority of cells in the column having a date or time format, and so these do not need to have the `:D` or `:T` suffixes. If you bring Excel data in via the clipboard, and the data set has less than 256 columns and 32000 rows, Genstat will also recognise columns as dates based on their cell formats. When Genstat sends a spreadsheet column containing dates with a 1 January 1900 base to the server, it automatically adjusts the values to have the base date of 1 March 1600.

Figure 1.24 shows some meteorological data from Dipton, New Zealand that has the monthly mean of the daily maximum temperatures, the total rainfall and the maximum wind speed (gust) for the month and time during the day that the gust was recorded. This is in an Excel file `DiptonWeather.xls` in the Genstat Data folder. You can see that the columns `Date` and `GustTime` have suffixes `:D` and `:T` to mark these as a date and time respectively.

	A	B	C	D	E	F	G
1	Year!	Month!	Date:D	Mean_Temperature	Rainfall	MaxGust	GustTime:T
2	1997	Jan	01-Jan-97	25.0	102.0	90.6	9:55:32 p.m.
3	1997	Feb	01-Feb-97	23.5	154.0	85.3	3:30:15 a.m.
4	1997	Mar	01-Mar-97	21.8	48.0	91.9	2:58:30 a.m.
5	1997	Apr	01-Apr-97	17.9	191.0	92.0	4:08:29 p.m.
6	1997	May	01-May-97	16.1	59.1	84.7	12:54:30 a.m.
7	1997	Jun	01-Jun-97	12.7	51.9	77.6	4:10:03 a.m.
8	1997	Jul	01-Jul-97	11.3	85.0	78.6	11:25:02 a.m.
9	1997	Aug	01-Aug-97	13.7	52.0	117.5	11:34:43 p.m.
10	1997	Sep	01-Sep-97	16.8	53.0	118.0	11:32:58 p.m.
11	1997	Oct	01-Oct-97	18.8	65.3	93.7	3:29:12 a.m.
12	1997	Nov	01-Nov-97	19.3	106.4	85.1	1:34:30 p.m.
13	1997	Dec	01-Dec-97	20.8	143.3	76.3	7:57:17 p.m.
14	1998	Jan	01-Jan-98	23.1	36.5	88.3	5:17:14 p.m.

Figure 1.24

Using the either the Open or Open Example Data Sets options of the File menu, the data can be loaded into a Genstat spreadsheet in the usual way. This gives the spreadsheet shown in Figure 1.25. Note how `Date` and `GustTime` are displayed in date format. If the `:D` suffix was missing from the column name for `Date`, then the column would be displayed as in Figure 1.26, where just the number of days is displayed with no date format.

Spreadsheet [DiptonWeather.xls] (Dipton Weather Data!A2:G193)

Row	Year	Month	Date	Mean_Temperature	Rainfall	MaxGust	GustTime
1	1997	Jan	01-Jan-97	24.95	102	90.6	21:55:32.42
2	1997	Feb	01-Feb-97	23.51	154	85.3	3:30:15.13
3	1997	Mar	01-Mar-97	21.76	48	91.9	2:58:30.32
4	1997	Apr	01-Apr-97	17.88	191	92	16:08:29.05
5	1997	May	01-May-97	16.14	59.1	84.7	0:54:29.63
6	1997	Jun	01-Jun-97	12.74	51.9	77.6	4:10:02.79
7	1997	Jul	01-Jul-97	11.34	85	78.6	11:25:02.33
8	1997	Aug	01-Aug-97	13.65	52	117.5	23:34:43.39
9	1997	Sep	01-Sep-97	16.81	53	118	23:32:58.19
10	1997	Oct	01-Oct-97	18.83	65.3	93.7	3:29:11.89
11	1997	Nov	01-Nov-97	19.34	106.4	85.1	13:34:30.17
12	1997	Dec	01-Dec-97	20.78	143.3	76.3	19:57:16.53
13	1998	Jan	01-Jan-98	23.06	36.5	88.3	17:17:13.54

Figure 1.25

Spreadsheet [DiptonWeather.xls] (Dipt

Row	Year	Month	Date
1	1997	Jan	35431
2	1997	Feb	35462
3	1997	Mar	35490
4	1997	Apr	35521
5	1997	May	35551
6	1997	Jun	35582
7	1997	Jul	35612
8	1997	Aug	35643
9	1997	Sep	35674
10	1997	Oct	35704
11	1997	Nov	35735
12	1997	Dec	35765
13	1998	Jan	35796

Figure 1.26

The format used to display the date or time can be changed using the Spread | Column | Edit attributes menu (shown in Figure 1.27). For a numerical column, the Numerical Format will be set to Date. To change the date format click the Date Format button and dialog shown in Figure 1.28 can be used. You can select a date or time format from the drop-down list. If the wrong base date has been selected, this can be changed using the Start date from option. There are 38 different date formats that can be used: the day and months can be displayed with a leading zero or not, the month can be shown as a number, a three letter abbreviation or with the full name and the year with two or four digits, the order or day, month and year changed and times can also be shown with dates or on their own. The default date format, and month names for other languages can be set in the Tools | Options menu on the Date Format tab, as shown in Figure 1.29.

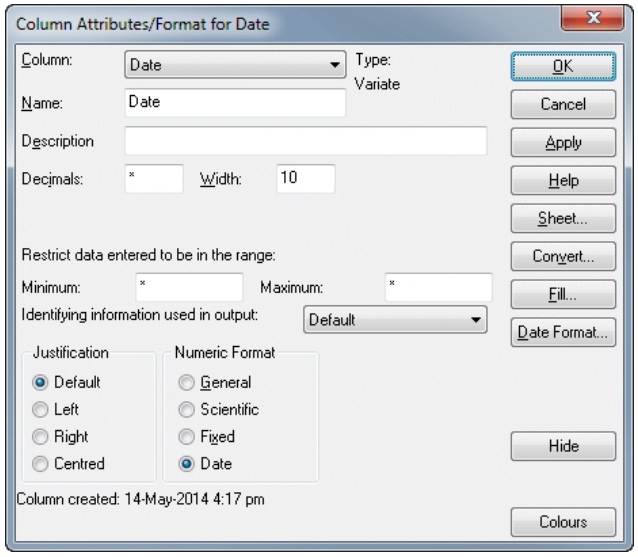


Figure 1.27

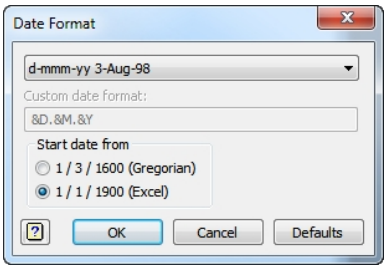


Figure 1.28

If for some reason you needed to change base date for a column, you can use the Spread | Calculate | Rebase Dates menu. If this is used on the column `Date` it would give the prompt sown in Figure 1.30. Normally this is not necessary as Genstat keeps track of he conversions need to save the data to an Excel file or to the server. The only time this would be required is if you wanted to add dates prior to 1/1/1900 to a column from an Excel file.

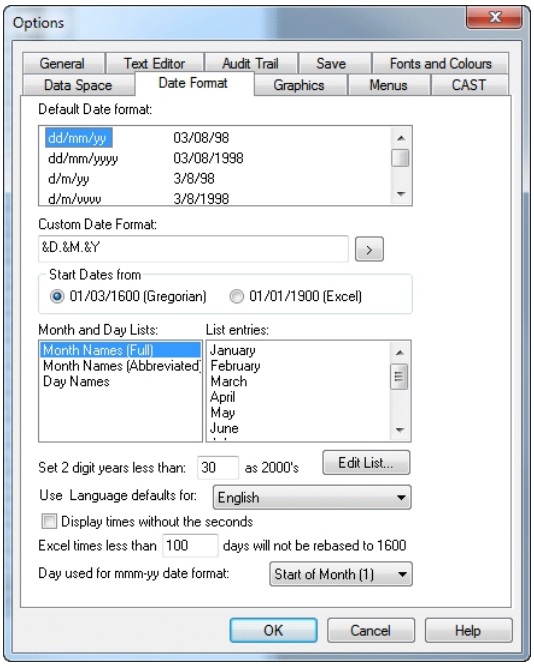


Figure 1.29

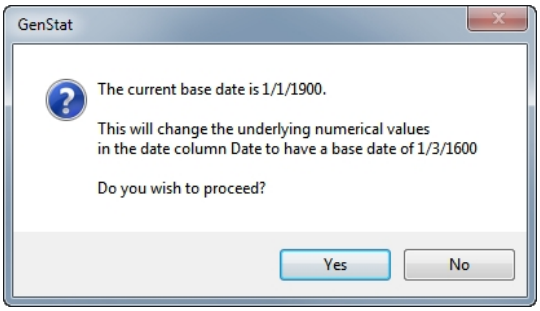


Figure 1.30

2 Data entry and validation

2.1 Entering data into a spreadsheet

The Genstat spreadsheet can also be used as a data-entry system. This includes a validation system, which may make it more reliable than the external alternatives.

The data shown below are taken from an experiment in New Zealand. Twelve sheep were divided into 4 "flocks" to follow 3 different drench programs. The initial weights of the sheep were recorded, and, after they were run for 3 months on their respective programs, their final weights were recorded.

Treatment	Rep	Weight in Kilograms	
		Initial	Final
Control	1	38	48
Control	2	31	42
Control	3	37	48
Control	4	34	41
Drenched once	1	36	52
Drenched once	2	35	50
Drenched once	3	38	52
Drenched once	4	32	49
Drenched twice	1	33	53
Drenched twice	2	34	49
Drenched twice	3	39	66
Drenched twice	4	36	57

To enter the data into a new spreadsheet, click on Spread in the menu bar and then click on New and Create, as shown in Figure 2.1. All the other options of the main Spread menu will be grey rather than black at this point, to show that they are not yet available (since the other menu options can only be selected for existing, active spreadsheets). This opens a menu containing a list of icons defining several types of spreadsheets that can be created. The last 6 icons in this list allow you to create blank spreadsheets for different types of data. The default spreadsheet type is for a Vector spreadsheet that allows columns of *variates* (numerics), *texts* (labels) and *factors* (grouped data) of equal length to be displayed simultaneously within a spreadsheet. When a spreadsheet is created, it can either be opened within a new

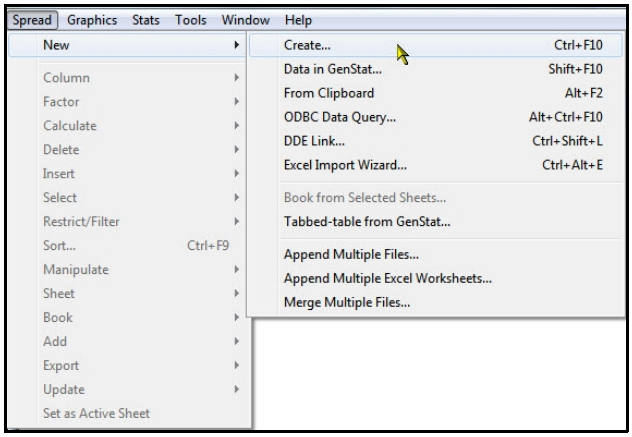


Figure 2.1

book or added as a page in existing book. The book that is to include the new spreadsheet is selected using the Create in Book list.

The data in our example will be in columns (or vectors) of variates and factors, so we have selected the Vector Spreadsheet icon, as shown in Figure 2.2. For a Vector spreadsheet you need to specify the number of rows and columns in the boxes provided. For this example we have entered 4 columns and 12 rows. It does not matter if you do not know the number of rows and columns needed initially for entering your data, you can easily insert or delete rows or columns at a later date. We have selected the New Book item from the Create in Book list to open the Vector spreadsheet within a new book.

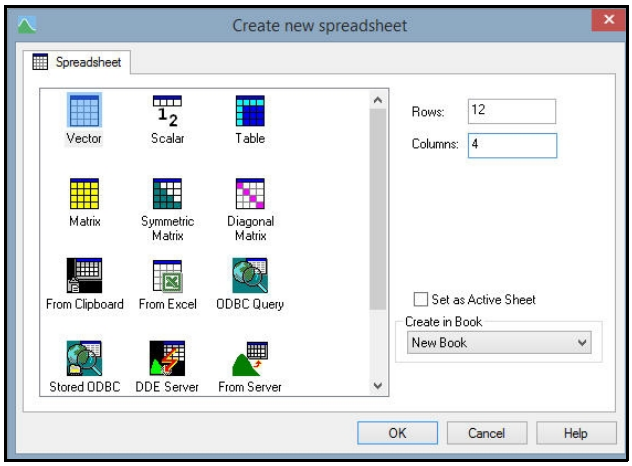


Figure 2.2

Clicking the OK button produces a blank spreadsheet within a single-paged book in a new window, as shown in Figure 2.3. By default, the 4 columns are initially created as variates and all the values are set as missing values represented by asterisks. The columns are labelled by default as `C1`, `C2`, `C3` and `C4`. If you enter data under these column names and transfer it to Genstat, four data structures will be created with the identifiers `C1`, `C2`, `C3` and `C4`. It is good practice to assign your own descriptive names to the columns. A column name must start with a letter or %, and the remaining characters can only be alphanumeric (A-Z, a-z, 0-9), or '%' or '_'. If you do use an illegal character in a column name, Genstat will convert these characters to valid ones. Column names can start with an underscore '_' but it is best not to use this as these columns will not be displayed in the data lists. Genstat uses hidden structures starting with an underscore for some system variables.

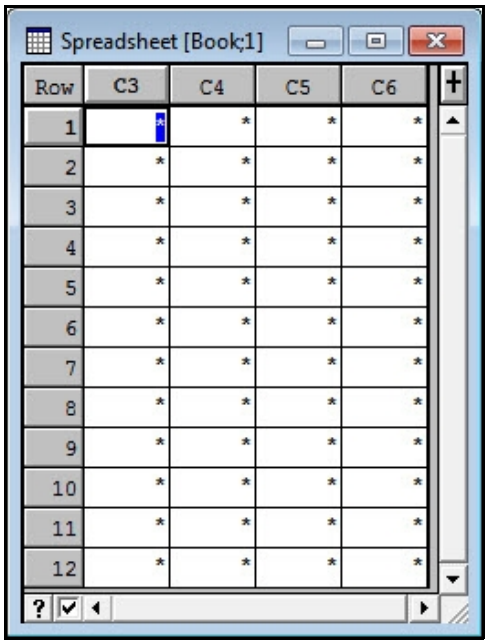


Figure 2.3

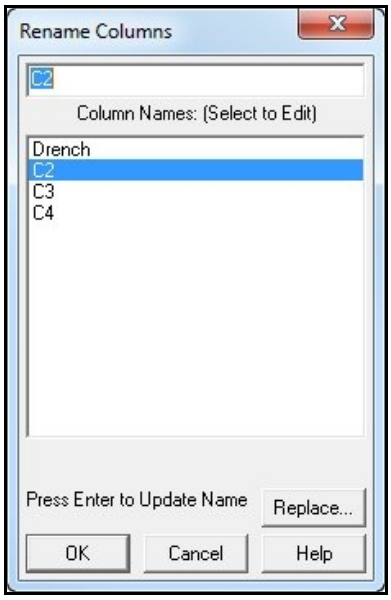


Figure 2.4

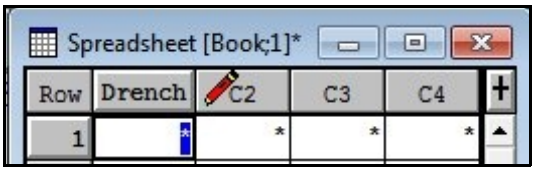


Figure 2.5

To rename the columns select the option Column from the Spread menu and then select Rename. Alternatively, press the Alt, Shift and R functions keys simultaneously. This opens the menu shown in Figure 2.4. To rename the columns simply select the column name to change in the list and enter the new name in the box at the top of the menu. Pressing the Enter key applies the changes and alters the name in the list of Column Names. In Figure 2.4 we have renamed C1 as Drench. An alternative way to rename the column is to move the cursor over the beginning of column heading until it changes to a pencil, as shown in Figure 2.5.

Clicking the left mouse button then opens a dialog where you can enter a new name. Figure 2.6 shows the column C2 being changed to Rep. In the same way we can change C3 to Lwt1 and C4 to Lwt2.

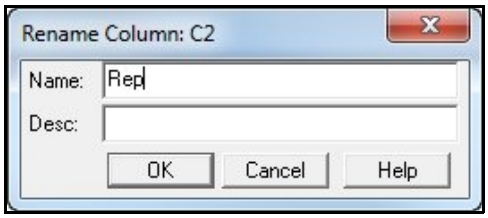


Figure 2.6

The column Drench contains grouped data, so we need to specify that the column is to be a factor before entering the labels. To convert the column to a factor, click anywhere on the column using the right-button on the mouse.

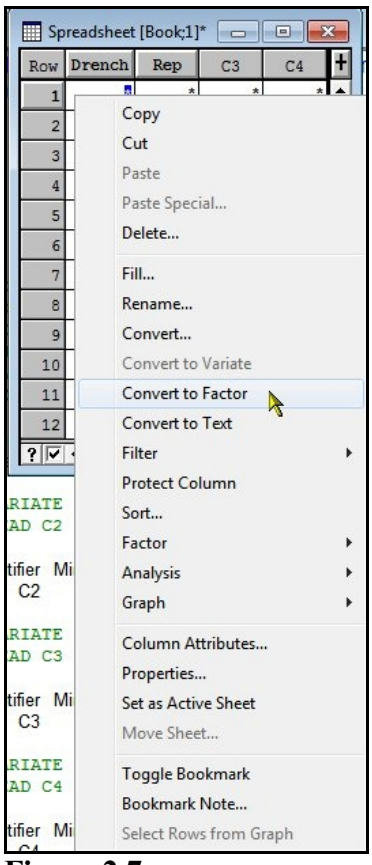


Figure 2.7

This pops up the menu shown in Figure 2.7. Selecting the option Convert to Factor opens the menu shown in Figure 2.8. Genstat has recognized that this is a new column that is being converted to a factor, and provides a menu to specify the levels and labels. The column Drench has three groups: Control, Drenched once and Drenched twice, so we have entered 3 in the Number of Levels box. We now want to change the labels to represent the three groups.

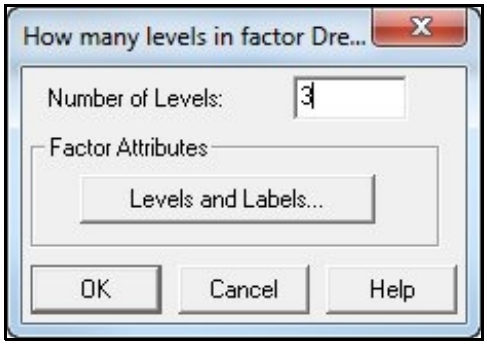


Figure 2.8

Clicking on the Levels and Labels button opens the menu in Figure 2.9. In this menu we enter `Control` for group 1 and press return or the down arrow to apply this label and for group 2 label this `Once` and finally for group 3 label this group as `Twice`. We can also colour the background of the groups different colours by clicking the colour wheel icon in the `Colour` column. This opens the colour selection dialog (Figure 2.10), which we can use to specify a colour by either selecting one of the predefined colours in a box (the custom colours are light shadings that are suitable for cell backgrounds) or by using the colour pallet area (rainbow coloured area) with the intensity slider at the right of the dialog.

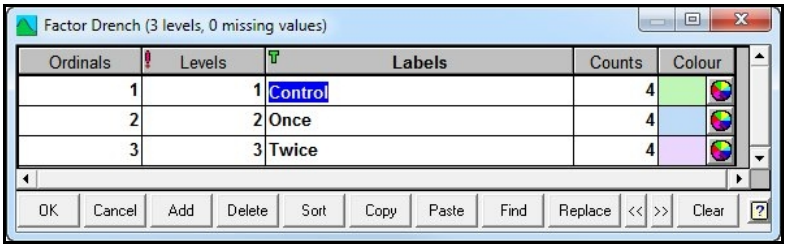


Figure 2.9

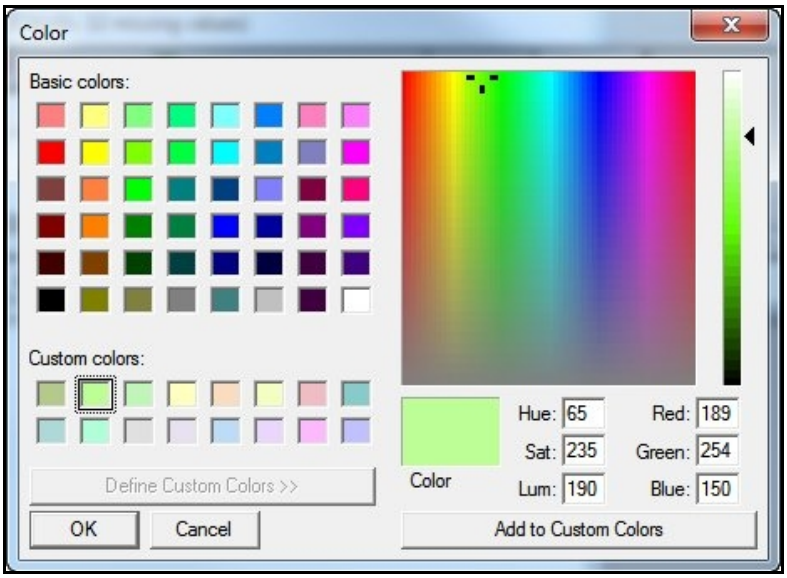


Figure 2.10

Clicking OK on this menu returns us to the spreadsheet where the column name now appears in italics and has a red ! at the start of the name (see Figure 2.11).

Figure 2.12

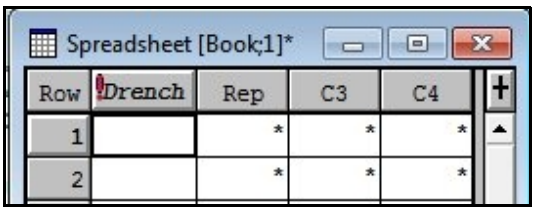


Figure 2.11

The labels can now be entered either by typing in the label name, or by double-clicking on a cell and selecting the appropriate label from the list, as shown in Figure 2.12. Alternatively, if you type the first character of the label and move to another cell, Genstat will fill in the rest of the name. For example, the letters `C`, `O` and `T` would be enough to specify the factor labels of `Control`, `Once` and `Twice` respectively. Note that if the labels begin with the same character then you will need to type as many characters as is required to distinguish between the labels. For example, if the factor contained 2 labels called `Farm` and `Field`, then

you would need to enter the first 2 characters (`Fa` or `Fi`) before moving onto a new cell. The case does not matter unless two items both have the same labels apart from case (e.g. `a` and `A` are both labels of the factor).

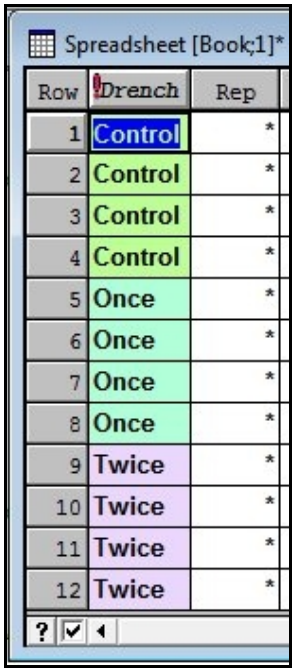


Figure 2.13

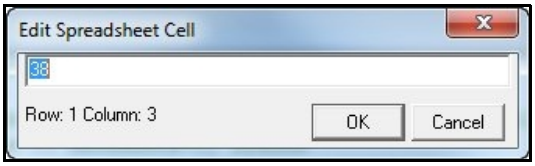


Figure 2.14

Figure 2.13 shows the column complete with the new factor labels.

We now enter the data into the columns `Lwt1` and `Lwt2`. We click on the cell for the first row of `Lwt1`, enter the value 38 and move to the next cell using the Enter key (alternatively you can use the down arrow key). We then type the value 31 in the second cell and so on. If a mistake is made and we want to edit individual characters within a cell, we can double-click on the cell. For example, double-clicking on the first cell in the column `Lwt1` opens the dialog box, shown in Figure 2.14, where we can change the value.

The column `Rep` contains patterned data with the values 1,2,3 and 4 repeated 3 times. There is a menu available to fill a column automatically with patterned data like this. Selecting Fill from the Calculate option of the Spread menu opens the menu shown in Figure 2.15. We have selected the column `Rep` and have entered the Start Value as 1, and the Ending Value as 4. Using the default option Fill to Bottom, the values 1,2,3,4 will be recycled until the bottom of the column. Clicking OK produces the spreadsheet shown in Figure 2.16.

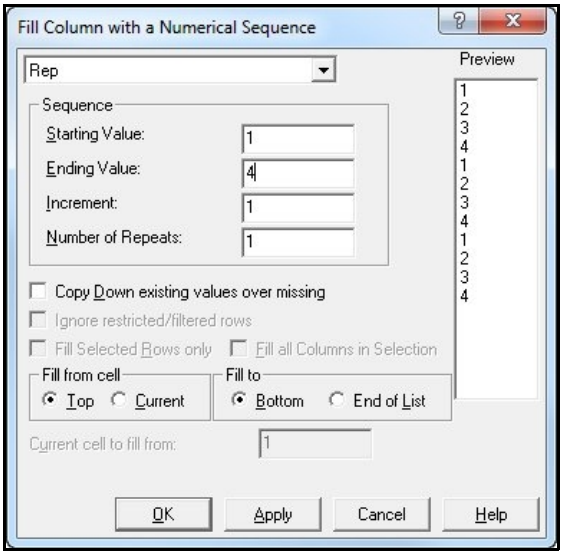


Figure 2.15

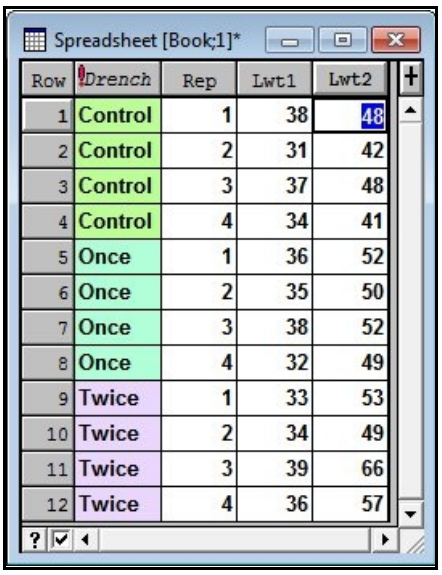


Figure 2.16

2.2 Data verification

When data are entered into a spreadsheet it is easy to mistype or enter an incorrect value. Genstat provides a facility for data verification within the spreadsheet through the double entry of the data. In this example we demonstrate how to check that the data within the columns `Lwt1` and `Lwt2` have been entered correctly. First we select Verify from the Sheet option of the Spread menu, which opens the menu shown in Figure 2.17. The columns that are to be verified are chosen by double-clicking on the name of the column in the list (or by selecting the column name and clicking the Verify button). When a column has been selected for verification, its name will be prefixed by the characters 'V:' in the list. In Figure 2.17 we have double-clicked on the names `Lwt1` and `Lwt2` to specify that they are to be verified.

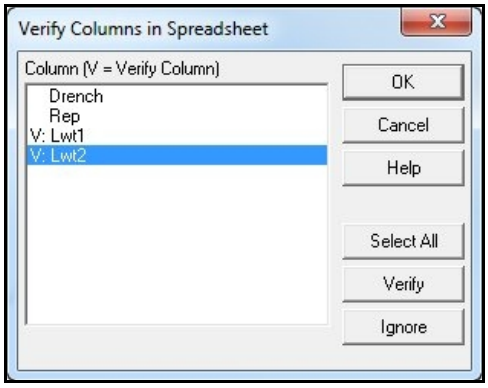


Figure 2.17

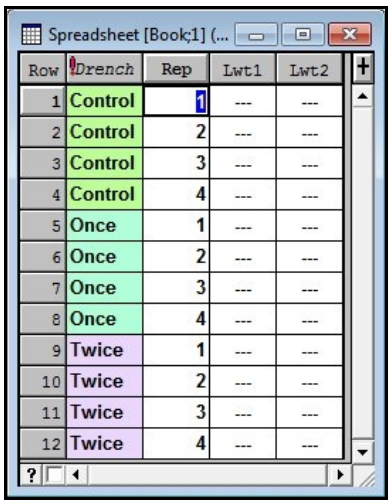


Figure 2.18

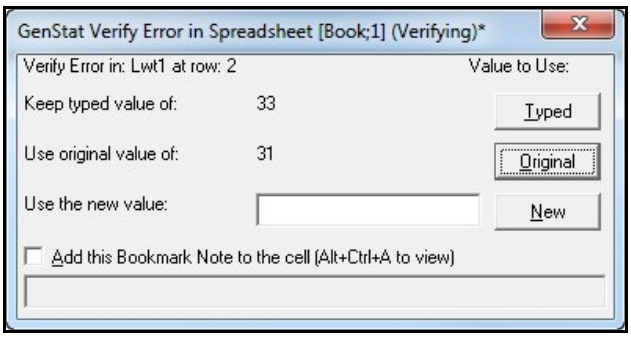


Figure 2.19

Clicking on OK changes the columns `Lwt1` and `Lwt2` in the spreadsheet to display three minus (−) characters in place of the values (see Figure 2.18). To verify the data, we now re-enter the values in these cells. First, we enter the value 38 in the first row of the column `Lwt1` and press enter to move the cursor to the next cell. As the value is correct in this cell, the value is redisplayed. Now in the second row we enter the value 33 and move to the next cell. On moving to the next cell Genstat has recognized that the value is different from the original value entered and displays the menu shown in Figure 2.19. Here you can specify the correct value and can add a comment to the cell if it is required. In our example we decide that the correct value should be 33, so we click on the Typed button to register this value.

On clicking the Typed button, a new spreadsheet is created containing a record of the mismatch in the data entry (see Figure 2.20). Each row within the new spreadsheet contains details of the column name, row, original value, new value typed and the new value. Any further mismatches in the data entry will be appended onto this spreadsheet. We then complete the data verification by entering the remaining values for the columns `Lwt1` and `Lwt2`.

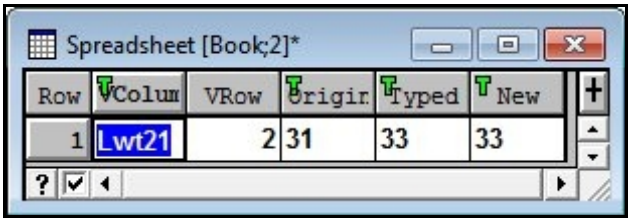


Figure 2.20

On entering the last value of the verification in row 12 of the column Lwt2 we are prompted with the menu shown in Figure 2.21. This menu allows you to set the verified columns as read-only to protect them from any further changes. Clicking on the Yes button changes the columns Lwt1 and Lwt2 to read-only, and provides a visual indication of this by changing the background on the column title to blue.

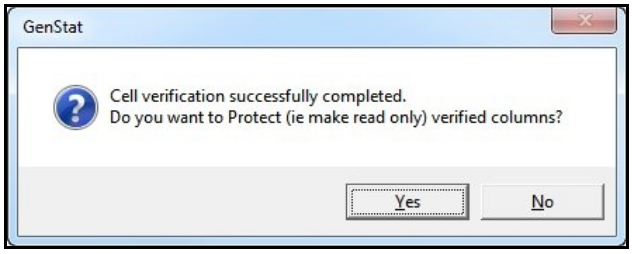


Figure 2.21

You can set or remove the protection for a column at any time using the Column/Sheet Protection menu. To remove the column protection on the columns Lwt1 and Lwt2 select Protection from the Column option of the Spread menu. This opens the menu shown in Figure 2.22. The columns within the spreadsheet are listed on the left of the menu and a protected column is identified in the list by having the prefix 'P:'. So, to remove the protection on the columns Lwt1 and Lwt2 double-click the names in the list to remove the prefix. Alternatively, selecting Lwt1 and Lwt2 within the list and clicking on the Unprotect button will remove the protection. You can protect a column in a similar way by double-clicking the name in the list or clicking the Protect button to include the prefix to indicate the column is to be protected. Clicking OK returns you to the spreadsheet and removes the blue background from the column titles.

Figure 2.22

Comparing spreadsheets is another form of data verification. You can compare two open spreadsheets within Genstat or you can compare a currently open Genstat spreadsheet with data from a foreign data source. For example, you could compare an open spreadsheet with another spreadsheet saved in GSH (Genstat Spreadsheet) format, or with data in an Excel file. The data set shown earlier in this section can also be found in the Genstat spreadsheet file called Drench.gsh. To illustrate the spreadsheet comparison facilities, we will now compare the data we have entered, with the data in the file. Selecting Compare from the Sheet option of the Spread menu opens the menu shown in Figure 2.23.The Data Source option identifies where the data that you wish to compare are located. The data we

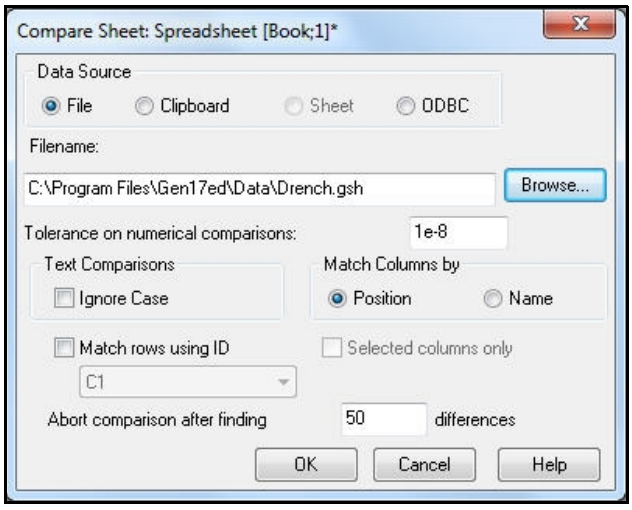


Figure 2.23

are comparing against are in a file so we select File. You need to enter the name of the file into the space provided or alternatively, you can select the file by clicking on Browse. The remaining options on the menu control how the comparison is to be done.

Leaving the default settings and clicking on OK pops up a dialog (Figure 2.24) to warn that the sheets are different, and prints a report in the Output Window, as shown below. There are two differences between the spreadsheet and the file `Drench.gsh`. The first difference reported is the record where the data value was changed during the data verification. The second indicates that in the current spreadsheet the column `Rep` is a *variate*, but in the file `Drench.gsh` this column has been saved as a *factor*.

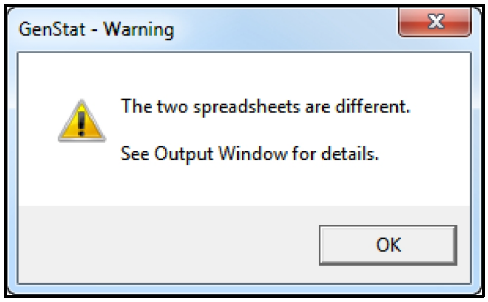


Figure 2.24

```
"Comparing Spreadsheets: New Data and Drench.GSH
   Column Types don't match: Rep = Variate vs Rep = Factor
   Mismatch on Lwt1 at row 2: 33 <> 31
 Spreadsheets are different. "
```

2.3 Inserting and deleting rows or columns

Columns and rows can be deleted using the Delete options on the Spread menu. To delete the column `Rep` click anywhere on the column and select Current Column from the Delete option on the Spread menu. Alternatively, you can use the mouse to delete the column `Rep`, we click on the column title using the left button and hold the mouse button down. The cursor will now appear as a hand containing a red cross. Drag the cursor over the spreadsheet and release the mouse (see Figure 2.25). Clicking Yes on the confirmation dialog will delete the column.

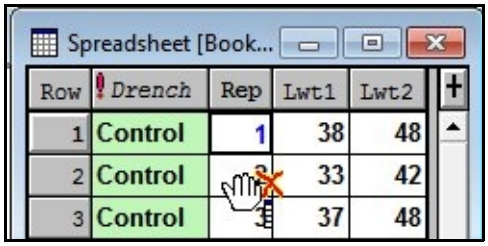


Figure 2.25

Rows can be deleted by clicking on the row number holding the mouse down, and then dragging it outside the spreadsheet. Figure 2.26 illustrates this being done with Row 2. You can select and drag multiple rows (or columns) for deletion in a similar fashion.

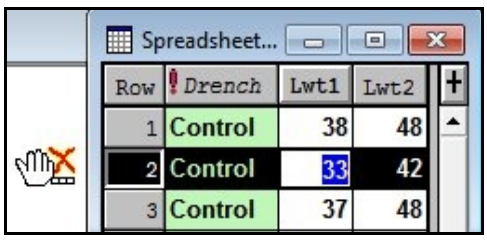


Figure 2.26

New columns and rows can be inserted using the facilities available within the Insert options on the Spread menu. To insert a new row at the bottom of the spreadsheet, click on the last row of the spreadsheet and select Row after current row from the Insert option on the Spread menu. This will add a new row as shown in Figure 2.27. New values are defaulted to missing values (represented by asterisks for numbers or empty cells for labels or strings).

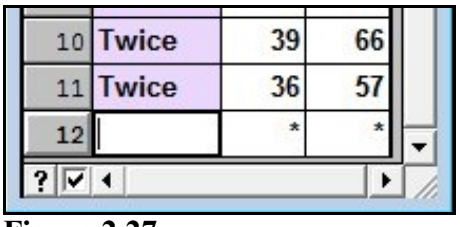


Figure 2.27

To insert a new column click in the Drench column and select Column after Current Column from the Insert option on the Spread menu. This opens the menu shown in Figure 2.28, where you can choose what type of data structure the new column will be, and set an initial value for each cell.

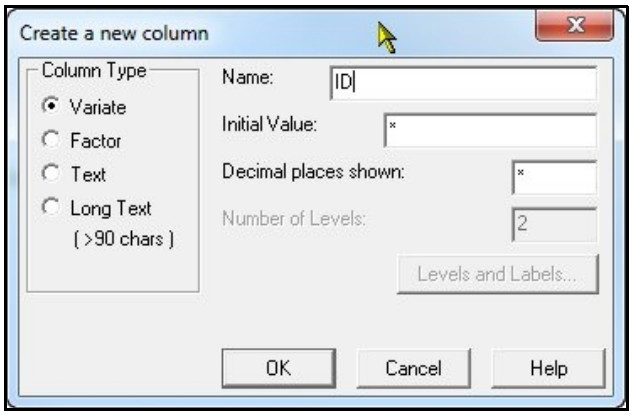
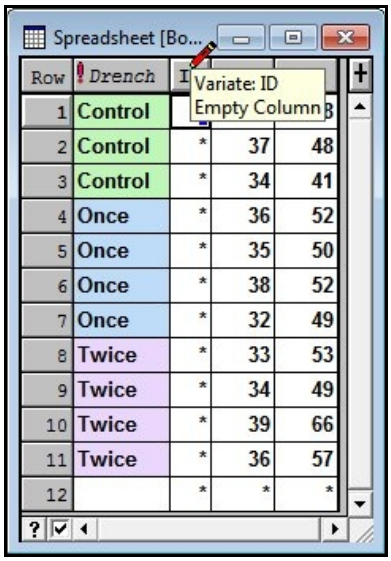


Figure 2.28



Row	Drench	ID		
1	Control	*	37	48
2	Control	*	34	41
3	Control	*	36	52
4	Once	*	35	50
5	Once	*	38	52
6	Once	*	32	49
7	Once	*	33	53
8	Twice	*	34	49
9	Twice	*	39	66
10	Twice	*	36	57
11	Twice	*	*	*
12				

Figure 2.29

Selecting Variate from the Column Type, entering the name ID and clicking OK produces a new column, initialized with missing values, as shown in Figure 2.29.

Another way of inserting a new column is to create a duplicate column. So, for example, if we want to duplicate the column Lwt1, we can select Duplicate from the Column options on the Spread menu.

This opens the menu shown in Figure 2.30 where we have selected the column Lwt1 and entered a new name for the duplicate column in the New Column Name field. You can create the duplicate column as a different type using the New Type options. Selecting the New Type as Variate and clicking OK inserts the duplicate column into the spreadsheet as shown in Figure 2.31.

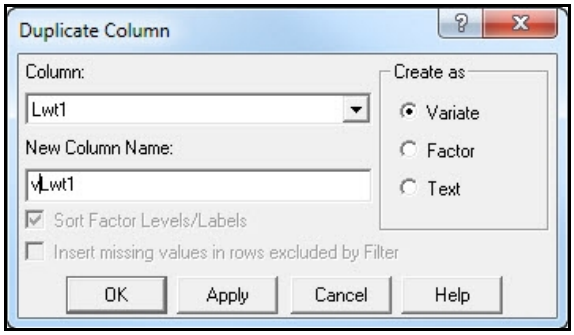
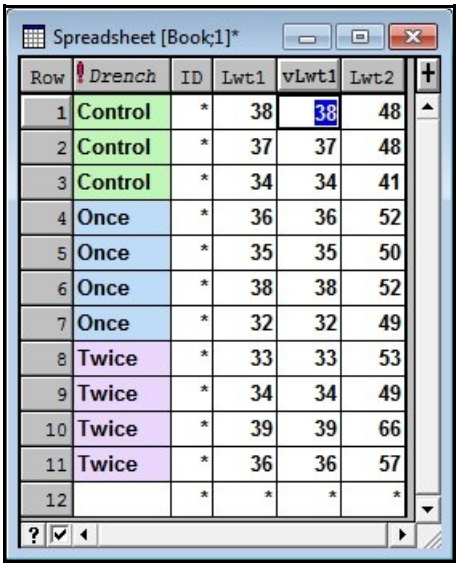


Figure 2.30



Row	Drench	ID	Lwt1	vLwt1	Lwt2
1	Control	*	38	38	48
2	Control	*	37	37	48
3	Control	*	34	34	41
4	Once	*	36	36	52
5	Once	*	35	35	50
6	Once	*	38	38	52
7	Once	*	32	32	49
8	Twice	*	33	33	53
9	Twice	*	34	34	49
10	Twice	*	39	39	66
11	Twice	*	36	36	57
12		*	*	*	*

Figure 2.31

2.4 Exercise

The following data are from an experiment assessing the durability of four different types of carpet, four machines were available to simulate the wear arising from daily use.

```
day machine carpet wear
1   1       d      38
1   2       a      17
1   3       c      38
1   4       b      39
2   1       a      19
2   2       d      22
2   3       b      26
2   4       c      35
3   1       b      41
3   2       c      54
3   3       a      11
3   4       d      36
4   1       c      59
4   2       b      36
4   3       d      22
4   4       a      16
```

Enter this data into a Genstat spreadsheet. Use the Fill menu from the Calculate option on the Spread menu to generate the `day` and `machine` information. Change the first 3 columns to factors and ensure the labels for carpet are a, b, c, d and e. Using the Verify menu from the Sheet option on the Spread menu, check that you have entered the data correctly. The data are stored in the file `Carpet.gsh`, compare your spreadsheet with this data set using the Compare menu. Close the spreadsheet and clear the data pool when you have finished.

3 Data manipulation

Before any statistical analyses are performed, the data may have to be manipulated into the correct form required for the analysis. This can sometimes be time consuming and awkward. We now show some advanced data handling techniques that make data manipulation easier.

3.1 Defining subsets of data values

When dealing with a large set of data, you often need to be able to select a subset of values to study, either temporarily, or for the remainder of a session. Genstat caters for this by allowing you to impose *restrictions* (*filters*) to define subsets of vectors (*variates*, *texts* or *factors*). The vectors keep all their original values, but subsequent commands working with the vectors will restrict their attention only to the subset.

One way of doing this is provided by the Genstat spreadsheet. For example, suppose for the drench data we wish to display a list of the sheep whose final weight is less than 51 kilograms. First, we close the spreadsheet we have been working with, and form a new spreadsheet in a new book containing only the columns `Drench` and `Lwt2`. Selecting New and then Data in Genstat from the Spread menu generates the Load Spreadsheet menu in Figure 3.1. In this menu we select `Drench` and `Lwt2` as the data to load, New Book from the Load in book list and click on Load.

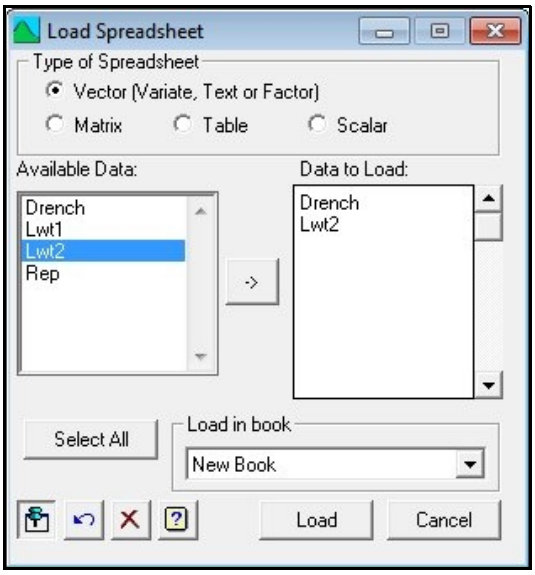


Figure 3.1

The resulting spreadsheet is shown in Figure 3.2. We now generate the menu in Figure 3.3 by selecting Restrict/Filter and By Logical Expression from the Spread menu. The Restrict Spreadsheet menu allows you to restrict or filter the data within a spreadsheet based on a logical expression. The Expression boxes define the condition, and the Restriction Type buttons indicate whether the restriction is formed by *including* or *excluding* the units (or rows) that satisfy the logical condition.

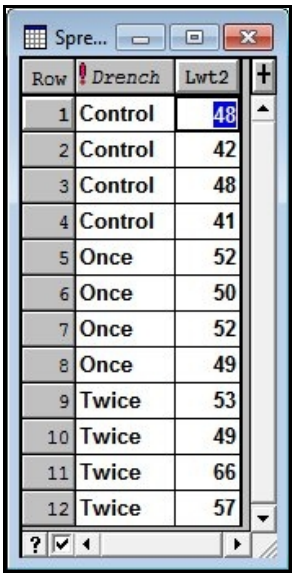


Figure 3.2

In our example we want to include all the units within the restriction that is where the units in the Lwt2 column are less than 51. To create the expression for this restriction we double-click Lwt2 from the Columns list into the first expression box. We then double-click the Less Than option from the Comparison list which puts a '<' symbol into the expression box and then type 51. We select the Include option from the Restriction Type box and click OK.

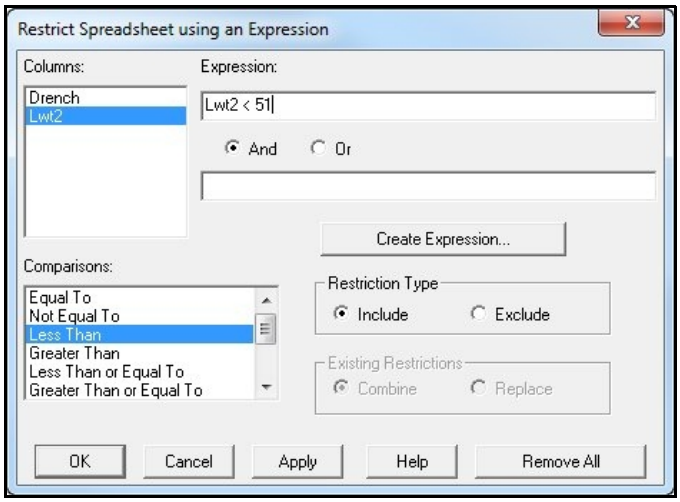


Figure 3.3

The resulting spreadsheet shown in Figure 3.4 now shows only the requested subset of units. The row heading now has stripes of red to indicate that the spreadsheet is restricted. The restriction is also shown in the status bar which now displays the number of rows as 7/12 to indicate 7 out of 12 rows are included by the current restriction (as shown in Figure 3.5). When we use these vectors in future, until we cancel the restriction, operations will be restricted to just the specified set of units. (This applies both to operations with menus and with commands.)

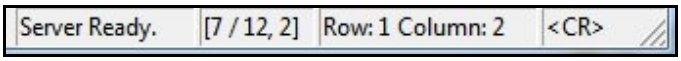


Figure 3.5

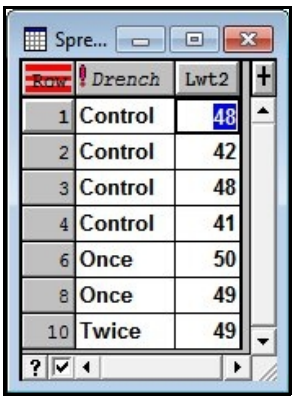


Figure 3.4

This is illustrated in Figure 3.6, where we use the PRINT directive, to print Lwt1 and Lwt2. Notice that, even though we only included Lwt2 in our restrictions as these two vectors are printed in parallel, the restriction is applied to both. Initial weights of the sheep (Lwt1) are only displayed for the filtered values for the final weights (Lwt2).

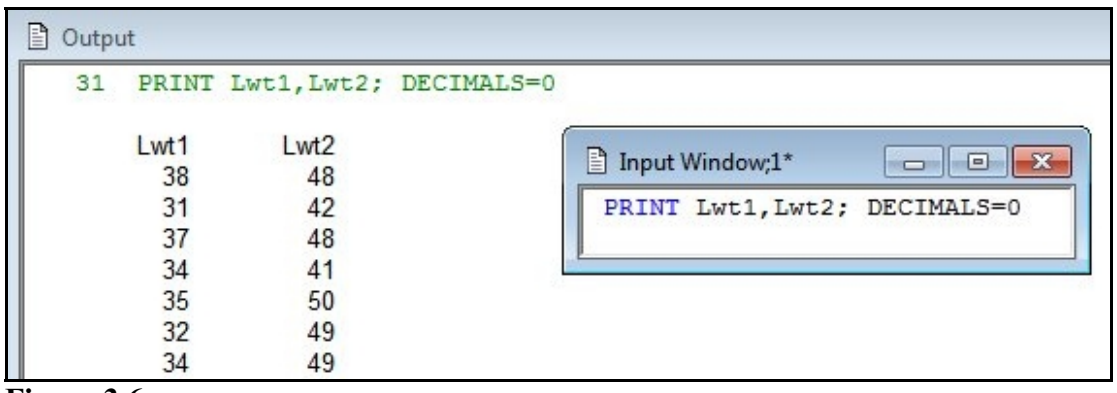


Figure 3.6

The restricted units are not discarded, and can be viewed in the spreadsheet in an alternative colour. To do this select Display Restricted Rows from the Restrict/Filter options on the Spread menu. This will display all the rows in the spreadsheet, but with the restricted rows shown in the red (default colour); see Figure 3.7.

You can also toggle the display of the restricted rows by clicking the '+' button in the top right-hand corner of the spreadsheet (positioned above the scroll bar as indicated by the cursor shown in Figure 3.7).

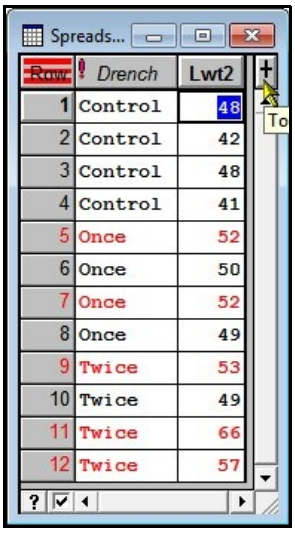


Figure 3.7

As the restricted units are not discarded, we can also change the restriction to look at some other set of units, or impose a further restriction. For example, say we now want to add to our restriction the condition that we want to identify the sheep whose treatment was to be drenched once. To combine a new restriction with the existing restriction, we could use the Restrict Spreadsheet using an Expression menu again or, alternatively as the column `Drench` contains grouped data (*factor*), we can use the Restrict on Factor menu, as shown in Figure 3.8.

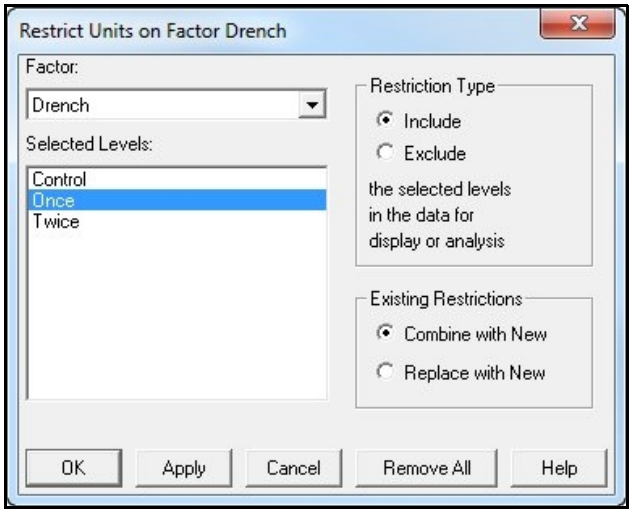


Figure 3.8

To open the menu, select the To Groups (factor levels) item from the Restrict/Filter option on the Spread menu. This menu displays the labels or levels of a factor, which you can select to filter the data by. Select `Once` from the Selected Levels and the Include option from the Restriction Type. To combine this restriction with the current subset select the Combine with New setting from the Existing Restrictions options. Clicking OK produces the spreadsheet shown in Figure 3.9.

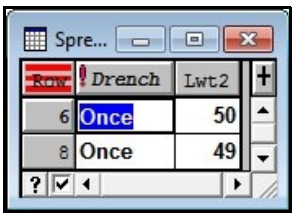


Figure 3.9

In creating our subset of data we have created one subset using a logical condition and then further restricted this set using a second condition. Using the Restrict Spreadsheet using an Expression menu you can create a restriction by combining the two logical conditions into a single condition using both the expression boxes.

Figure 3.10 shows how to do this for our example. First, we remove the current restriction by clicking on the Remove All to ensure we are using the complete set of data. Now, as before, we enter the condition for `Lwt2` less than 51 in the first expression box. Then, in the second box, we enter the condition for the restriction: `Drench .in.'Once'`. The ".in." operator, which is explained in Section 2.7, can be inserted by double-clicking on Inclusion in the list of Comparisons. To combine these two conditions we have selected the And option between the boxes; that is, we want to include into our restriction sheep whose final weight is less than 51 kilograms and that have been drenched once. Clicking on OK produces the same spreadsheet as shown in Figure 3.10.

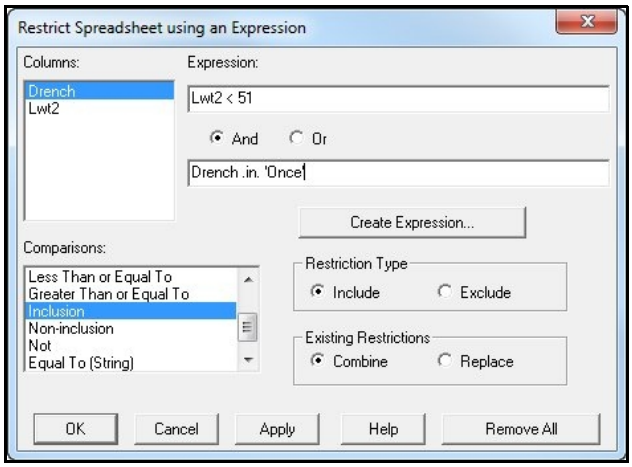


Figure 3.10

To restore the data to its original form at any time you need to remove the restriction applied to the data. You can do this be selecting Remove All from the Restrict/Filter option of the Spread menu or by clicking the clear restriction button in the toolbar as shown in Figure 3.11.

Figure 3.11

If you want to store a subset of the units in a vector rather than restricting the original data set, you can use the Subset menu. This is opened by selecting the Subset option on the Data menu. You can also define the restriction by specifying the rows in the spreadsheet explicitly. The rows are selecting using the Select line of the Spread menu, and the Restrict/Filter menu then allows you to indicate how these are to generate the restriction.

Whichever way the restriction is defined within the spreadsheet, it is imposed within Genstat using the `RESTRICT` directive and this provides an alternative if you wish to define very complicated restrictions or to restrict vectors too large to be displayed in a spreadsheet.

3.2 Exercise

The file `Computer.gsh` contains the number of personal computers sold in a shop during each month of the year 2001, together with the prices charged. Using the Restrict/Filter options on the Spread menu subset the data to display only the rows where the months the price is greater than £1100.

Build up the subset further by filtering the rows where the number sold is less than 15.

Remove the restriction from the spreadsheet and restrict the data again this time using both the conditions at the same time (you will need to use both Expression fields of the Restrict using an expression menu). Remove the restriction when you have finished.

3.3 Sorting data

The spreadsheet allows you to reorder the units of a list of vectors according to one or more index vectors. To illustrate this we will reintroduce the spreadsheet `Drench.gsh`. First we close all the currently open spreadsheets either by using the Close option on the File menu or by clicking the "X" button at the top right-hand corner of the spreadsheet windows. We now open the file `Drench.gsh` using the Open option of the File menu, which produces the spreadsheet shown in Figure 3.12. We now want to sort the data in the spreadsheet by specifying the final weights in ascending order. To do this, select Sort from the Spread menu; this produces the menu shown in Figure 3.13.

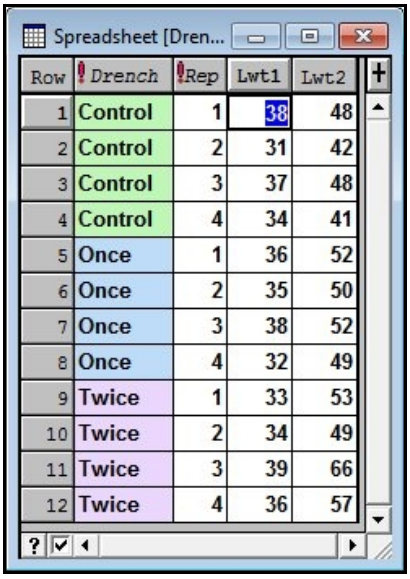


Figure 3.12

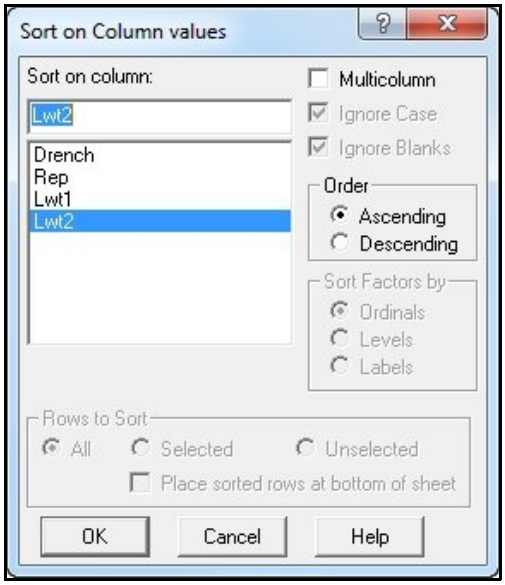


Figure 3.13

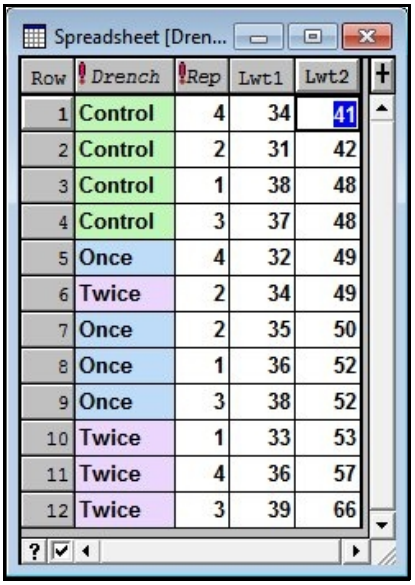


Figure 3.14

We have selected `Lwt2` from the Sort on column list to be the index for the sort, and selected Ascending from the Order options. Clicking OK produces the spreadsheet shown in Figure 3.14, where the rows are reordered, so that the values in the `Lwt2` column are in ascending order. If you have textual columns, you can sort these alphabetically.

You can also do multi-column sorts, where you specify an ordering based on a series of columns. The columns are then sorted using the first column, then rows that have equal values in the first column are sorted according to a column, and so on. To illustrate this we will sort the data in alphabetical order for `Drench` and then in ascending order by `Lwt1` within each drench group.

Using the Sort menu we select Drench as the first column that we are going sort by and select Labels from the Sort Factors By option to sort the factor in order of its labels. Selecting the Multicolumn option adds the text Key;1 to the column Drench in the Sort on Column list. The Key;1 tells us that this is the first column that we are going to sort by. Now select the column Lwt1, this adds Key;2 to the text, telling us that this is the second column by which the data will be sorted (see Figure 3.15). Clicking on OK produces the spreadsheet shown in Figure 3.16.

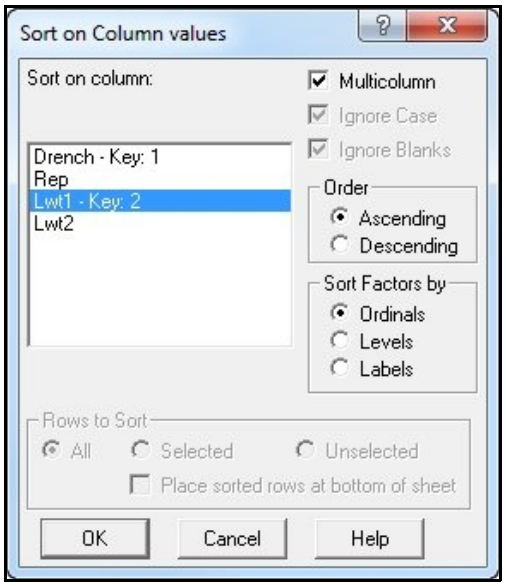


Figure 3.15

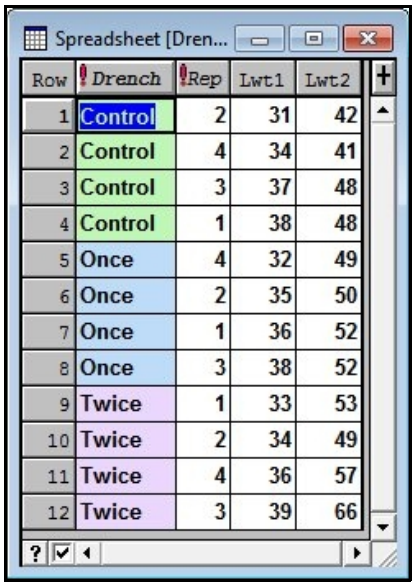


Figure 3.16

You can also sort a selection within a spreadsheet. For example, to sort the final weights for Rep 4 in descending order we first need to make a selection of the rows containing Rep 4. To make a multiple selection, click on the first row of the selection, then hold the control key down and click on the second row of the selection, and so on (keeping the control button selected). Figure 3.17 shows the selection of all the rows for Rep 4. Opening the Sort menu when a selection has been made enables some additional options at the bottom of the menu, as shown in Figure 3.18.

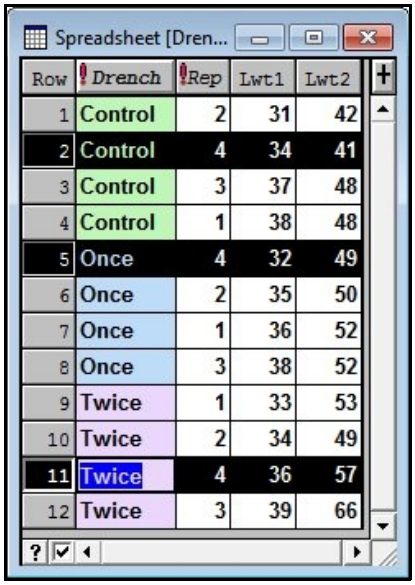


Figure 3.17

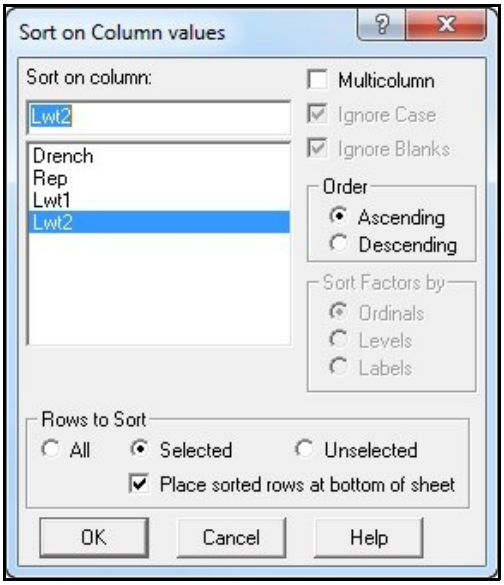


Figure 3.18

Select Lwt2 from the Sort on column list, Ascending from the Order options and Selected from the Rows to Sort options. When sorting a selection of rows, it is useful to group the results together to see how they have been sorted. You do this at the bottom of the spreadsheet by selecting the Place sorted rows at bottom of sheet from the Rows to Sort options.

Figure 3.19 shows the results of this process on our example spreadsheet.

Other facilities for sorting data are provided by the Genstat SORT directive.

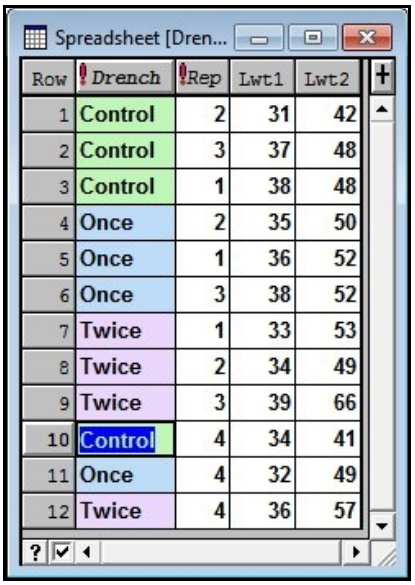


Figure 3.19

3.4 Exercise

The file Computer.gsh, already investigated in Section 2.2, contains the number of personal computers sold in a shop during each month of the year 2001, together with the prices charged.

Sort the spreadsheet in descending order according to the number of computers sold. Sort the spreadsheet again, this time using the months in alphabetical order.

Sort the spreadsheet with multiple indexes, firstly by the price and then by the number of PCs sold.

Close the spreadsheet and clear the data pool when you have finished.

3.5 Appending, stacking and unstacking data

We first show how to append data to a spreadsheet. This is particularly useful when your data are split across two data files, or on separate worksheets within a spreadsheet. The following example demonstrates how to append data that are stored on different worksheets within an Excel file. The file `Toysales.xls` contains a subset of data of yearly sales data over 3 years of a toy company for the sale of toy dogs and kittens. The data set includes the location of the shop, the number of toys sold and the price per unit. The worksheet Dog Sales contains the figures for the toy dogs, the worksheet Kitten Sales contains the figures for sale of their kitten toy during the same period, and the worksheet Dog and Kitten Sales contains data on both of these. First, we load the toy dog sales data file into a spreadsheet. We click on the Open option of the File menu, select the file `Toysales.xls` and click on OK. This uses the Excel import wizard, described in Chapter 1. Here we simply need to select the worksheet Dog Sales in the Select Excel Worksheet for Import menu (Figure 3.20), select New Book in the Add to Book list, and click Finish.The resulting spreadsheet is shown in Figure 3.21.

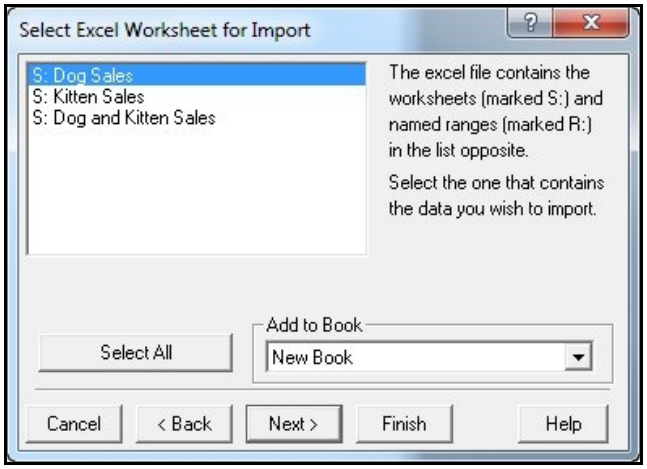


Figure 3.20

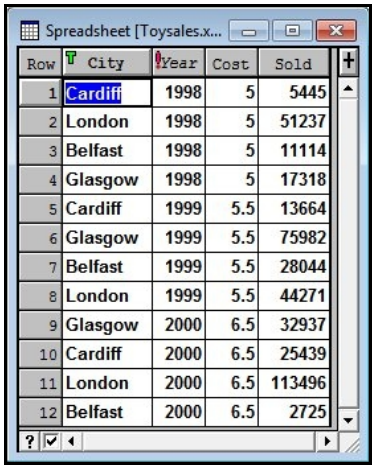


Figure 3.21

To append the data for the toy kitten sales we need to use the Append Data to Sheet menu (see Figure 3.22) which is selected using Append from the Manipulate options on the Spread menu. We select File as our data source and use the Browse button to select the file `Toysales.xls` and put the name and path of the file into the File box. We then select Name for the Match Columns by option as we want to match the columns from the Genstat spreadsheet file by their column names. To identify the different data sets within the spreadsheet we enter the name `Toy` in the Record Source in Factor box. This will create a new factor in the spreadsheet where each level of the factor represents the different appended data sets. By default these are simply the numbers 1 and 2, however, you can specify labels for these by entering names into the Factor Label boxes. In Figure 3.22 we have entered `Kitten` to label the appended data

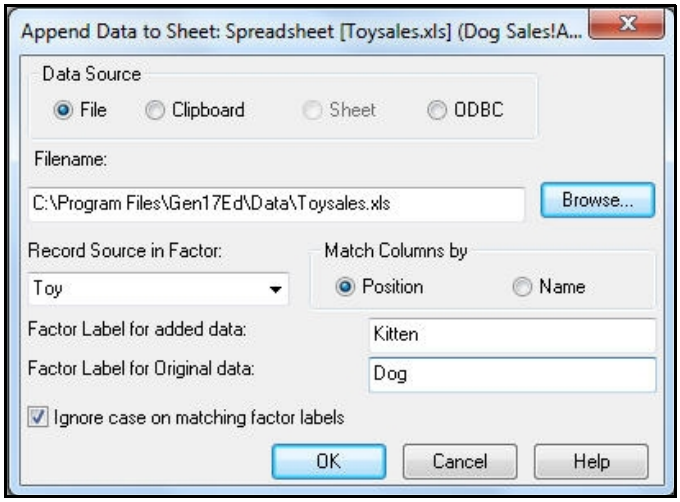
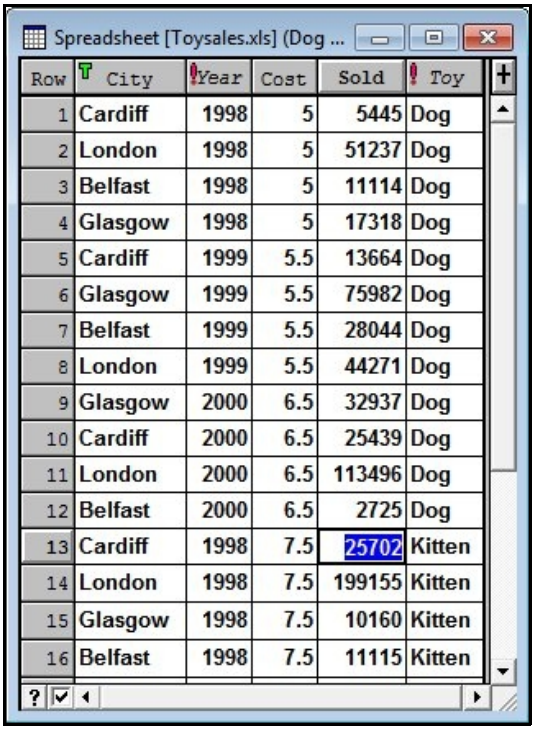


Figure 3.22

and Dog to represent the original data.

Clicking OK produces the Select Excel Worksheet for Import menu again (Figure 3.20). This time we select the Kitten Sales worksheet, and click Finish. Genstat reads the data from the file, appends the values onto the current spreadsheet, and creates a new factor Toy using the labels Dog and Kitten to represent the different data sets (see Figure 3.23).



Row	City	Year	Cost	Sold	Toy
1	Cardiff	1998	5	5445	Dog
2	London	1998	5	51237	Dog
3	Belfast	1998	5	11114	Dog
4	Glasgow	1998	5	17318	Dog
5	Cardiff	1999	5.5	13664	Dog
6	Glasgow	1999	5.5	75982	Dog
7	Belfast	1999	5.5	28044	Dog
8	London	1999	5.5	44271	Dog
9	Glasgow	2000	6.5	32937	Dog
10	Cardiff	2000	6.5	25439	Dog
11	London	2000	6.5	113496	Dog
12	Belfast	2000	6.5	2725	Dog
13	Cardiff	1998	7.5	25702	Kitten
14	London	1998	7.5	199155	Kitten
15	Glasgow	1998	7.5	10160	Kitten
16	Belfast	1998	7.5	11115	Kitten

Figure 3.23

An alternative menu is available for appending data from different worksheets within a single Excel file. To illustrate this we now close the sheet shown in Figure 3.23 and select Append Multiple Excel Worksheets from the New submenu on the Spread menu. This opens the menu shown in Figure 3.24 where you can select one or more worksheets or ranges to be appended into a single Genstat spreadsheet. In this menu we have selected both the Kitten Sales and Dog Sales worksheets in the Select Multiple Sheets/Ranges list. We then select Name for the Match Columns by option as we want to match the columns from the Genstat spreadsheet file by their column names.

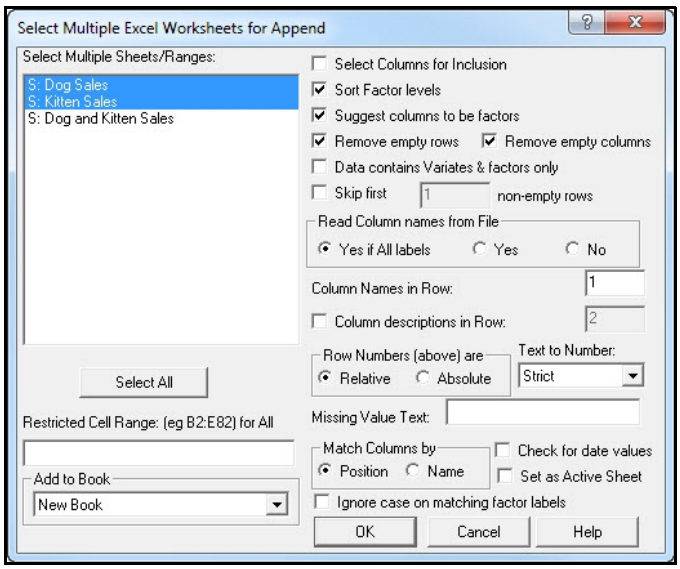


Figure 3.24

Clicking OK produces the sheet shown in Figure 3.25. The only difference between this sheet and the one shown in Figure 3.23 is that a source factor column called Worksheet has been automatically generated in this example where the labels of this factor are the names of the worksheets.

We now close this sheet, and will next open the data in the third worksheet, Dog and Kitten Sales. Genstat has two menus that enable you to easily stack or unstack your columns of data within a spreadsheet. We will first look at how you can stack columns together.

Spreadsheet [Toysales.xls] (Dog Sales!A2:D13)*

Row	WorkSheet	City	Year	Cost	Sold
1	Dog Sales	Cardiff	1998	5	5445
2	Dog Sales	London	1998	5	51237
3	Dog Sales	Belfast	1998	5	11114
4	Dog Sales	Glasgow	1998	5	17318
5	Dog Sales	Cardiff	1999	5.5	13664
6	Dog Sales	Glasgow	1999	5.5	75982
7	Dog Sales	Belfast	1999	5.5	28044
8	Dog Sales	London	1999	5.5	44271
9	Dog Sales	Glasgow	2000	6.5	32937
10	Dog Sales	Cardiff	2000	6.5	25439
11	Dog Sales	London	2000	6.5	113496
12	Dog Sales	Belfast	2000	6.5	2725
13	Kitten Sales	Cardiff	1998	7.5	25702
14	Kitten Sales	London	1998	7.5	199155
15	Kitten Sales	Glasgow	1998	7.5	10160
16	Kitten Sales	Belfast	1998	7.5	11115
17	Kitten Sales	London	1999	7.99	181643

Figure 3.25

The data in the worksheet Dog and Kitten Sales are shown in Figure 3.26. There are six columns; the location (now a factor), year sold, two columns of sales and two columns of prices. We want to stack the two columns of sales (SoldDog and SoldKitten) together and the two columns of prices (CostDog and CostKitten) together.

Spreadsheet [Toysales.xls] (Dog and Kitten Sales!A2:F13)

Row	City	Year	CostDog	SoldDog	CostKitten	SoldKitten
1	Cardiff	1998	5	5445	7.5	25702
2	London	1998	5	51237	7.5	199155
3	Belfast	1998	5	11114	7.5	11115
4	Glasgow	1998	5	17318	7.5	10160
5	Cardiff	1999	5.5	13664	7.99	34450
6	Glasgow	1999	5.5	75982	7.99	20652
7	Belfast	1999	5.5	28044	7.99	14684
8	London	1999	5.5	44271	7.99	181643
9	Glasgow	2000	6.5	32937	8.5	59921
10	Cardiff	2000	6.5	25439	8.5	65965
11	London	2000	6.5	113496	8.5	188046
12	Belfast	2000	6.5	2725	8.5	25871

Figure 3.26

To stack the columns we select Stack from the Manipulate options in the Spread menu, which produces the menu shown in Figure 3.27. We want to stack two columns together at a time, so we enter 2 into the Number of columns to stack together box. We enter the factor name Toy into the Record column source in Factor box; this creates a new column containing a factor where each level will represent a column that has been stacked. We click in the Stack Columns list and then select the two columns CostDog and CostKitten by clicking the -> button. This puts the names of these columns into the Stack Columns list and prefixes them with a '1'. The 1 indicates the

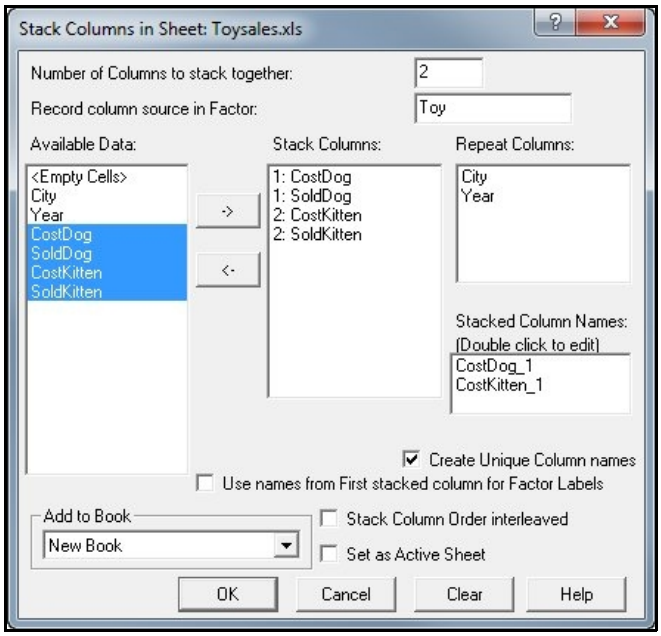
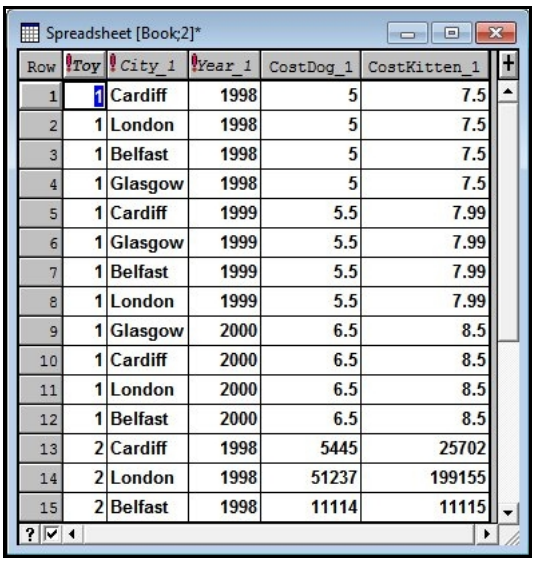


Figure 3.27

columns that we wish to stack together in our first stacked column. We now select the two columns SoldDog and SoldKitten, clicking on the -> button to copy these into the Stack Columns list. This time the names are prefixed by a 2 indicating that these columns will be placed into the second stacked column. You can include repeated columns in the stacked spreadsheet. For example, we have selected the columns City and Year to be repeated for each level of the stacking by clicking in the Repeat Columns box and then double-clicking on their names in the Available data box.

Clicking OK produces a new spreadsheet shown in Figure 3.28. This spreadsheet consists of five columns: a column for the source factor (Toy), a repeated column for the city, another repeated column for the year, and the two stacked columns with the costs and the numbers sold. The new spreadsheet creates default names, with suffixes "_1" for the repeated and stacked columns. However, you can change these using the Column Rename menu (obtained by selecting Rename from the Column option of the Spread menu.

Now suppose that we want to unstack columns in the stacked spreadsheet, so that we have a column of data for each year.


Spreadsheet [Book;2]*

Row	Toy	City_1	Year_1	CostDog_1	CostKitten_1
1	1	Cardiff	1998	5	7.5
2	1	London	1998	5	7.5
3	1	Belfast	1998	5	7.5
4	1	Glasgow	1998	5	7.5
5	1	Cardiff	1999	5.5	7.99
6	1	Glasgow	1999	5.5	7.99
7	1	Belfast	1999	5.5	7.99
8	1	London	1999	5.5	7.99
9	1	Glasgow	2000	6.5	8.5
10	1	Cardiff	2000	6.5	8.5
11	1	London	2000	6.5	8.5
12	1	Belfast	2000	6.5	8.5
13	2	Cardiff	1998	5445	25702
14	2	London	1998	51237	199155
15	2	Belfast	1998	11114	11115

Figure 3.28

The Unstack menu (Figure 3.29) is opened by selecting Unstack from the Manipulate option of the Spread menu. This menu splits up single columns into multiple columns based on the levels of an unstacking factor. In this example the unstacking factor is Year_1, which we created using the stack menu. So we double-click the name Year_1 from the Available Data list to put it into the Unstacking Factor box. The columns are unstacked so that the rows of each level of the unstacking factor become a new column. There are 3 levels for the factor Year, so the resulting spreadsheet should contain 3 columns for each unstacked column. Click in the Unstack Columns list, then highlight the names CostDog_1 and SoldDog_1. Now click the -> button to transfer them across to the Unstack columns list. The ID Factors box allows you to specify factors to identify the rows within each year, to ensure that these correspond across columns. (This is important here, as the cities are not in the same order for every year.)

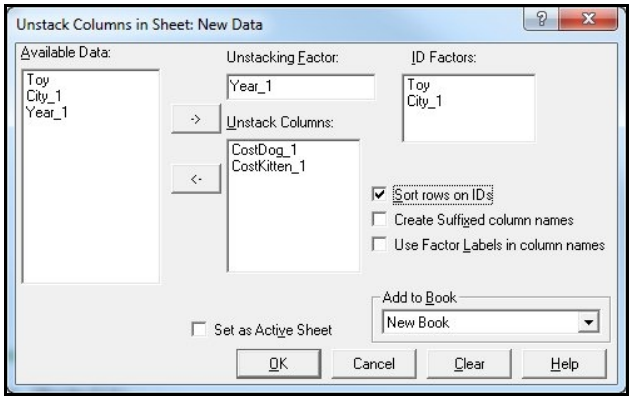


Figure 3.29

Spreadsheet [Book;7]*

Row	Toy_1	City_2	CostDog_101	CostDog_102	CostDog_103	CostKitten_101	CostKitten_102	CostKitten_103
1	1	Belfast	5	5.5	6.5	7.5	7.99	8.5
2	1	Cardiff	5	5.5	6.5	7.5	7.99	8.5
3	1	Glasgow	5	5.5	6.5	7.5	7.99	8.5
4	1	London	5	5.5	6.5	7.5	7.99	8.5
5	2	Belfast	11114	28044	2725	11115	14684	25871
6	2	Cardiff	5445	13664	25439	25702	34450	65965
7	2	Glasgow	17318	75982	32937	10160	20652	59921
8	2	London	51237	44271	113496	199155	181643	188046

Figure 3.30

Clicking OK produces the spreadsheet in Figure 3.30, where there are 3 columns of prices and sales for each year. As with the Stack menu, Genstat has given the columns default names (which can be changed using the Column Rename menu).

If you have data open in two spreadsheets, you can merge them together in different orders or at different levels of aggregation using the Merge menu. To illustrate this menu we will match together two sets of data where the data has been stored in different files. The files `Health1.gsh` and `Health2.gsh` contain some data carried out on university students. The file `Health1.gsh` contains measurements of their height, weight, age and gender, whilst the file `Health2.gsh` contains data on their pulse rates before and after exercising. Both files contain a column with the students ID, which will be used as an indicator to merge the spreadsheets. To merge the data from these files they both need to be open within spreadsheets in Genstat. So, we open both of them using Open from the File menu (Figures 3.31 and 3.32).

Spreadsheet [Health1.gsh]

Row	ID	Height	Weight	Age	Gender
1	1	173	57	18	female
2	2	179	58	19	female
3	3	167	62	18	female
4	4	195	84	18	male
5	5	173	64	18	female
6	6	184	74	22	male
7	7	162	57	20	female
8	8	169	55	18	female
9	9	164	56	19	female
10	10	168	60	23	male
11	11	170	75	20	male
12	14	187	59	18	male

Figure 3.31

Spreadsheet [Health2.gsh]

Row	ID	Exercise	Pulse1	Pulse2
1	1	mod	86	88
2	2	mod	82	150
3	3	high	96	176
4	5	low	90	88
5	6	low	78	141
6	7	mod	68	72
7	8	mod	71	77
8	9	high	68	68
9	10	mod	88	150
10	11	high	76	88
11	12	low	74	76
12	13	mod	70	71

Figure 3.32

Genstat warns that `Health2.gsh` has a column `ID` with the same name as a column in `Health1.gsh` (Figure 3.33). This is deliberate: `ID` is the column that will be used to merge the sheets. We can check the box Don't show this warning again to stop this appearing in future.

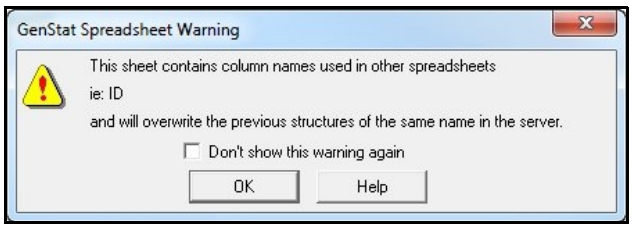


Figure 3.33

Clicking in the spreadsheet Health1.gsh and then selecting Merge from the Manipulation option on the Spread menu opens the menu shown in Figure 3.34.

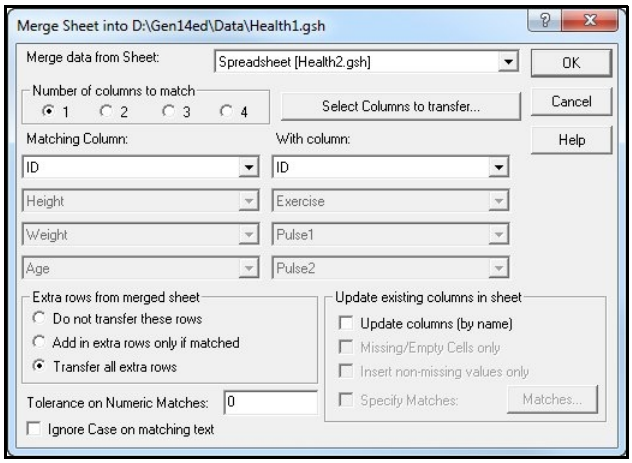


Figure 3.34

We wish to merge in the data from the spreadsheet Health2.gsh so we select this from the Merge data from Sheet list. Here, we simply wish to merge the two sheets using the student's ID, so we select ID from the Matching Column list and also from the With Column list. If we only wanted to merge a subset of columns from the sheet Health2.gsh, we could select them by clicking on the Select Columns to Transfer button. The options at the bottom of the menu allow you to control how the rows are updated in the spreadsheet and how to update existing columns in both spreadsheets.

Leaving the menu with the default settings produces the merged spreadsheet shown in Figure 3.35. Where a student ID was found in one sheet, but not the other, missing values are used to complete the row. For example, students 4 and 14 were found in the spreadsheet Health1.gsh, but not Health2.gsh so these rows have missing values inserted for the columns merged from the spreadsheet Health2.gsh. The reverse can be seen for the students 12 and 13.

Spreadsheet [Health1.gsh]*

Row	ID	Height	Weight	Age	Gender	Exercise	Pulse1	Pulse2
1	1	173	57	18	female	mod	86	88
2	2	179	58	19	female	mod	82	150
3	3	167	62	18	female	high	96	176
4	4	195	84	18	male		*	*
5	5	173	64	18	female	low	90	88
6	6	184	74	22	male	low	78	141
7	7	162	57	20	female	mod	68	72
8	8	169	55	18	female	mod	71	77
9	9	164	56	19	female	high	68	68
10	10	168	60	23	male	mod	88	150
11	11	170	75	20	male	high	76	88
12	14	187	59	18	male		*	*
13	12	*	*	*		low	74	76
14	13	*	*	*		mod	70	71

Figure 3.35

An alternative menu is available for merging data from multiple files where the files can be in different file formats. Selecting Merge Multiple Files from the New submenu on the Spread menu opens the menu shown in Figure 3.36. To illustrate this menu we will merge the two spreadsheets Health1.gsh and Health2.gsh.To select the two files we click on the '...' button, make a multiple selection of the two filenames and click OK.This places the names of the two files in the Filename list. We wish to merge the spreadsheets using the student's id, so we have entered ID into the Identifying Columns for Merge list. Clicking Open on this menu will produce a spreadsheet identical to the spreadsheet shown in Figure 3.35.

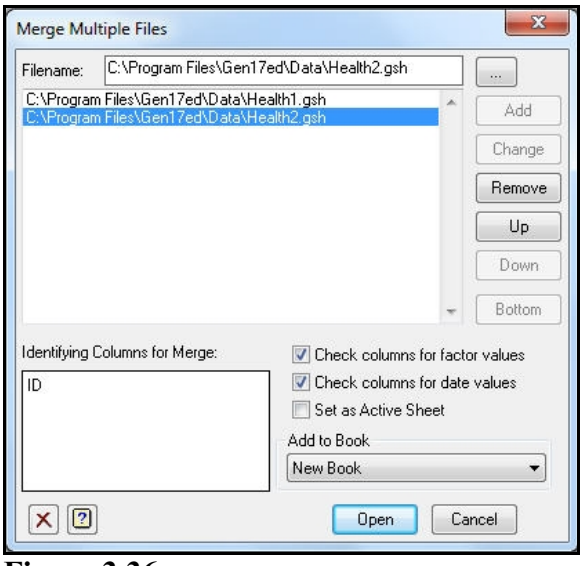


Figure 3.36

Other data manipulation methods available via the Manipulate options of the Spread menu include transposing, duplicating or converting spreadsheets.

3.6 Exercise

Experiments on cauliflowers in 1957 and 1958 provided data on the mean number of florets (y) in the plant and the temperature (x). Open the spreadsheet from file `Floret.GSH`, and stack the columns `y1` & `y2` together and `x1` & `x2` together.

Draw a scatter plot of the mean number of florets against the temperature. Redraw the graph, but this time enter the source factor (created from the stack) into the Groups box to highlight the two different groups. (Hint: click on the 2D Scatter Plot option of the Graphics menu on the menu bar, and follow the resulting wizard to plot the graph.)

4 Calculations and summaries

Once you have your observed data in a spreadsheet, you will often want to add calculated information to the spreadsheet. This may be calculating a new column from other columns, summaries across columns, creating new factors, or summarizing across by groups.

4.1 Calculating numerical columns

If you wish to calculate a new column using data from other columns, select the Column item in the Calculations submenu from the Spread menu. This allows you enter numerical expressions that will add a new column to the spreadsheet. If you open the Genstat spreadsheet file `Grazing.gsh` (which can be found in the `Data` folder), and then use this menu (Figure 4.1) you can calculate the change in live weight as the difference between the columns `Final_Lwt` and `Initial_Lwt`. To enter this expression, double-click Final_Lwt in the available data list, and then type a - (minus symbol) or click the - button and finally click `Initial_Lwt` in the available data list. To give a name to the column of results, click in the Save Result In box and type the name `Lwt_Change`. Clicking Run will create the column in the spreadsheet. If you use the Calculations item from the Spread menu then you will get the same dialog except that the Display in Spreadsheet and the current sheet may not be selected in the dialog. These will be set to the current spreadsheet if the menu is opened from this spreadsheet. The new column is added to the end of the spreadsheet, with the background column name in yellow.

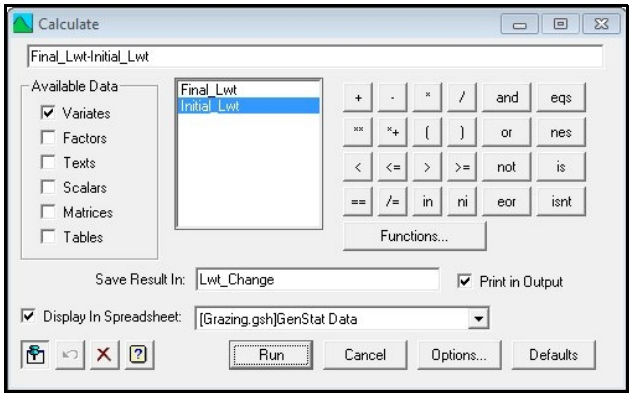
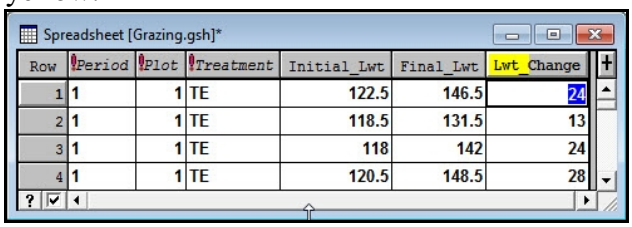


Figure 4.1

The resulting spreadsheet with the added column is shown in Figure 4.2.



Row	Period	Plot	Treatment	Initial_Lwt	Final_Lwt	Lwt_Change
1	1	1	TE	122.5	146.5	24
2	1	1	TE	118.5	131.5	13
3	1	1	TE	118	142	24
4	1	1	TE	120.5	148.5	28

Figure 4.2

If you wish to examine the calculation used for the column at a later date, then right clicking on the column and using the Column attributes menu item will show the expression.

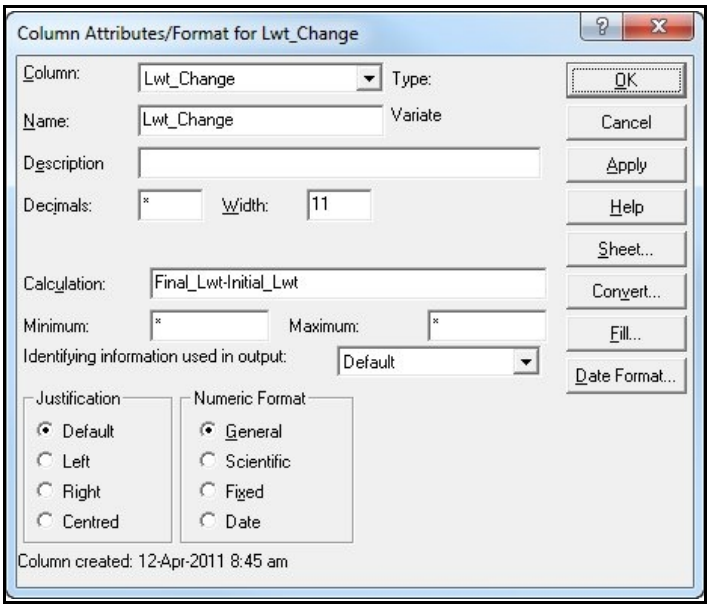


Figure 4.3

This dialog is shown in Figure 4.4. You may edit the calculation, so for example if the grazing period was four weeks and you wanted the calculation units to be grams per day, you could change the expression to

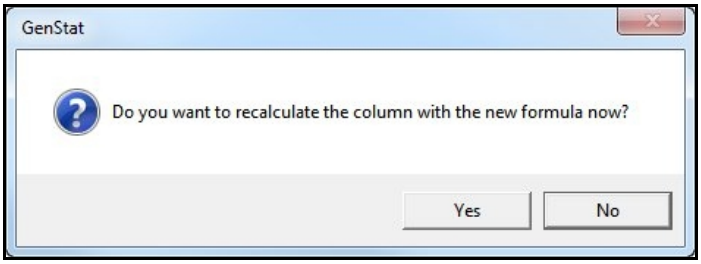


Figure 4.4

```
1000*(Final_Lwt-Initial_Lwt)/28.
```

If you change the expression, then when you click OK, you will get a prompt to update the columns values shown to the right. Also if you edit any values in columns used in the expression, you will be updated to update the calculated column's values with the dialog shown in Figure 4.5. The detection of required updates to calculated columns can be changed using the Spreadsheet Options item in the Tools menu. The item "Prompt to update calculated columns" on the General tab controls this behaviour.

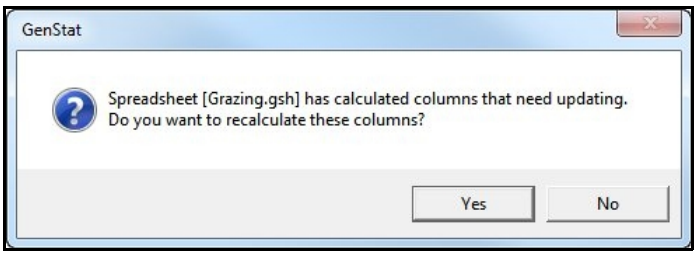


Figure 4.5

Individual cells can also be calculated in a one-off fashion (as the calculations are not saved) by entering an expression following an equals symbol in a spreadsheet cell (as in Figure 4.6).

When focus moves off the cell, the calculation will be evaluated and the numerical result put into the cell (as in Figure 4.7). Not all Genstat functions are supported. Also values of other cells in the spreadsheet can be used by using qualified column names such as `C1$[1]`.

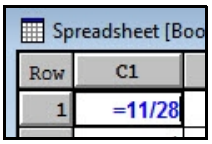


Figure 4.6

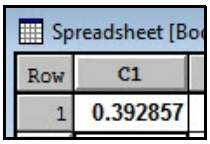


Figure 4.7

4.2 Creating text columns

There are two menus that allow you to create new text columns from other columns. These are the items Text Split and Combine Text in the Calculations submenu of the Spread menu. The first takes sections of text from an existing column and splits it into new columns, and the second combines text from columns into a new text column. Sometimes you may need to use combinations of these two menus to get the text you require. For example you could split a column into separate text columns to extract just the parts of the text you need, and then put these separate text items back into a single text column. These menus do not just work on text columns but can be used for the text displayed in a variate or factor column (e.g. the factor labels).

The spreadsheet file `Rhizotron Expt.gsh` (which can be found in the `Data` folder) contains the experimental layout for some electrophoresis gels that have been run on soil from 4 rhizotrons (containers for growing plant roots); see Figure 4.8.

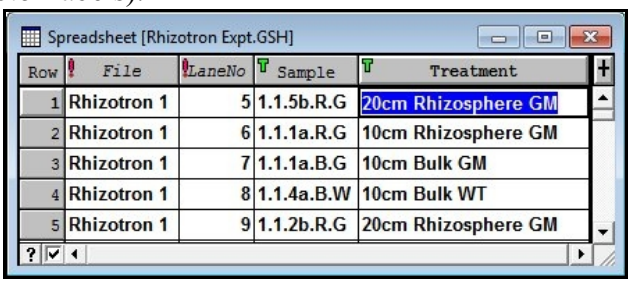


Figure 4.8

Set the focus on the `Treatment` column which contains text giving the depth, source and plant type (GM = genetically modified, and WT = wild type), and then use the Text Split menu to split this column into 3 parts. The sections of text are separated by spaces so we use this as the option for Split using item. We specify that the Number of splits to save is 3, and then give the 3 columns the names `Depth`, `Soil` and `Plant`. As we probably want to use these in either ANOVA or tabulation, we can select the option Convert saved columns to factors to make them into factors.

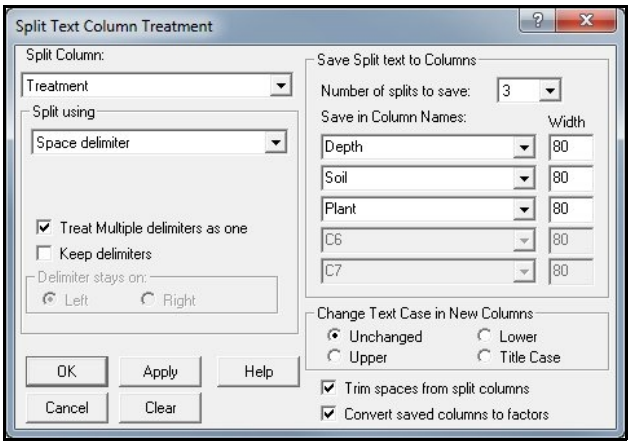


Figure 4.9

When you press the OK button, these columns are added to the spreadsheet as shown in Figure 4.10.

Treatment	Depth	Soil	Plant
20cm Rhizosphere GM	20cm	Rhizosphere	GM
10cm Rhizosphere GM	10cm	Rhizosphere	GM
10cm Bulk GM	10cm	Bulk	GM
10cm Bulk WT	10cm	Bulk	WT

Figure 4.10

The Combine Text menu (Figure 4.11) can be used to combine columns into a single text column. The columns that are combined can be of any type (text, variate or factor) and the currently displayed text in the column will be used. Using the same file as in the previous section (Rhizotron Expt.gsh) we can combine the information from the columns File and LaneNo to create a unique id for each row in the spreadsheet. We set the Number of columns to combine as 2, and select the two columns File and LaneNo To avoid having the text Rhizotron in the id, we can set the starting position as 10 for the File column. Setting the width as 1 will just take the first character from position 11. In this case it will not make any difference as there is just one character at there. We would get the same result if we use the start as position 10 with length 2, as the option Trim spaces before combining columns would have removed the leading space at position 10. We type a new column name ID in the Save into text column box. Clicking the OK button creates the new column ID in the spreadsheet as shown in Figure 4.12.

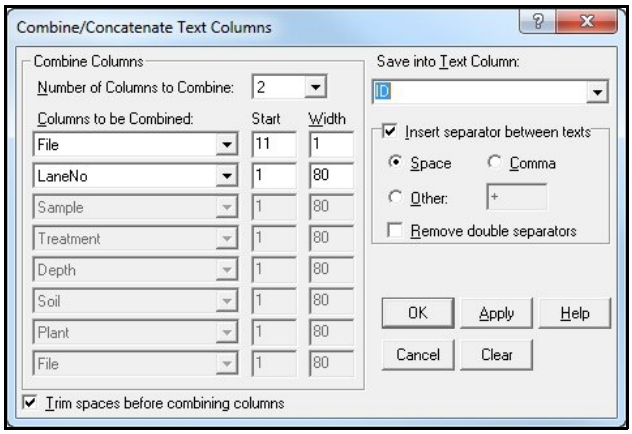


Figure 4.11

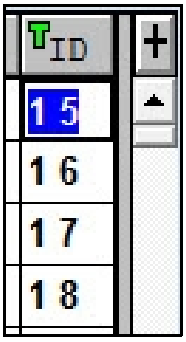
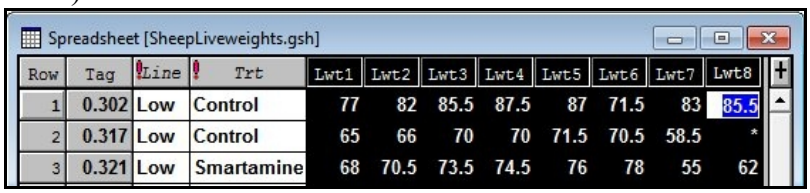


Figure 4.12

4.3 Summaries across rows

The item Row summaries in the Calculations submenu of the Spread menu allows you to form summaries across selected columns for each row in the spreadsheet. The summaries may be means, totals, minima, maxima, variances or standard deviations among others. The columns to be summarized across are selected by clicking their headings. If no columns are selected, you can summarize across all the columns or just columns of a certain type (variate or factor).

Open the Genstat spreadsheet file SheepLiveweights.gsh (which can be found in the Data folder). This contains eight live weights (Lwt1 - Lwt8) of sheep from two lines allocated to 2 treatment groups taken over a year. Now select the 8 live weight columns. You can speed this up by selecting the first one, and then holding down the Shift key while selecting the last one (as shown in Figure 4.13).



Row	Tag	Line	Trt	Lwt1	Lwt2	Lwt3	Lwt4	Lwt5	Lwt6	Lwt7	Lwt8
1	0.302	Low	Control	77	82	85.5	87.5	87	71.5	83	85.5
2	0.317	Low	Control	65	66	70	70	71.5	70.5	58.5	*
3	0.321	Low	Smartamine	68	70.5	73.5	74.5	76	78	55	62

Figure 4.13

Now open the Row summaries menu (Figure 4.14) to obtain the dialog to the right. Choose the Row summary to calculate as Mean and enter the Save in column name as MeanLwt.

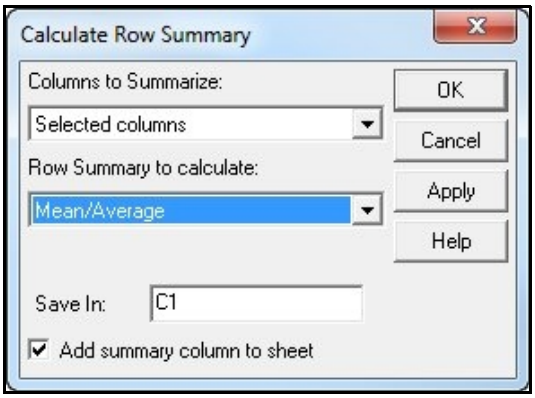


Figure 4.14

Clicking OK creates this column in the spreadsheet, as shown in Figure 4.15. If you have multiple summary statistics you wish to calculate, you can press the Apply button and then change the Row summary to calculate and the Save in items for each statistic you wish to calculate.

Lwt6	Lwt7	Lwt8	MeanLwt
71.5	83	85.5	82.375
70.5	58.5	*	67.3571
78	55	62	69.6875

Figure 4.15

Note if you right click the new column, and choose the Column Attributes item in menu, this will display the calculation that was used to create the column as shown in Figure 4.16.

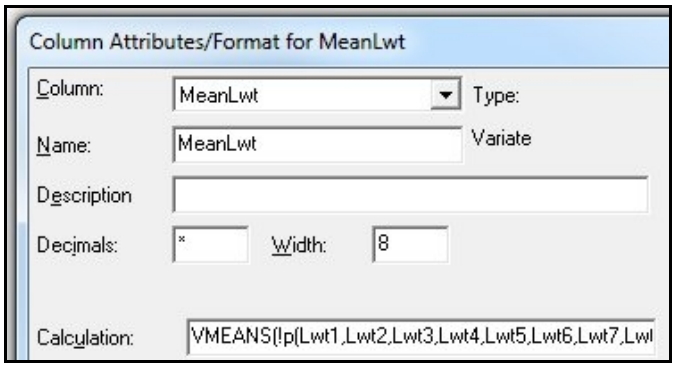


Figure 4.16

You can create a permanent pointer to the columns Lwt1-8 using the Create Pointer item in the Sheet sub-menu of the Spread menu. Select the 8 columns as previously and then this menu item will open the dialog show in Figure 4.17. You need to give a name to the pointer (here it's called Lwt). Now if you open the Calculate Row Summary dialog there will be an extra item in the drop-down list which will be this pointer as shown in Figure 4.18.

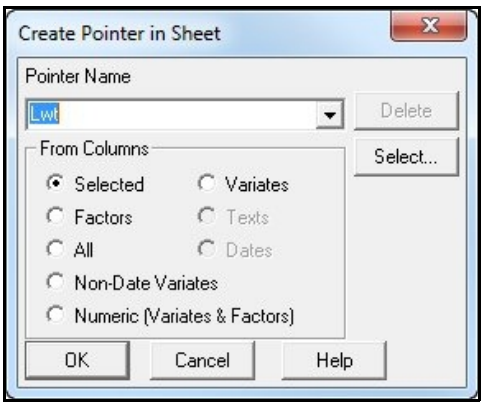


Figure 4.17

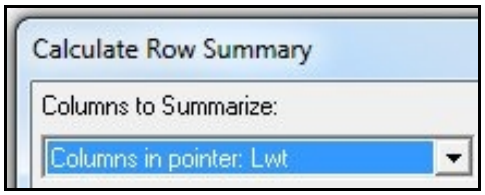


Figure 4.18

4.4 Forming Factors

If you have a variate that you wish to allocate to groups with values within a given ranges being allocated to groups then this can be done with the item Code to Groups in the Calculations submenu of the Spread menu. The spreadsheet file `New Zealand Income Survey.gsh` (which can be found in the `Data` folder) contains 200 randomly selected cases from a survey on personal income for New Zealanders; see Figure 4.19.

Spreadsheet [New Zealand Income Survey.GSH]

Row	Personid	Gender	Qualification	Age	Hours	Income	Marital	Ethnicity
1	1	female	school	15	4	87	never	European
2	2	female	vocational	40	42	596	married	European
3	3	male	none	38	40	497	married	Maori
4	4	female	vocational	34	8	299	never	European
5	5	female	school	45	16	301	married	European

Figure 4.19

If we want to categorize the variate `Age` into 3 groups for those < 18, 18-40 and 40+ then we put our cursor into a cell in the Age column and use the Code to Groups menu. This opens the dialog in Figure 4.20. We need to change the Number of Groups to 3 and then put in the breakpoints of 18 and 40. We do not need enter a value in the first cell of the Break Points column, but need to enter the values 18 and 40 in cells 2 and 3. If you do not edit the labels, then they will change to reflect the break points you have entered. The labels can be then changed to Youth for the < 18 group and 40+ for the >= 40 group. We set the Groups Column Name to `Age_Group`. Clicking OK If you untick the item Links are Lower Boundaries, then you put in the upper bound of each group, rather than the lower bound, starting from cell 1.

Code to Groups Column Age

Break Points	Levels	Labels	Counts
< 18	1	Youth	14
18	2	18 - 40	145
40	3	40+	41

Groups Column Name: Age_Group

Number of Groups: 3

Limits are Lower boundaries

OK | Cancel | Reset | = Ranges | = Counts

Figure 4.20

Age	Age_Group
15	Youth
40	40+
38	18-40
34	18-40
45	40+
45	40+
36	18-40
35	18-40

Figure 4.21

If you want to re-code individual items in a factor, variate or text to a new value then you need to use the item Recode in the Calculations submenu of the Spread menu. This gives a list of all the unique items in the column and then for each item you can specify what it's new value is to be. This can be used for changing values like 0 or missing value (*) to a new value, or to group levels of a factor together, or define sets of values for factor groups. Recoding the factor column `Qualification`, we can group the 4 levels (`none`, `school`, `vocational` and `degree`) into 2 levels `none/school` and `post school` into a new factor. To do this we click on the column we want to recode, and then use the Recode menu item. For each unique item in the column we provide a new value. It's easiest to type the first item in and then press Ctrl+C to copy it onto the clipboard. Then for each item that you want to recode to that value, click on its New Value cell or use the up/down arrows to put focus there, and press Ctrl+V to paste the value there. If the Recode to Numeric is ticked, then the new values will be numeric. When the Code as a Factor item is ticked, the resulting column will be a factor, otherwise it will be a text or variate depending on the setting of Recode to Numeric.

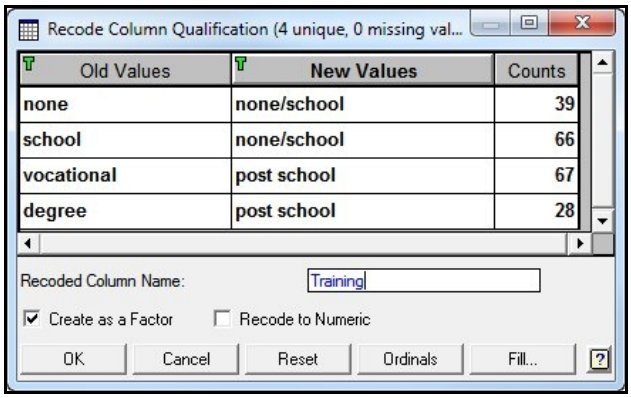


Figure 4.22

Qualification	Training
school	none/school
vocational	post school
none	none/school
vocational	post school
school	none/school
degree	post school

Figure 4.23

Figure 4.22 shows the dialog completed to group the 4 levels into 2 as specified previously. The new column name is set to training, and when you click OK, this will add the column to the spreadsheet just after the column it was created from, as in Figure 4.23.

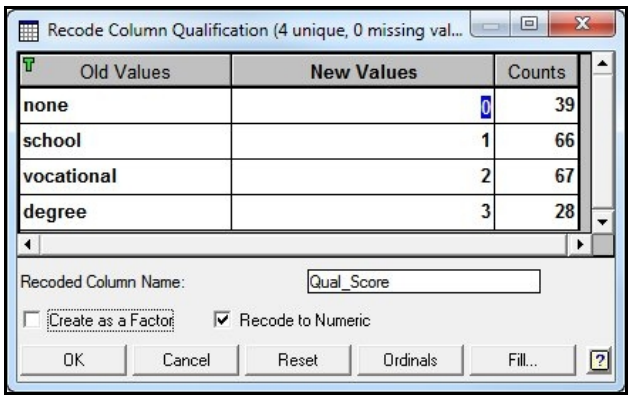


Figure 4.24

Qualification	Qual_Score
school	1
vocational	2
none	0
vocational	2
school	1
degree	3

Figure 4.25

If we wanted to create a variate from the `Qualification` column we could again use the Recode menu item, and tick Recode to Numeric and untick Code as a Factor. The dialog showing this is in Figure 4.24, where the resulting column is named `Qual_Score`. Clicking OK adds this column to the spreadsheet as shown in Figure 4.25.

If you have two or more factors and you want to create a factor that indexes all the combinations of these factors you would use the item Product/Combine in the Factor submenu of the Spread menu. It saves time to select the factors you wish to combine by clicking on their column headings before opening this menu, as this fills in the factors and creates a default name for the new column. So if we want to create a factor that indexes both Gender and Martial status, we can select these as in Figure 4.26, and then use the Product/Combine menu to get the dialog shown in Figure 4.27.

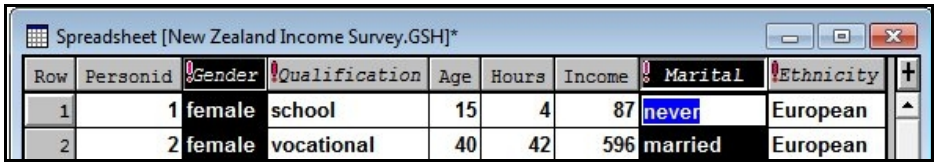


Figure 4.26

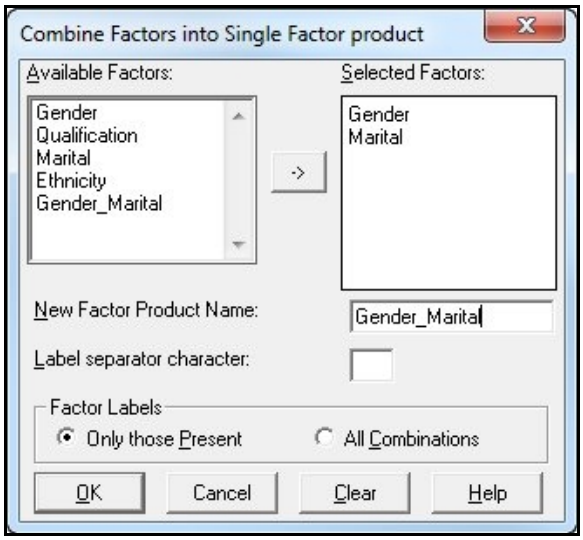


Figure 4.27

Gender_Marital
female never
female married
male married
female never
female married
male married

Figure 4.28

Clicking OK then creates this factor at the end of the spreadsheet as shown in Figure 4.28. The calculation that created the factor is stored in the column attributes. The yellow background indicates that this is a calculated column.

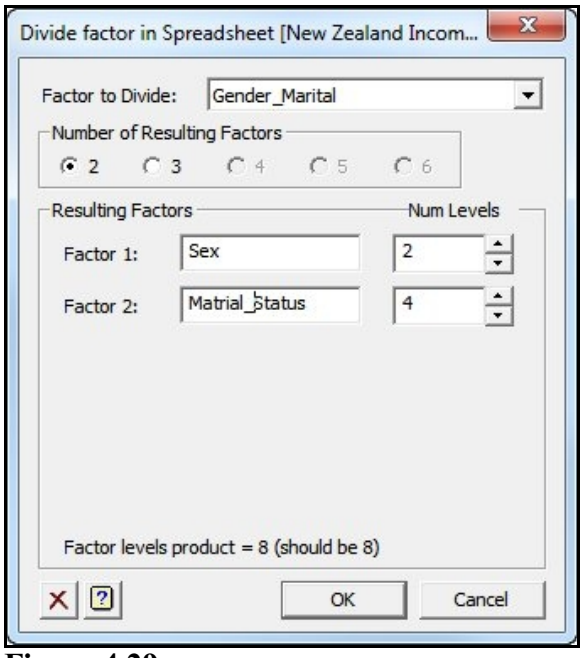


Figure 4.29

Gender_Marital	Sex	Matrial_Status
female never	1	1
female married	1	2
male married	2	2
female never	1	1
female married	1	2

Figure 4.30

The opposite action to combining factors is to divide a factor. The item Divide in the Factor submenu of the Spread menu allows you to spit up a factor into multiple component factors, provided the factor order follows a set order with all combinations present. To divide the factor we have just created, `Gender_Marital` back into its two components, we put the focus on this column and then use the Divide menu to open the dialog shown in Figure 4.29. We enter the names for the two new factors as `Sex` and `Martial_Status`. The first factor Sex has two levels and show we have to change the default value for Num Levels to 2 from the 4 that is entered by default. The value for the Num Levels for `Martial_Status` automatically changes to 2 from 4 as the product of these must be 8. Clicking OK then creates these two factors at the end of the spreadsheet as shown in Figure 4.30.

The factors do not have labels, so these would need to be manually entered using the item Edit Levels and Labels in the Factor submenu of the Spread menu. This dialog is shown for editing the factor column Sex in Figure 4.31.

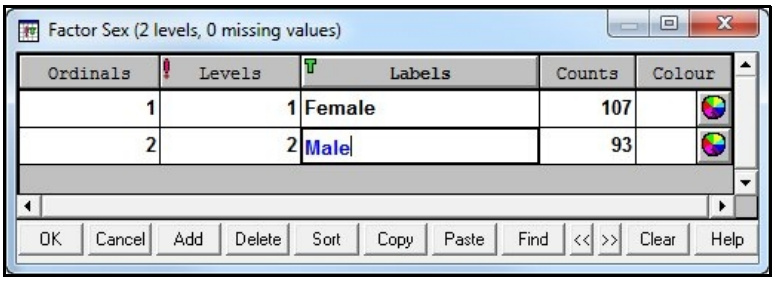


Figure 4.31

Sometimes you have factors in several spreadsheets, or two factors in one sheet that you want to use the same labels and levels in each, with the same order of labels. The item Standardize Levels in the Factor submenu of the Spread menu allows you to ensure that two factors have the same levels and labels in the same order. To illustrate this menu, we will open the two Excel files (Grazing 1.xls and Grazing 2.xls). Use the File Open menu to select these two files, and click Finish on the resulting dialogs to obtain the two sheets. Note the labels for Treatment in Grazing 2.xls are in lower case. Now opening the Standardize Levels menu gives the dialog in Figure 4.32. Selecting the two Treatment columns by double-clicking these, and then select the Case of Labels as Upper. Clicking OK standardizes the factor labels.

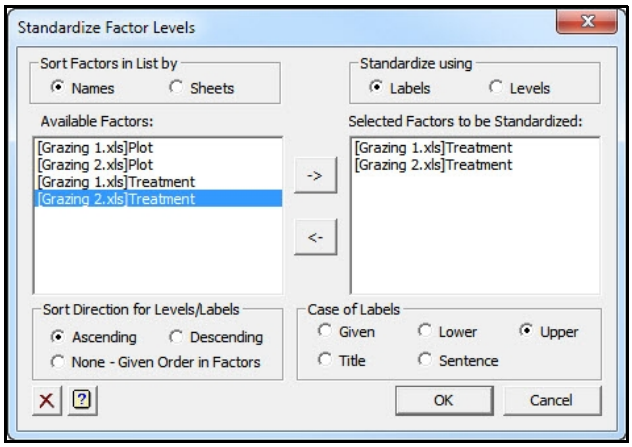


Figure 4.32

If you standardize with the Treatment factor in Grazing 3.xls, this gives the note in Figure 4.33 which indicates the factors were already harmonized.

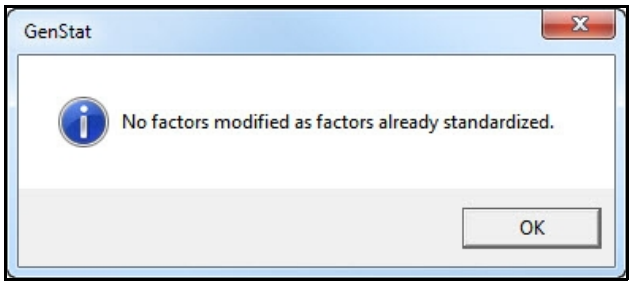


Figure 4.33

4.5 Summaries over groups

If you have a data that you wish to form summaries over groups and put these results into a vector spreadsheet, then this can be done with the item Summary Stats in the Calculation submenu of the Spread menu. If you want the results to be in a table rather than a vector spreadsheet, you would use the item Summary Tables in the Summary Statistics submenu of the Stats menu. Using the spreadsheet file New Zealand Income Survey.gsh (see Figure 4.19), if we want the means of Age, Hours and Income for the combinations of the Gender, Qualification and Marital factors and the counts in each group, we would open this dialog, as shown in Figure 4.34. To select the required summary, we would double-click the three factors in the Factors list or select these and click the Arrow -> button, then click the Counts button and finally select the three variates, Age, Hours and Income, and then click the Mean button. Clicking the OK button produces the spreadsheet in Figure 4.35, which contains the requested summaries by group. The columns are prefixed with a letter and then an underscore for each statistic (m_ for means etc.), and the factors with f_. This is to avoid the summaries overwriting the original data in the Genstat server. If the Merge into Original sheet option is ticked, then the summary columns would be merged back into the sheet the dialog was opened from, rather than being put into a new sheet.

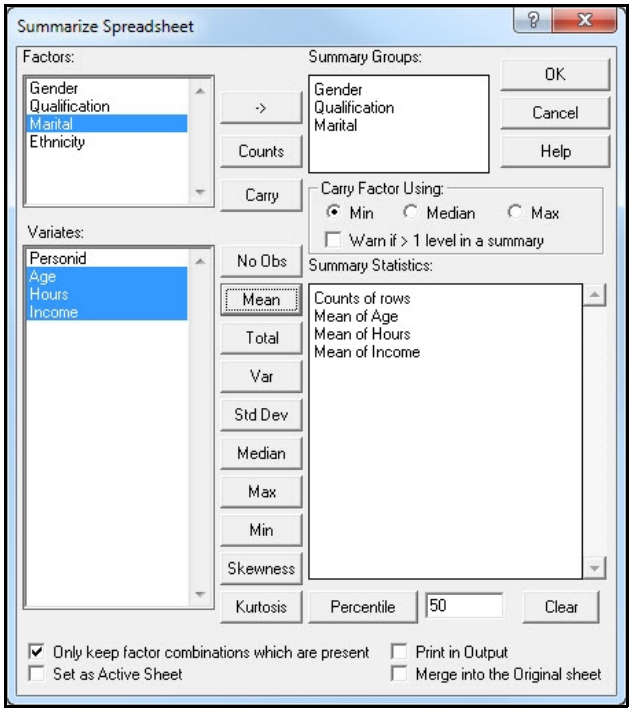


Figure 4.34

Spreadsheet [Book;2]*

Row	f_Gender	f_Qualification	f_Marital	r_Count	m_Age	m_Hours	m_Income
1	female	none	never	8	28.125	23.75	341.25
2	female	none	married	9	37	23.4444	337.111
3	female	none	previously	1	43	25	266
4	female	none	other	4	32.5	28.75	361.25

Figure 4.35

4.6 Subsets of rows

If you wish to remove rows, take a random subset of rows or split a sheet up into multiple sheets there are various ways of doing this. One way is to make a restriction to include just the rows required in the new data set, and then either delete the restricted rows using Restricted rows item of the Delete sub-menu of the Spread menu or duplicate the sheet with just the included rows in the new sheet using the Duplicate item of the Manipulate sub-menu of the Spread menu. It normally is safer to duplicate the spreadsheet keeping the original data, rather than just deleting the rows, as if you forget to save the spreadsheet into a new file the deleted rows data will be lost. If random selection of rows is required, then you can make a random restriction using the Random Rows item of the Restrict/Filter sub menu of the Spread menu or a random subset of rows using the Split/Subset item of the Manipulate sub-menu of the Spread menu.

Using the spreadsheet file `New Zealand Income Survey.gsh` (see Figure 4.19), we can take 100 random rows from the 200 in the spreadsheet as follows.

First we restrict to 100 random rows using Random Rows item of the Restrict/Filter sub menu of the Spread menu. The dialog for this is shown in Figure 4.36. Clicking OK filters the spreadsheet to include 100 of the rows. We can then use the Duplicate item of the Manipulate sub menu of the Spread menu to obtain the dialog shown in Figure 4.37. It is important that the item Only duplicate unrestricted data is ticked (normally if the spreadsheet is restricted, this will be ticked by default). Clicking OK creates the spreadsheet shown in Figure 4.38.

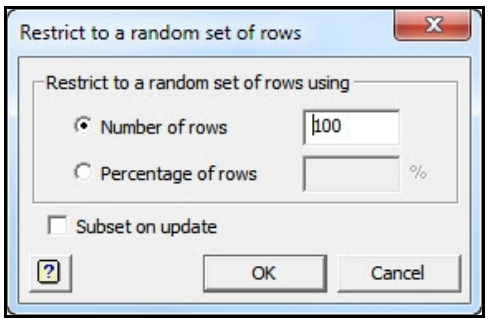


Figure 4.36

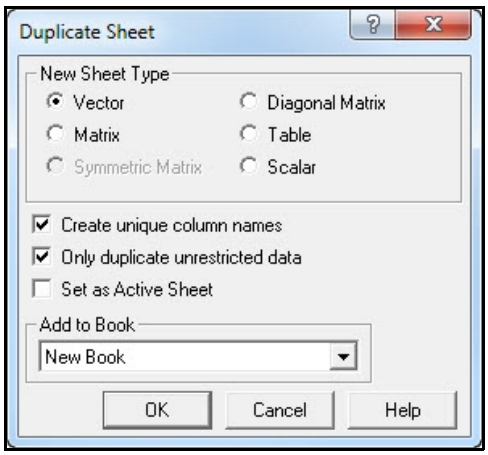


Figure 4.37

Spreadsheet [Book;2]*

Row	Personid_1	Gender_1	Qualification_1	Age_	Hours_1	Income_1	Marital_1
1	3	male	none	38	40	497	married
2	5	female	school	45	16	301	married
3	6	male	degree	45	50	1614	married
4	7	female	none	36	12	201	other
5	8	male	degree	35	45	934	previously

Figure 4.38

We can now use the Split/Subset item of the Manipulate sub-menu of the Spread menu. This gives the dialog in Figure 4.39. We need to select the Operation as Subset into a single sheet, and Split sheet using as Random sampling. The number of samples should be 100 and Weighting should be <Equal>. Clicking OK will create a sheet like that in Figure 4.38 but the names will end in _2 rather than _1 so that they are unique. If we had the Operation as Split into a multiple sheets, we would have obtained two sheets with a random 100 rows in one and the remaining 100 rows in the second sheet.

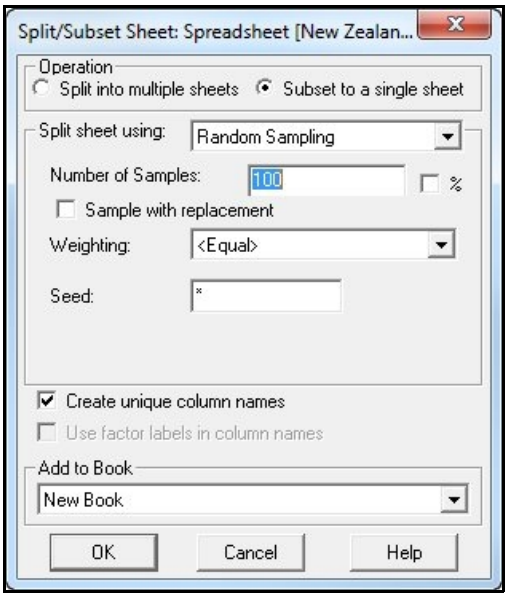


Figure 4.39

We can also use the Split/Subset menu to split the sheet into two data sets. If we wanted to have a sheet for males and another for females we could use this menu and complete the dialog as shown in Figure 4.40. Here we the Operation as Split into a multiple sheets, and Split sheet using as Factor Groups and the Factor as `Gender`. We select Factor Groups to keep as <All levels> so that each level of `Gender` will have its only sheet. If you just selected some of the levels, then only those selected will have sheets created for them. Selecting a single level would be equivalent to setting Operation as Subset into a single sheet. Clicking OK will create two sheets like that in Figure 4.41.

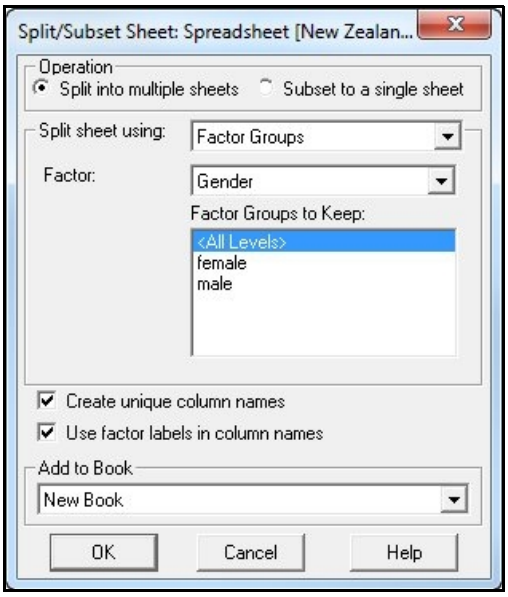


Figure 4.40

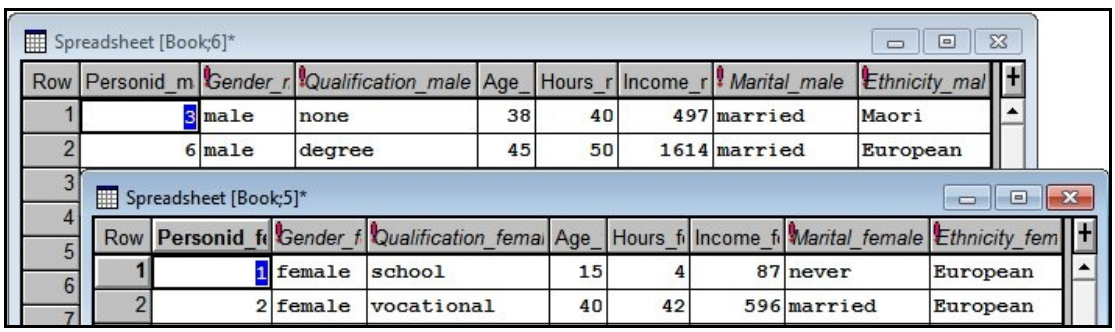


Figure 4.41

5 Spreadsheet tables

There are a number of menus that produce results in tables. A table is a set of values that are classified by a set of factors. A Genstat spreadsheet can contain multiple one-way tables (provided they all have the same classifying factor) or a single table with two or more classifying factors. If a table has three or more factors, one of these factors can be distributed across the tabs in a book to give a tabbed-table. Tables can be manipulated in a number of ways within the spreadsheet. The following section will explain how to manipulate tables within a spreadsheet.

5.1 Creating tables from Genstat menus

Several of the Genstat Statistics menus (e.g. the Frequency and Summary Tables items in Summary Statistics submenu or the items in the Analysis of Variance submenu of the Stats menu) can create tables. These menus generally have a Store button that opens a dialog to specify what to save and whether to display this in a spreadsheet. We will look at creating some summary tables from the Summary Statistics menu using the spreadsheet file `New Zealand Income Survey.gsh` (see Figure 4.19). Open the spreadsheet file and then the Summary Statistics menu. This will give the dialog shown in Figure 5.1. Now if we want the means of `Income` over the `Gender` groups, we fill in the menu as shown (by double-clicking `Income` and then `Gender`). Now click the Store button which opens the dialog in Figure 5.2. We complete this to store the tables for the means, standard deviations and counts in 3 tables. In the Store dialogs we have selected and give a Genstat structure name for each item we are storing from the analysis. To display these saved structures in one or more spreadsheets, we tick the Display Tables in Spreadsheet option. Then we close the Summary Table Store Options dialog by clicking OK. The tables are not saved until we click the Run button on the top menu. Some menus have Save buttons rather than Store buttons and these are enabled after the main menu has been run, and then create the structures when the Save dialog is closed. The Store button options need to be specified before the main menu is run and so do not create the structures when they are closed. So now clicking the Run button gives us the spreadsheet in Figure 5.3. Note how we have the three one-way tables in the same spreadsheet. We want to have some overall summaries across both genders, (margins), we need to tick the Set Margin option. Setting this and clicking the Run button gives us the spreadsheet in Figure 5.4. Note how the margin cells are shaded. The colour used for this can be changed using the Options item in the Tools menu, and on the Fonts and Colours tab, select the Spreadsheett item in the Show Settings list, and then select the Margins Background item in the Display Items list to specify a new colour.

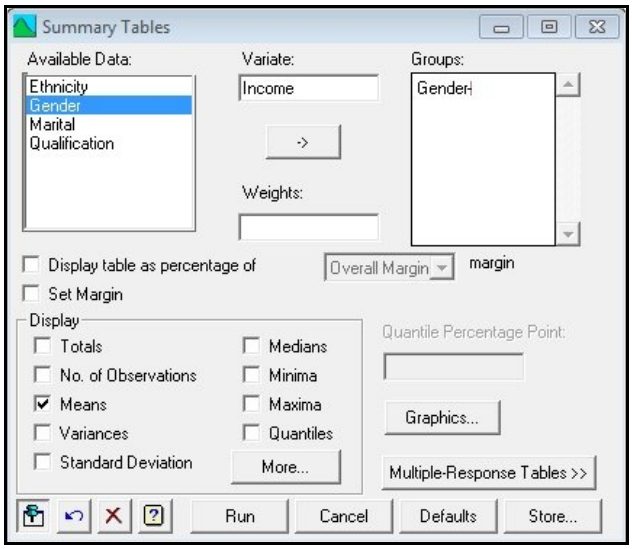


Figure 5.1

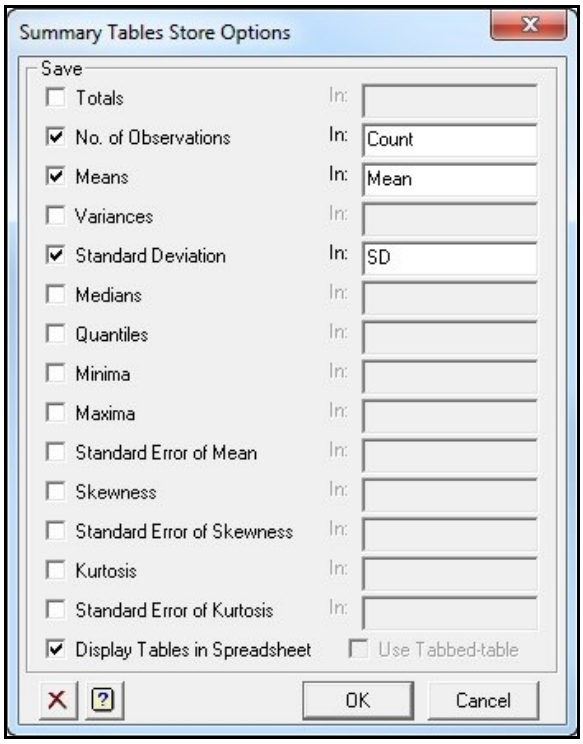


Figure 5.2

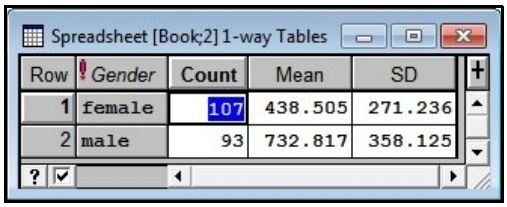


Figure 5.3

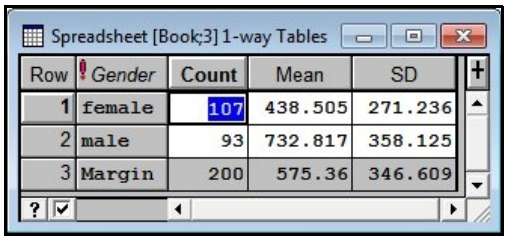


Figure 5.4

Now if we want a cross tabulation of `Gender` by `Martial` status we can add to the Groups box by double-clicking it in the Available Data list. This will now create a two-way table. A spreadsheet can only hold a single two or more way table so now clicking run will give us three spreadsheets for the counts, means and standard deviations. The resulting spreadsheets are shown in Figure 5.5.

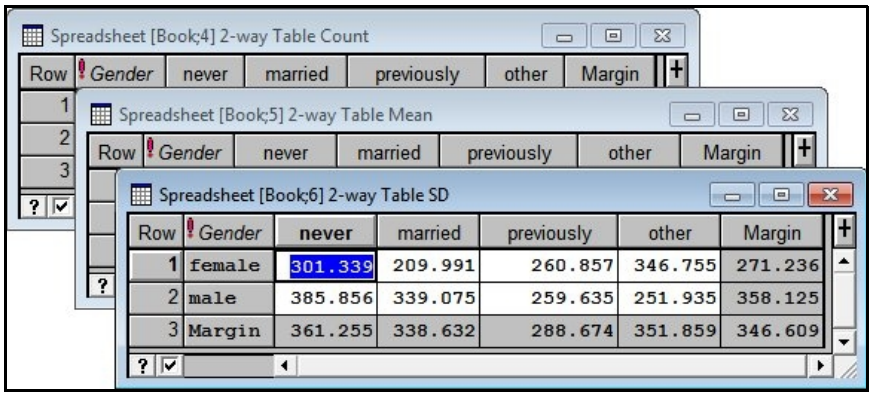


Figure 5.5

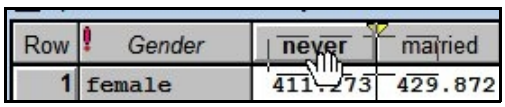


Figure 5.6

If you would prefer the table to display with the martial status groups down the columns rather than across the rows you can reformat the table by either dragging the Gender column across to a position after one of the martial status columns (as shown in Figure 5.6) or alternatively by dragging the '`never`' column before the Gender column.

This gives the spreadsheet displayed in Figure 5.7.

Spreadsheet [Book;8] 2-way Table Mean*

Row	Marital	female	male	Margin
1	never	411.273	629.455	520.364
2	married	429.872	831.613	607.786
3	previously	525.917	797.333	642.238
4	other	479	833.333	630.857
5	Margin	438.505	732.817	575.36

Figure 5.7

If you want to change the number of decimal places in a column, use the Edit Attributes item in the Column submenu of the Spread menu and set this. To do this for all the columns at once, select them by clicking on their headings before using this menu, and you will obtain a dialog that will allow you to change the attributes of the all selected columns (as in the dialog shown in Figure 5.8).

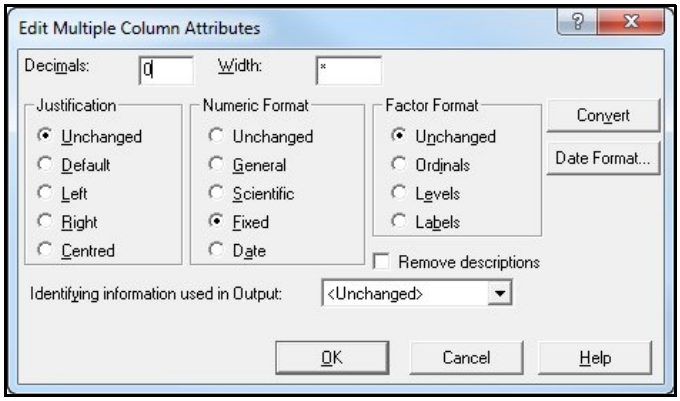


Figure 5.8

Tables with up to 9 classifying factors can be displayed in a spreadsheet. To create a 3 way table with the extra classifying factor `Qualification` we can double-click this it in the Available Data list to add it to the Groups box. Clicking run will create 3 spreadsheets again (the means table is shown in Figure 5.9). Now the first two classifying factors are in the first two columns and the last factor is across the rows. If you wish to suppress the display of some margins you can do this with a filter.

Spreadsheet [Book;11] 3-way Table Mean

Row	Gender	Marital	none	school	vocational	degree	Margin
1	female	never	341.25	270.111	518	813	411.273
2	female	married	337.111	385.846	506	562.5	429.872
3	female	previously	266	369.5	479.25	883.333	525.917
4	female	other	361.25	301.25	777	772	479
5	female	Margin	339.773	322.077	523.229	779.182	438.505
6	male	never	381.571	470.333	811.7	1171.83	629.455
7	male	married	669.714	783	813.5	1059.29	831.613
8	male	previously	609	533	829	886.75	797.333
9	male	other	569.5	*	908.714	*	833.333
10	male	Margin	535.706	530.556	835.219	1058.41	732.817
11	Margin	never	360.067	377.923	640.375	1028.3	520.364
12	Margin	married	482.625	496.167	642.667	948.889	607.786
13	Margin	previously	437.5	402.2	629.143	885.286	642.238
14	Margin	other	430.667	301.25	879.444	772	630.857
15	Margin	Margin	425.179	407.364	672.239	948.714	575.36

Figure 5.9

The rows to be hidden can either be individually selected, and then the Set As Excluded rows item in the Selected Rows sub-menu of the Restrict/Filter sub-menu of the Spread menu used. Alternatively if all the margins of a factor were to be hidden, you could put the focus on the cell showing Margins and use the Values not equal to the current cell item in the Restrict/Filter sub-menu of the Spread menu to hide all the margins of this factor. Figure 5.10 shows the table with the margins of `Gender` hidden (and the number of decimals places for the columns set to 0).

Spreadsheet [Book;14] 3-way Table Mean*

Row	Gender	Marital	none	school	vocational	degree	Margin
1	female	never	341	270	518	813	411
2	female	married	337	386	506	563	430
3	female	previously	266	370	479	883	526
4	female	other	361	301	777	772	479
5	female	Margin	340	322	523	779	439
6	male	never	382	470	812	1172	629
7	male	married	670	783	814	1059	832
8	male	previously	609	533	829	887	797
9	male	other	570	*	909	*	833
10	male	Margin	536	531	835	1058	733

Figure 5.10

Figure 5.11

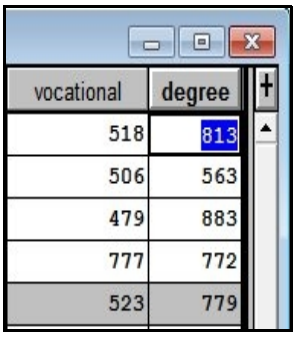


Figure 5.12

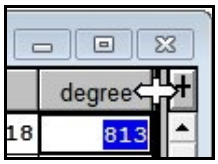


Figure 5.13

Columns in a spreadsheet can be hidden with the Hide/Show item in the Column sub-menu of the Spread menu. If you open this menu this will give you the dialog shown in Figure 5.11. You just need to double-click any column in the list to change its attribute from Show to Hide or vice-versa. Here in the dialog the Margin for `Qualification` is hidden. Figure 5.12 shows the table spreadsheet with the margin column hidden. When there is a hidden column the dividing line between columns is thicker. If you double-click the divider between the two shown columns where the column is hidden the columns between will be shown again. Figure 5.13 shows this being done for the final column to redisplay the hidden margin column.

Rows and columns in tables can be moved by dragging them with the mouse. The factor defining the table has to be renamed so that the original factor that was used to create the table is not redefined with a different label order.

If a table in a spreadsheet does not have margins, they can be calculated at a later date using the Table Margins in the Calculate sub-menu of the Spread menu. This opens the dialog shown in Figure 5.14. Note if there are not equal numbers of observations in the cells making up the table, then some summary statistics (e.g. means) may not be the same as the margins from the Summary Statistics menu as all cells are given equal weighting when forming margins through this menu.

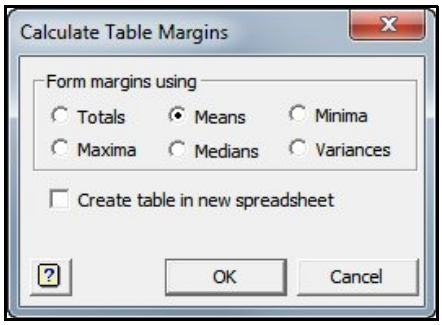


Figure 5.14

5.2 Tabbed-tables

If you have three or more classifying factors in a table, then one of the factors can have its groups displayed across the tabs in a spreadsheet book. You can do this either by specifying that a tabbed-table be created in one of the summary menus or by using the Tabbed-table from Genstat item in the New submenu of the Spread menu. Using the first approach, with the data in the previous section and the three classifying factors of `Gender`, `Martial` and `Qualification`, we use the Store button and Select the Tabbed-table option as shown in Figure 5.15. Clicking Run on the Summary Tables menu now gives us the spreadsheet shown in Figure 5.16 which was the genders across the tabs. The first group in the Groups box is used across the tabs.

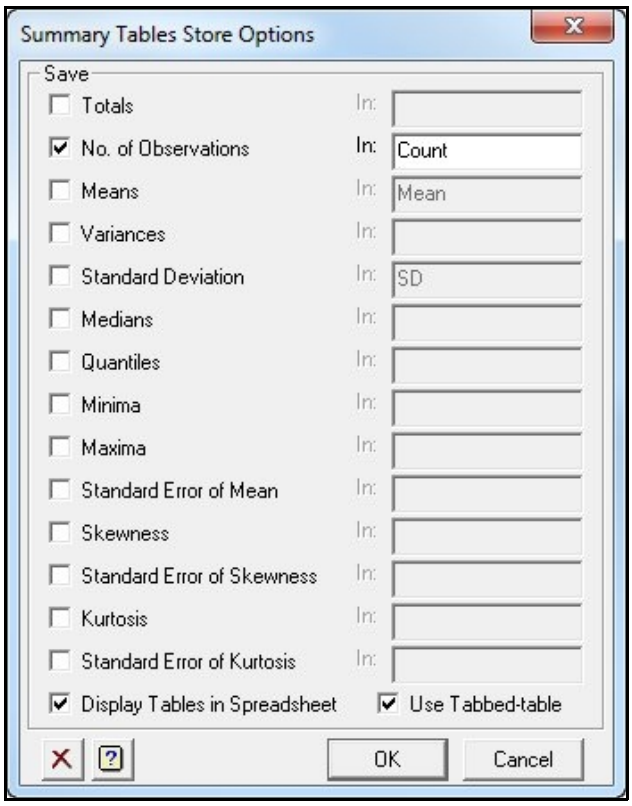
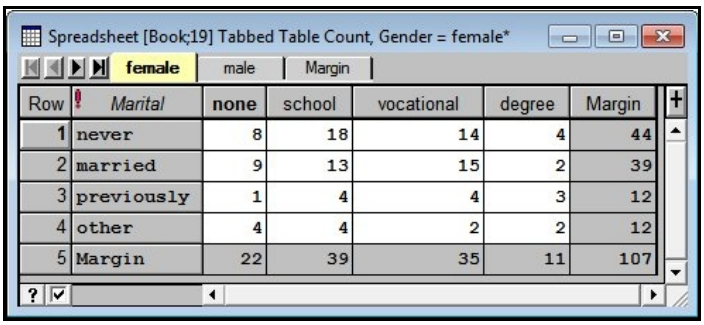


Figure 5.15

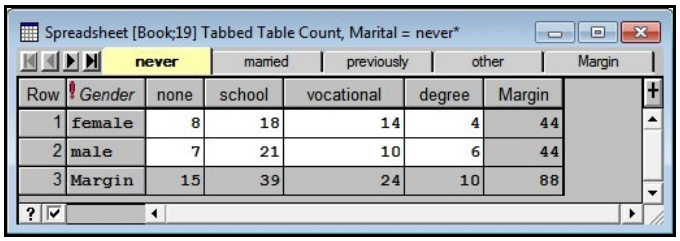


Row	Marital	none	school	vocational	degree	Margin
1	never	8	18	14	4	44
2	married	9	13	15	2	39
3	previously	1	4	4	3	12
4	other	4	4	2	2	12
5	Margin	22	39	35	11	107

Figure 5.16

If you want to reorder the classifying factors in the table, then you can use the Reorder table item in the Manipulate submenu of the Spread menu. This opens the dialog shown in Figure 5.17. You can use the Up and Down button to reorder the factors in the table. In a tabbed-table the first margin is shown across the tabs. Clicking OK creates the table shown in Figure 5.18.

Figure 5.17

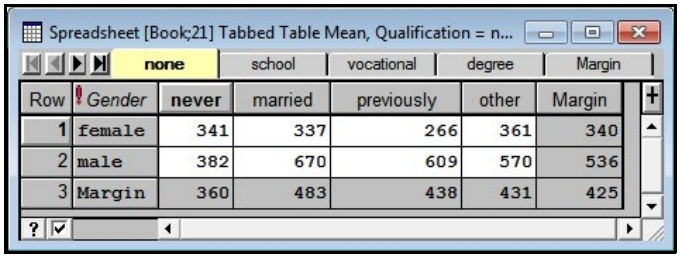


Row	Gender	none	school	vocational	degree	Margin
1	female	8	18	14	4	44
2	male	7	21	10	6	44
3	Margin	15	39	24	10	88

Figure 5.18

The second method uses the Tabbed-table from Genstat item in the New submenu of the Spread menu. This opens the dialog shown in Figure 5.19 which displays all the tables 3+ way tables in the Genstat server. One of these is double-clicked in the Available Tables list to put it in the Selected table box. Then moving the focus to the Factor across tabs gives a list of the factors in Genstat, and double-clicking an item in the Available factors list puts its groups across the tabs in the book. You must be careful to specify a factor that is a classifying factor of the table, otherwise you will get an error. Clicking OK on this menu filled in as shown gives the spreadsheet displayed in Figure 5.20.

Figure 5.19



Row	Gender	never	married	previously	other	Margin
1	female	341	337	266	361	340
2	male	382	670	609	570	536
3	Margin	360	483	438	431	425

Figure 5.20

The other use of tabbed-tables is to put multiple 2+ way tables into a single table. If we selected the Multiple 2+ way tables in the Create Tabbed-table Book dialog, then we can put the statistics in the `Count`, `Mean` and `SD` tables into a single tabbed-table as shown in Figure 5.21. Clicking the OK button then provides the table in Figure 5.22 with the various statistics across the tabs. This now has all the values from the three tables in a single Genstat table with an extra classifying factor for the Statistics (`Tabs_1`).

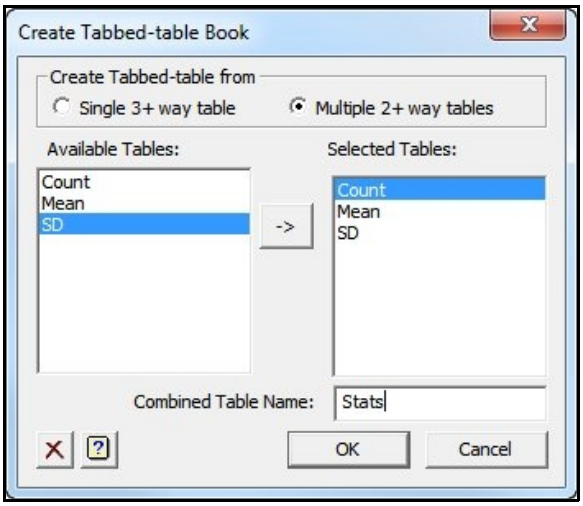
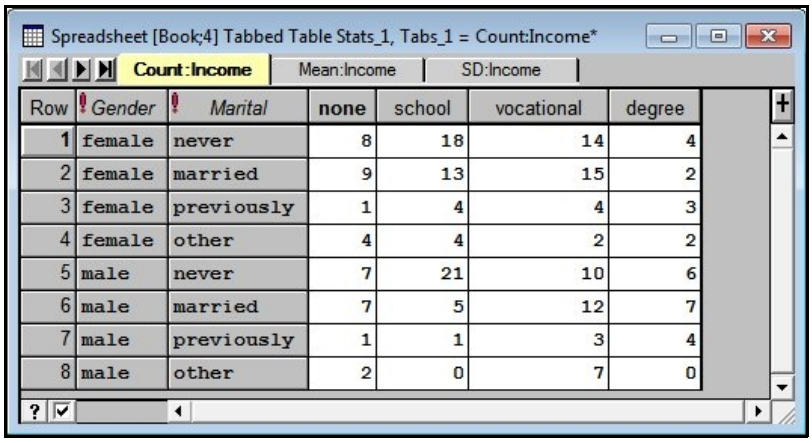


Figure 5.21



Row	Gender	Marital	none	school	vocational	degree
1	female	never	8	18	14	4
2	female	married	9	13	15	2
3	female	previously	1	4	4	3
4	female	other	4	4	2	2
5	male	never	7	21	10	6
6	male	married	7	5	12	7
7	male	previously	1	1	3	4
8	male	other	2	0	7	0

Figure 5.22

We can now use this table with other spreadsheet menus to display the statistics side by side in rows or columns. For example, using the Reorder Table menu shown previously allows us to display statistics side by side. Setting the factor order as shown in Figure 5.23 and then clicking OK gives the spreadsheet displayed in Figure 5.24.

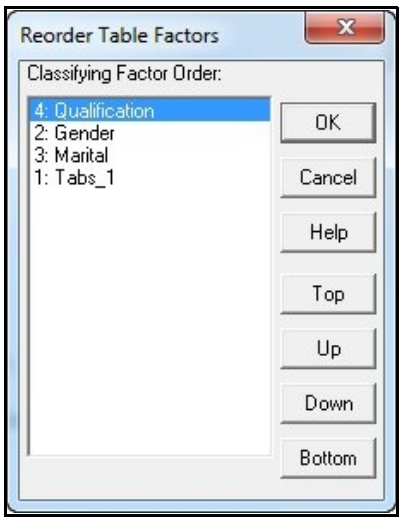
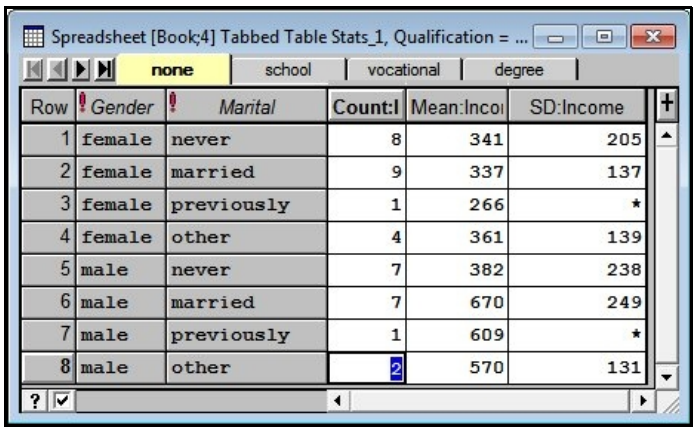


Figure 5.23



Row	Gender	Marital	Count:I	Mean:Inco	SD:Income
1	female	never	8	341	205
2	female	married	9	337	137
3	female	previously	1	266	*
4	female	other	4	361	139
5	male	never	7	382	238
6	male	married	7	670	249
7	male	previously	1	609	*
8	male	other	2	570	131

Figure 5.24

Putting multiple statistics into a single table can also be done in the Summary Tables menu if multiple statistics are selected in the Store dialog and the Display Tables in Spreadsheet and Use Tabbed-table options are both ticked as in Figure 5.25. This generates the same spreadsheet as in Figure 5.24. The combined table name is not specified by the user, but defaults to `Table_1`.

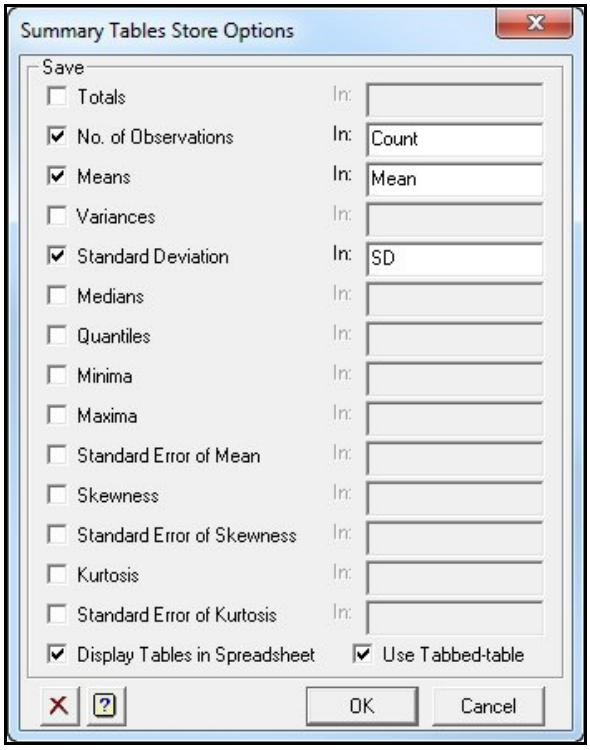


Figure 5.25

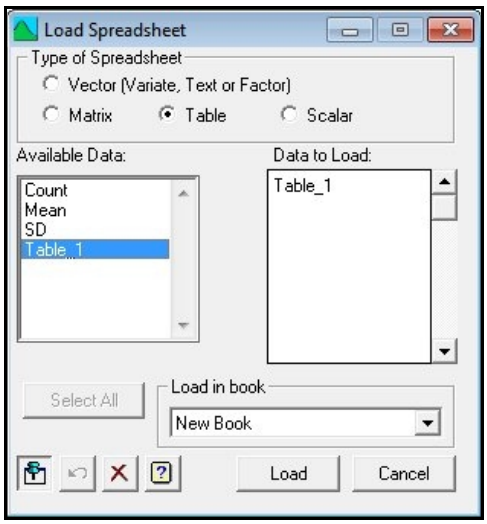


Figure 5.26

You can now put this combined table in a single page spreadsheet table if you use the Data in Genstat item in the New sub-menu of the Spread menu. Select the Table item in the Type of Spreadsheet group box and then double-click `Table_1` in the Available Data list (as shown in Figure 5.26) and then click the Load button. This will give the spreadsheet as shown in Figure 5.27. As before you, can drag the classifying factors around to redisplay the table in different ways. Renaming the labels in the `Tabs_1` column to remove the `:Income` part of each label and then moving the `Tabs_1` column to a position after the `Martial` column gives the table shown in Figure 5.28.

Spreadsheet [Book;7] 4-way Table Table_1*

Row	Tabs_1	Gender	Marital	none	school	vocational	degree
1	Count:Income	female	never	8	18	14	4
2	Count:Income	female	married	9	13	15	2
3	Count:Income	female	previously	1	4	4	3
4	Count:Income	female	other	4	4	2	2
5	Count:Income	male	never	7	21	10	6
6	Count:Income	male	married	7	5	12	7
7	Count:Income	male	previously	1	1	3	4
8	Count:Income	male	other	2	0	7	0
9	Mean:Income	female	never	341	270	518	813
10	Mean:Income	female	married	337	386	506	563
11	Mean:Income	female	previously	266	370	479	883
12	Mean:Income	female	other	361	301	777	772
13	Mean:Income	male	never	382	470	812	1172
14	Mean:Income	male	married	670	783	814	1059
15	Mean:Income	male	previously	609	533	829	887
16	Mean:Income	male	other	570	*	909	*
17	SD:Income	female	never	205	217	258	497
18	SD:Income	female	married	137	159	267	26
19	SD:Income	female	previously	*	228	45	122
20	SD:Income	female	other	139	383	75	537

Figure 5.27

Spreadsheet [Book;7] 4-way Table Table_1*

Row	Gender	Marital	Qualification	Count	Mean	SD
1	female	never	none	8	341	205
2	female	never	school	18	270	217
3	female	never	vocational	14	518	258
4	female	never	degree	4	813	497
5	female	married	none	9	337	137
6	female	married	school	13	386	159
7	female	married	vocational	15	506	267
8	female	married	degree	2	563	26
9	female	previously	none	1	266	*
10	female	previously	school	4	370	228
11	female	previously	vocational	4	479	45
12	female	previously	degree	3	883	122
13	female	other	none	4	361	139
14	female	other	school	4	301	383
15	female	other	vocational	2	777	75
16	female	other	degree	2	772	537
17	male	never	none	7	382	238
18	male	never	school	21	470	285
19	male	never	vocational	10	812	250
20	male	never	degree	6	1172	404
21	male	married	none	7	670	249
22	male	married	school	5	783	251
23	male	married	vocational	12	814	286
24	male	married	degree	7	1059	476
25	male	previously	none	1	609	*
26	male	previously	school	1	533	*
27	male	previously	vocational	3	829	289
28	male	previously	degree	4	887	278
29	male	other	none	2	570	131
30	male	other	school	0	*	*
31	male	other	vocational	7	909	228
32	male	other	degree	0	*	*

Figure 5.28

6 Bookmarking and comments

Sometimes it is useful to insert place holders into your text windows or spreadsheets. This is particularly useful if you have a large spreadsheet or text file open, and want to quickly go to a particular cell or line. To bookmark a text window or spreadsheet you can use the Bookmark option on the Search menu. For spreadsheets there is an additional menu available that allows you to bookmark particular numerical values. To illustrate the bookmark facilities we will open the sulphur data, stored in the Genstat spreadsheet file `Sulphur.gsh` (which can be found in the `Data` folder).

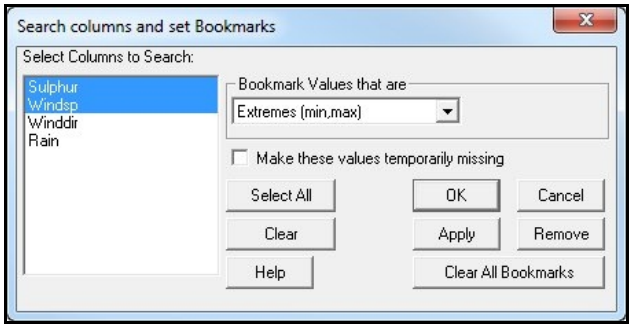
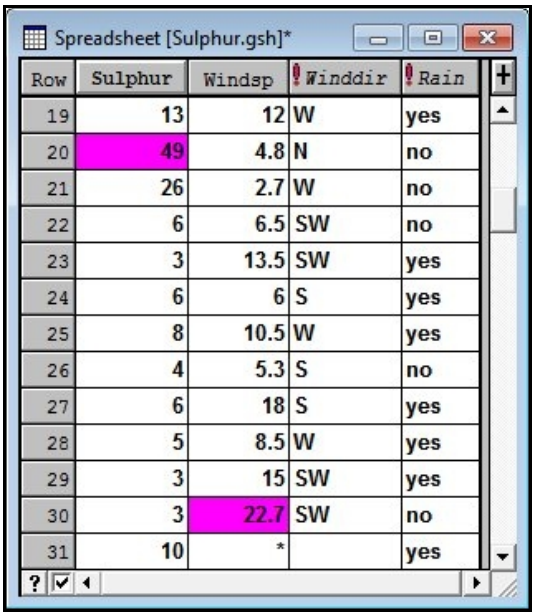


Figure 6.1

In this example we will bookmark the maximum and minimum values within the columns `Sulphur` and `Windsp`, so that these values can be identified quickly. We select By Value from the Bookmark options on the Edit menu, which opens the menu in Figure 6.1. Here we have selected both `Sulphur` and `Windsp` from the Select Columns list, and the Extreme values (max, min) from the Bookmark Values list.

Clicking OK produces the spreadsheet in Figure 6.2, where the bookmarked cells are shown in a user-defined colour (by default magenta). You can navigate to these cells by selecting Next on the Bookmark option on the Edit menu. Each time you select this menu option, the cursor will move to the next bookmark within the spreadsheet.



Row	Sulphur	Windsp	Winddir	Rain
19	13	12	W	yes
20	49	4.8	N	no
21	26	2.7	W	no
22	6	6.5	SW	no
23	3	13.5	SW	yes
24	6	6	S	yes
25	8	10.5	W	yes
26	4	5.3	S	no
27	6	18	S	yes
28	5	8.5	W	yes
29	3	15	SW	yes
30	3	22.7	SW	no
31	10	*		yes

Figure 6.2

You can add comments to individually bookmarked cells in a spreadsheet by adding a note. Clicking in the bookmarked cell within the column `Sulphur` at row 20, and selecting Add Note from the Bookmark option on the Edit menu opens the menu in Figure 6.3.

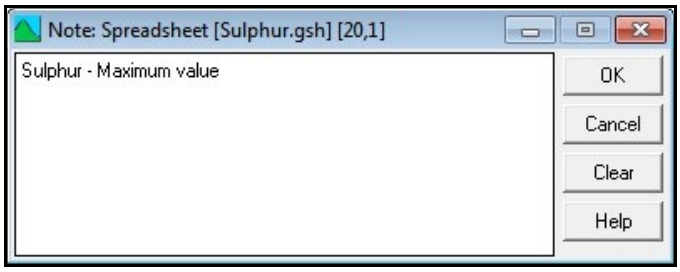


Figure 6.3

This displays a small, resizable text editor where a comment can be supplied for the bookmarked cell; by default this supplies a note based on the option selected from the Bookmark Values list. In this example the default note specifies that the cell is the maximum value for `Sulphur`. If a spreadsheet containing bookmarks is saved into a Genstat spreadsheet file, the bookmarks will be retained when the file is opened again.

Another feature, useful for visually displaying values that fall into different categories or conditions, is the Conditional Formatting menu. To open this menu select Conditional Formatting from Column on the Spread menu. The menu (shown in Figure 6.4) allows you to set up to 3 different conditions to format the data. You can also choose the colour in which to display the data that satisfies each of the conditions. In Figure 6.4 we have set different colours to represent different amounts of Sulphur in the air (small values in red and larger amounts in blue or green). Note that we have set the condition for greater than or equal to 20 before greater than or equal to 10. This is to ensure that the values greater than 10 but less than 20 are shown in blue. Clicking OK would redisplay the values within the column Sulphur in the chosen colours.

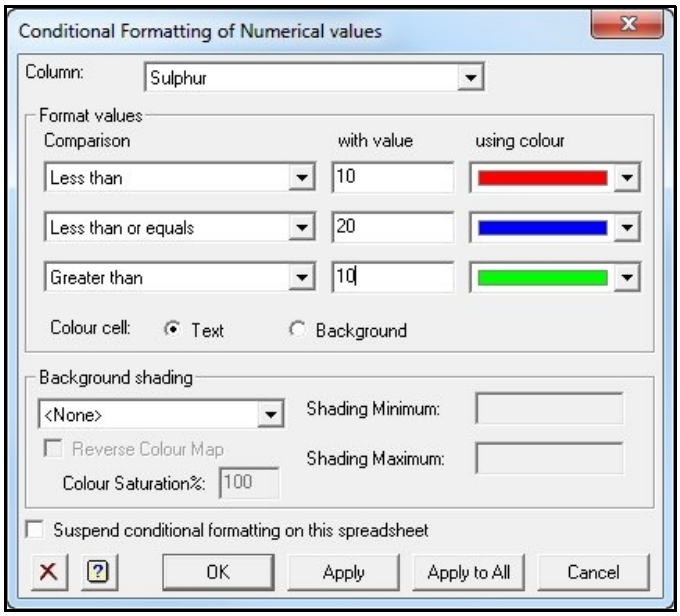


Figure 6.4

Individual spreadsheet cells can also be made temporarily missing (the value is retained in the spreadsheet but is set to missing in any calculations or statistical analysis), so that individual values can be excluded from an analysis. An example of where this could be useful is in an Analysis of Variance where, if you restricted out a row you could get a fault that the design is unbalanced. Analysis of Variance is discussed further in Chapter 6 of the Introduction to Genstat for Windows Guide (the menu to open this in the Guides submenu of the Help menu). In Chapter 3 of the introduction guide, when the values were transformed to logarithms, Genstat produced a warning that it could not calculate the logarithm of 0. So, this is a case, where we may wish to make this value temporarily missing before making a transformation.

To make this value temporarily missing select Temporary missing values from Column on the Spread menu. This opens the menu shown in Figure 6.5, where we have selected Sulphur and entered row 1 (where the value 0 is located).

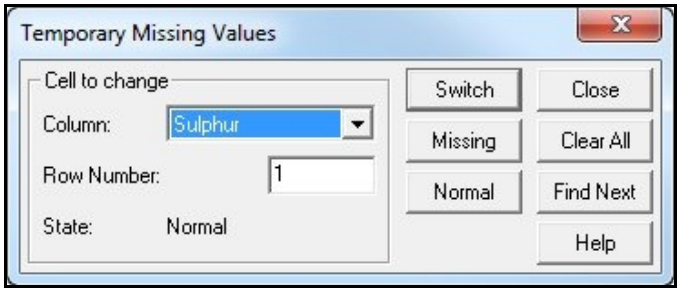


Figure 6.5

Clicking on the button Missing changes the cell to be temporarily missing, and clicking OK produces the spreadsheet in Figure 6.6. The temporary missing cell is disabled and has an asterisk appended to the value in the cell.

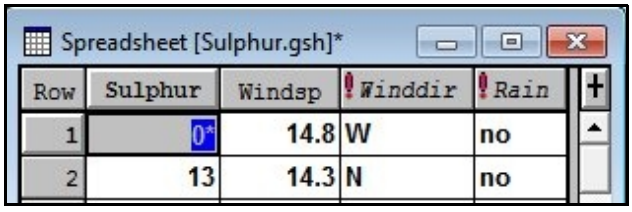


Figure 6.6

Alternatively, you can either toggle the status of the current cell with the Alt+F8 key, or click the Temporary Missing button on the toolbar (Figure 6.7).

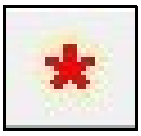


Figure 6.7

6.3 Exercise

The file `Ant.xls` contains data from an insecticide trial for killing ants. Five types of insecticide were used on each of three types of bait. The data has been entered on different sheets in the Excel file. Open the sheet Baits 1 & 2 from the file and then append the data from the Sheet Bait 3 to make the complete data set.

Bookmark the column time to show the maximum and minimum values (You can access the bookmark menu using the By Value option in the Bookmark section of the Spread menu). Clear these bookmarks.

Using the Conditional Format menu highlight the values for `insecticide` 2 in blue and the values for `insecticide` 4 in red. Clear the conditional formatting.

7 Working with spreadsheet books

Within Genstat you can have multiple spreadsheets contained within a single *book*. Each spreadsheet is contained on a separate page in a book and is known as a *sheet*. Each page has an associated tab displayed above the sheet containing the name of the sheet. However, if a book only contains one sheet, then no tab is displayed. For example, each of the Genstat spreadsheet files opened in previous chapters is opened as a single paged book and hence no tab is displayed.

To illustrate how to create a new spreadsheet in a book we open the file `Health1.gsh`. In this example we have some additional information about the year of the experiments that we wish to store on another sheet for the students 1,10 and 14. To add a new sheet in a book we select the Add Sheet option from the Book submenu on the Spread menu which opens the menu shown in Figure 7.1.The data in our example will be in two columns of variates with 3 rows, so we have selected the Vector spreadsheet icon, and entered 2 columns and 3 rows. We wish to create the new sheet in the book containing the existing health data, so to do this we select the file `Health1.gsh` from the Create in Book list.

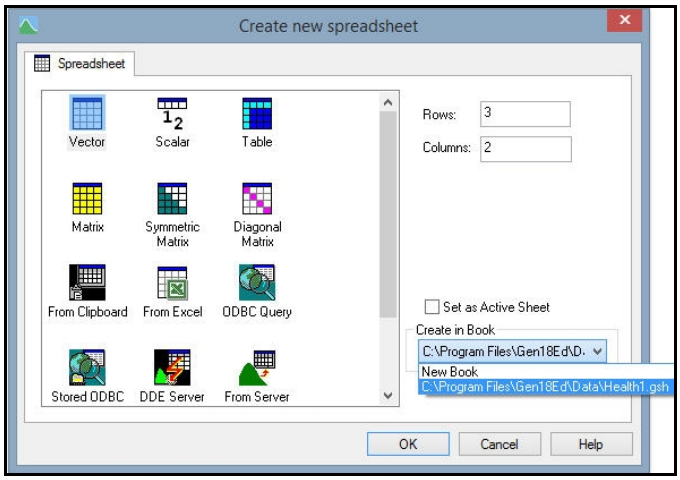


Figure 7.1

Clicking OK adds a new sheet to the book called `Sheet2` as shown in Figure 7.2, where we have entered the data and renamed the columns to `ID` and `Year`. Within a book the current sheet is identified by the highlighted tab. In Figure 7.2 this is `Sheet2`.You can change to display another sheet within a book by clicking on the tab for the sheet or by clicking the arrow navigation buttons on the top left of the window.

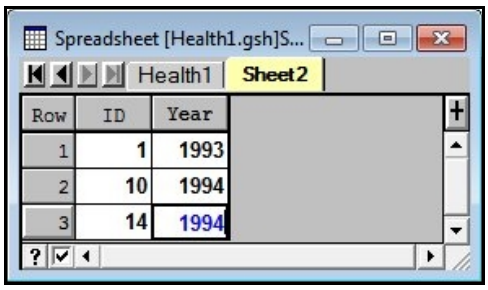


Figure 7.2

Alternatively, you can switch between sheets by selecting the Display Sheets item from the Book submenu on the Spread menu. This produces the menu shown in Figure 7.3 where we have highlighted `Health1` as the sheet that we now wish to display.

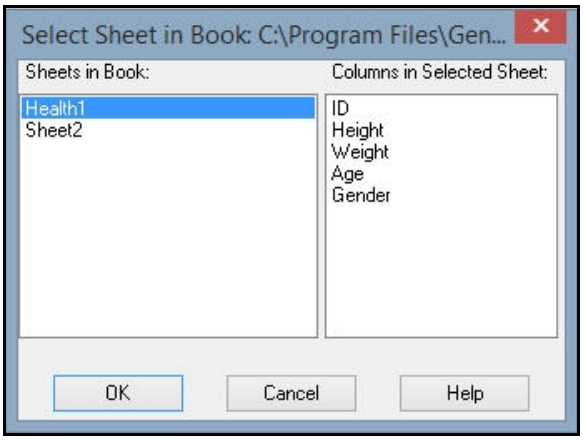


Figure 7.3

When a new sheet is added to a book, it is given a name by default: for example, Sheet1 for the first sheet, Sheet2 for the second sheet, and so on. You can change the name of a sheet by double-clicking on a tab. Figure 7.4 shows the menu that is opened after double-clicking on the `Sheet2` tab. Here we are changing the name of the sheet to `Year`.

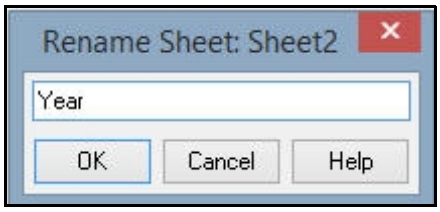


Figure 7.4

Spreadsheets can be moved or copied into books. To illustrate this we will add the data from the file `Health2.gsh` into a new sheet in the current book. To do this, we first open the file `Health2.gsh`, and then select the Move Current Sheet item from the Book submenu on the Spread menu. This opens the menu shown in Figure 7.5. In this menu, we have selected the Health1.gsh book as the one to which the data from `Health2.gsh` are to be moved. Clicking OK closes the `Health2.gsh` spreadsheet, and creates a new sheet in the book called `Health2` that contains the data from `Health2.gsh` as shown in Figure 7.6.

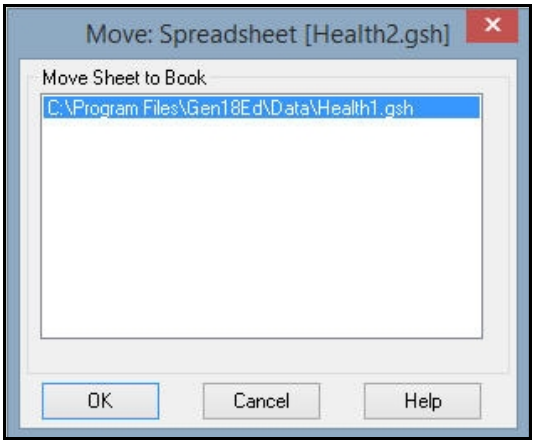


Figure 7.5

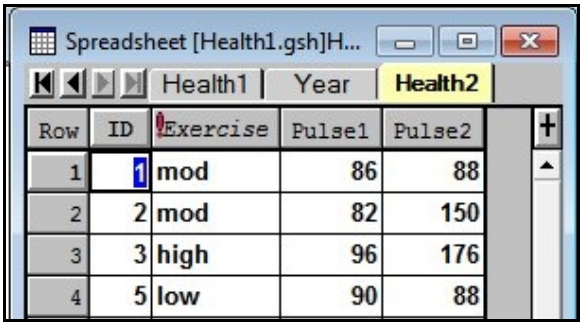


Figure 7.6

Sheets can be deleted from books using the delete menu available by selecting the Delete Sheet item from the Book submenu of the Spread menu. Alternatively, you can use the mouse to delete sheets. Click on a tab using the left button and hold the mouse down. Drag the cursor onto the sheet and, when it changes its appearance to a hand with a red cross, release the mouse (see Figure 7.7). Clicking Yes on the confirmation dialog will then delete the sheet.

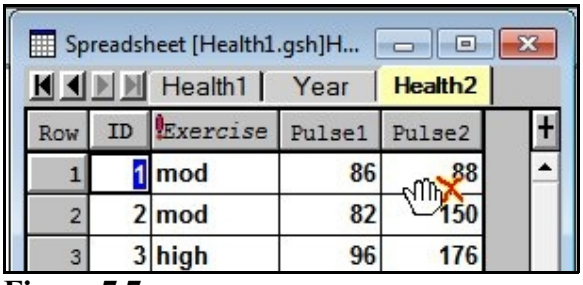


Figure 7.7

Similarly, sheets within a book can be reordered using the mouse. Click on the tab for the sheet that is to be moved using the left button, and hold the mouse down. Drag the cursor along the tabs. Its appearance will change to show a hand with a small sheet, and a small yellow triangle will appear above the tabs indicating the position where the sheet will be placed. When the yellow triangle is in the required position, release the mouse. Figure 7.8 shows `Health2` being dragged into the position between the other two sheets. An alternative way to reorder the sheets within a book is to use the menu available by selecting the Reorder Sheets item from the Book submenu on the Spread menu.

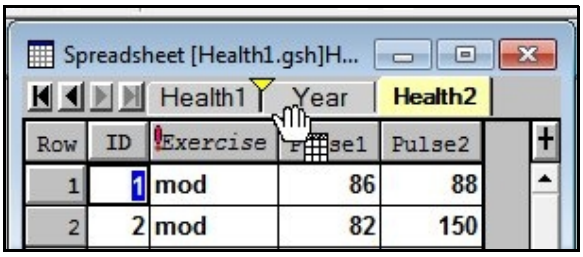


Figure 7.8

Sheets can be split into new books using the mouse or menus. To split the sheet called `Health2` into a new book, we select the Split item from the Book submenu of the Spread menu. This opens the menu shown in Figure 7.9. Here we have selected the sheet `[Health1.gsh]Health2`, and clicked on the '->' button to move this name into the Selected Sheets list. The Action for Selected Sheets option is set to Move, so that the sheet will be removed from the book.

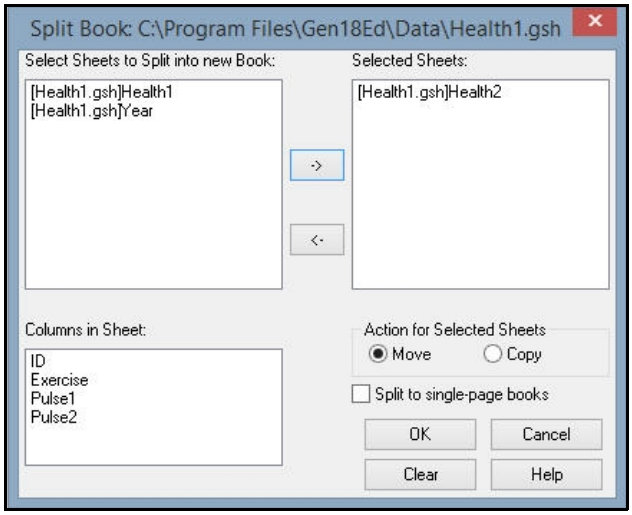


Figure 7.9

Clicking OK splits the book into two books (see Figure 7.10).

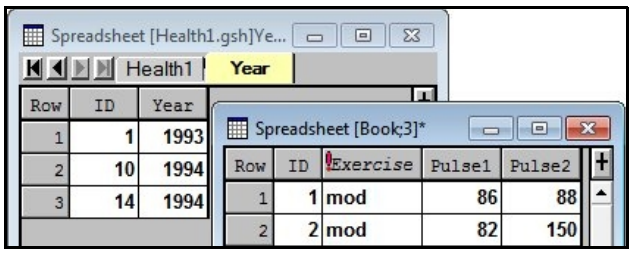


Figure 7.10

Sheets can also be split from books by left-mouse clicking on a tab, dragging the cursor off the spreadsheet and releasing the mouse (see Figure 7.11).

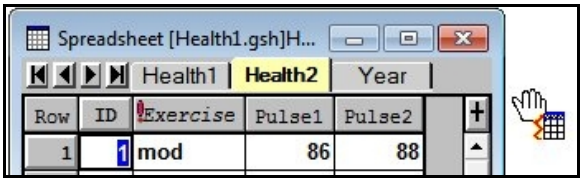


Figure 7.11

Other facilities that are available when working with books include the ability to open multiple files directly into a book. Also, some menus include an Add to Book or Create in Book option that allows you to create new sheets within either existing books or new books.

The best way to save Genstat books is as a Genstat Book file (*.gwb). In this format all the sheets are saved to the file together with any associated information. Also, Genstat books can be saved directly into multiple worksheets within an Excel file (*.xls). Alternatively, individual sheets from a book can be saved in a Genstat Spreadsheet (*.gsh) file.

8 Reading and writing data to databases

Genstat has facilities for reading and writing to databases using Open DataBase Connectivity (ODBC). ODBC is a Microsoft standard to allow a common method of accessing databases made by different software packages. The ODBC interface is built into Windows, and the common ODBC drivers are installed as standard in all Windows versions from Windows 95 second edition onwards. Genstat is able to query any data source that has an ODBC interface. This includes all main database systems (Access, Oracle, Informix, SQL Server, dBase, FoxPro, Paradox) and many spreadsheets (Excel, Quattro etc.). It is possible to use ODBC to read a data file from a package that is not even installed on your PC.

An ODBC link can be defined using either the ODBC/Data Sources Applet within the Control Panel or when you initially start an ODBC Data Query within Genstat. We will demonstrate how to create a link using the ODBC Data Query facilities within Genstat. Selecting ODBC Data Query from New on the Spread menu opens the menu shown in Figure 8.1. This menu shows all the current ODBC connections currently available on your PC. Connections to databases using ODBC are made by creating Data Source Names (DSN). A DSN stores all the information about how to connect to the data source and is stored permanently on a PC once it has been created. There are three types of DSN available, and the DSN you choose will depend on how you want to access to the database. The three types of DSN are as follows:

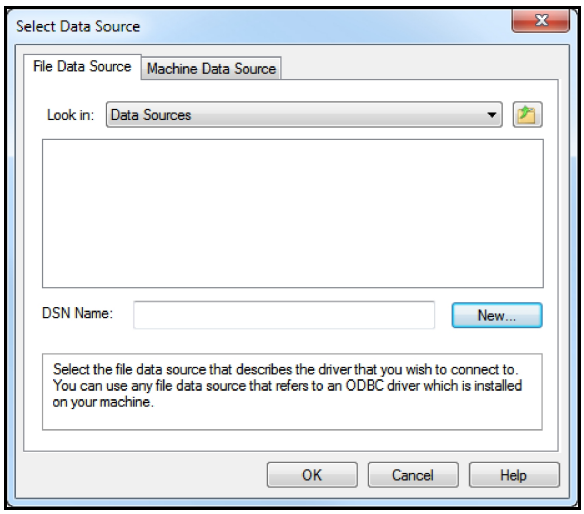


Figure 8.1

1 **User DSN** - This type of DSN can only be accessed by the current user who initially created it. So any other user (i.e. with a different username and password) working in the same PC will not be able to access the database.
2 **System DSN** - This type of DSN is specific to a computer. So any user of the computer will have access to the database using this type of DSN.
3 **File DSN** - This type of DSN is created as a file (*.dsn), which can be copied to any computer. Anyone who can access the file containing the DSN information can then access the database.

Note: If you are using a 64-bit version of Windows, there are two versions of ODBC available: 32-bit `C:\Windows\SysWoW64\Odbcad32.exe`, and 64-bit `C:\Windows\System32\Odbcad32.exe`. The 32-bit version of Genstat will see only the 32-bit version of ODBC and the 64-bit version of Genstat will see only the 64-bit version of ODBC. If you have a 32-bit version of Office installed (the default even on a 64-bit version of Windows), the Office ODBC drivers will be in the 32-bit version of ODBC. Therefore, if you are using the 64-bit version of Genstat you will not be able to see the 32-bit Office ODBC drivers for Access or Excel. On a 64-bit version of Windows, 32-bit Genstat by default installs in `C:\Program Files (x86)\Gen18Ed` and 64-bit version in `C:\Program Files\Gen18Ed`. In the following examples, the 32bit version of Genstat has been used, but if using the 64-bit version of Genstat with 64-bit Office, the directories used in the examples will change as explained above (i.e. `C:\Program Files\Gen18Ed` will be used rather than `C:\Program Files (x86)\Gen18Ed`).

On the menu in Figure 8.1 the file DSN's are listed under the File Data Source tab and the User and System DSN's are listed under the Machine Data Source tab. We will now illustrate how you can create a File DSN for a MS Access data file. The file we will connect to is called `cardata.mdb` and contains data on 33 cars recorded in 1997. Selecting the File Data Source and clicking on New opens the menu shown in Figure 8.2. This menu lists all the ODBC drivers currently available on the PC. We are connecting to an Access database file, so we select the Microsoft Access Driver (*.mdb) from the list and click Next to proceed. This opens the menu shown in Figure 8.3, where a descriptive name for the DSN can be supplied. We enter `Car Data` in the space provided and click Next, which opens the menu in Figure 8.4. This menu gives you a summary of the choices you made; if you want to change any details you can click on Back. Clicking on Finish creates the DSN with the choices shown in the menu.

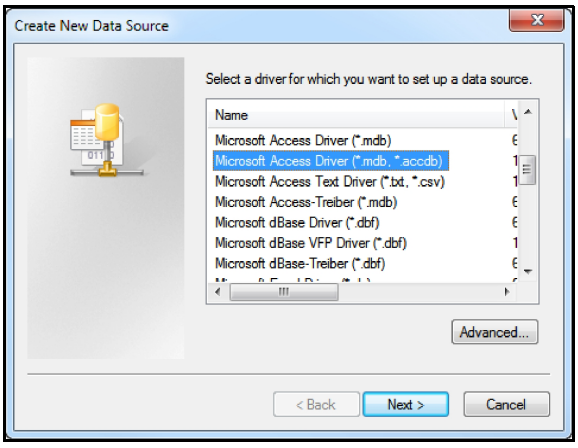


Figure 8.2

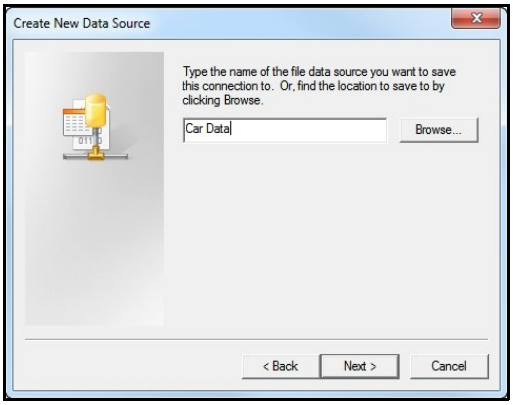


Figure 8.3

After clicking Finish you will be prompted with some additional menus depending on which ODBC driver you are connecting to. These menus are specific to the ODBC driver and are used for specifying information for the driver connection to the data source plus any other driver specific options. The Access Setup menu is shown in Figure 8.5. Here we need to specify the file name that we want to connect to (`cardata.mdb`).

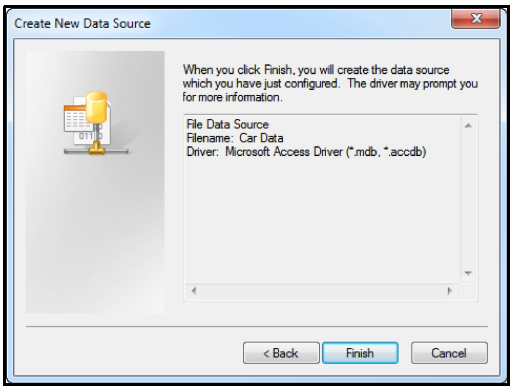


Figure 8.4

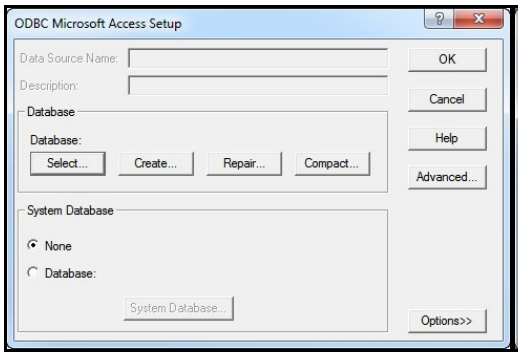


Figure 8.5

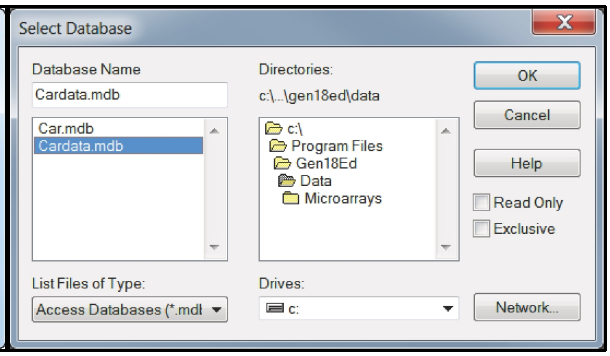


Figure 8.6

Clicking on Select opens a browse menu (see Figure 8.6) where we have selected the file `cardata.mdb`. Clicking on OK selects the file and displays the name in the Database options (see Figure 8.5). If the database is password protected, you can click on the Advanced button to specify a username and password associated with the database. If the database is password protected and you do not provide one using the Advance button, then you will be prompted for a password each time you try to connect. Clicking OK on the Access Setup menu completes the DSN and enters it into the list of File Data Sources on the Select Data Source menu.

To initiate the ODBC connection to Genstat, we double-click on the Car Data DSN from the list of File Data Sources. Figure 8.7 shows the resulting menu where you are provided with a list of all the tables and views within the DSN. Selecting a table or view from the Table list displays all the columns within that table or query within the Available Columns list. Merged data across tables are not supported with the interactive interface, but can be handled by editing the resulting SQL statement generated by Genstat. An alternative way to access merged data across tables is to create a view within the database itself as these are displayed in the Table list. You can then select the columns as you would with a single table. In Figure 8.7 we have selected the table `CarData`, and from the Available Columns, we have made a multiple selection `Car`, `Price` and `Max_MPH`. We have then clicked on the `->` button to copy the selected columns across to the Selected Columns list.

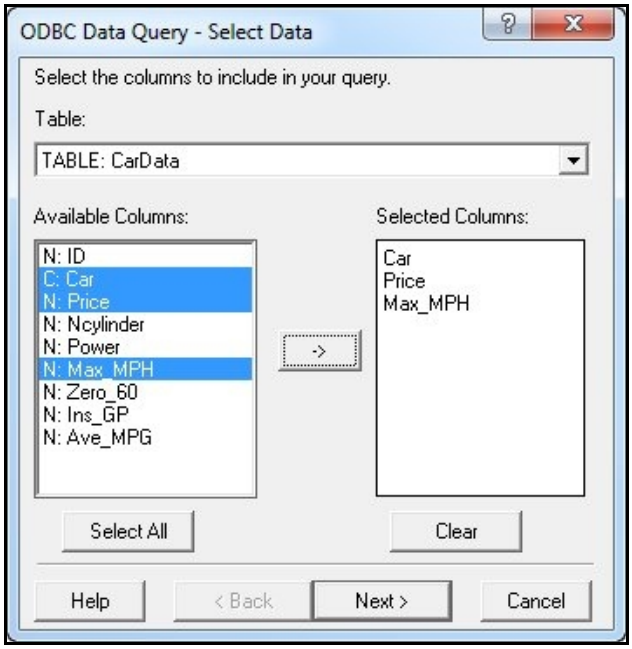


Figure 8.7

Clicking on Next opens a Filter menu as shown in Figure 8.8. Within this menu you can choose a subset of rows from the database based on a logical condition. The condition is entered into the space provided and you can use the lists of available columns, functions and operators to help build the expression. For example, we want to filter to those rows where the price for the cars is less than £10,000. Double-clicking on the name `Price` from the Available Column list puts the name in the edit field for the expression. Similarly double-clicking on 'less than' in the Operators list puts a '<' symbol into the expression. Finally we type 10000 and click Next to continue.

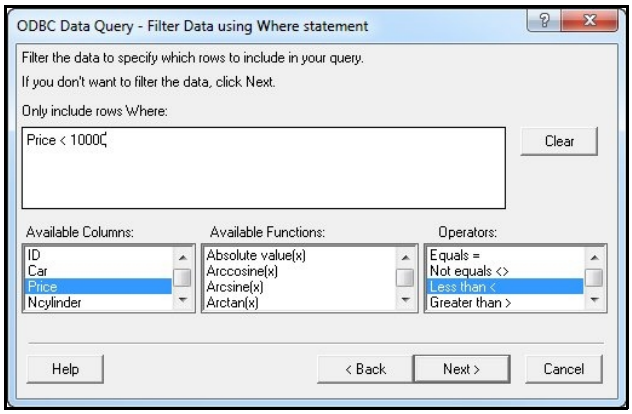


Figure 8.8

Figure 8.9 shows the final menu in the process; this specifies how you want to run the query. You can simply run the query by selecting the Run the SQL Query option. Alternatively, you can view the generated SQL statement, by selecting the View or Edit the SQL Query option as shown in Figure 8.9. Selecting the option to Save the Query makes the box below available for us to enter the file name

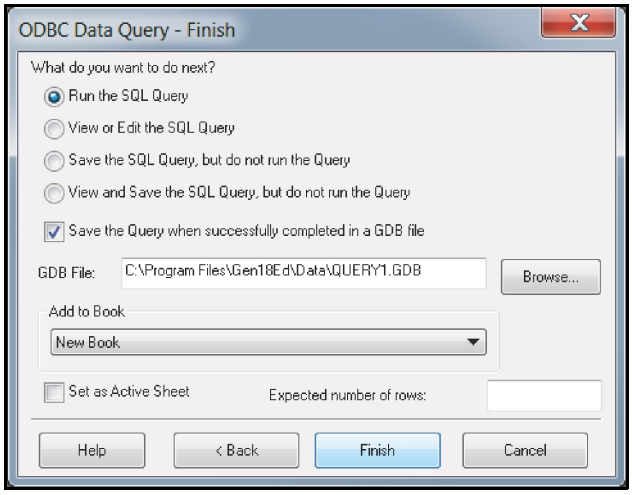


Figure 8.9

```
C:\Program Files (x86)\Gen18Ed\Data\QUERY1.GDB.
```

This will save the whole ODBC query process within a file called a Genstat GDB file. A GDB file can be opened using the Open option on the File menu, and will automatically run the query on the ODBC Server specified within the file.

Clicking Finish opens the SQL View menu shown in Figure 8.10 that displays the SQL statement generated by the query. You are able to edit the SQL within this window if you wish, before actually running the query. Clicking on the Tables or Columns buttons will open lists that can be used to construct the query. When editing the SQL, any column names containing non alphanumeric characters must be surrounded by quotes. For example, the column name `Max_MPH` contains an underscore (_), so the name has been surrounded by quotes in the generated SQL statement.

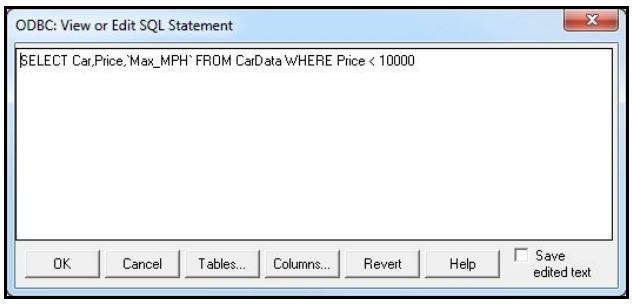
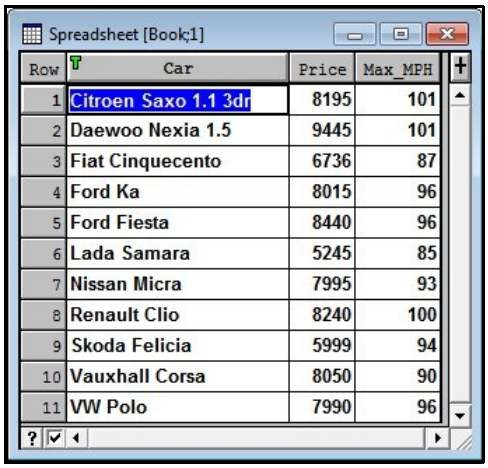


Figure 8.10

Using the generated SQL statement and clicking OK produces the spreadsheet shown in Figure 8.11.

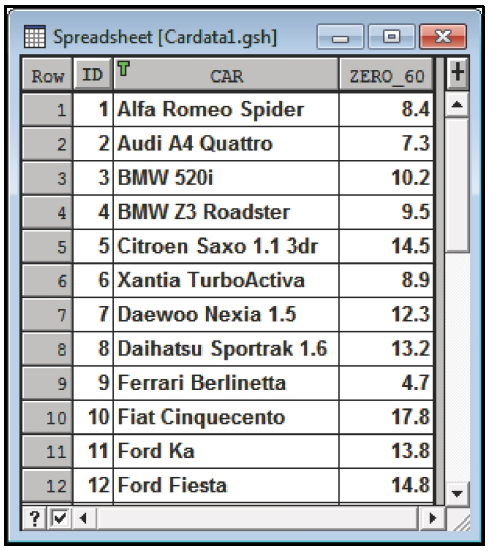
A Genstat spreadsheet can be written to a database using ODBC, provided you have the correct access rights to do this. There are three ways in which you can write to a database: create a new table, add new rows into a table, and update existing rows within a table.



Row	Car	Price	Max_MPH
1	Citroen Saxo 1.1 3dr	8195	101
2	Daewoo Nexia 1.5	9445	101
3	Fiat Cinquecento	6736	87
4	Ford Ka	8015	96
5	Ford Fiesta	8440	96
6	Lada Samara	5245	85
7	Nissan Micra	7995	93
8	Renault Clio	8240	100
9	Skoda Felicia	5999	94
10	Vauxhall Corsa	8050	90
11	VW Polo	7990	96

Figure 8.11

To illustrate these methods we will write data to the Car Data database in MS Access. The file `Cardata1.gsh`, shown in Figure 8.12 contains data on the time required to accelerate from 0 to 60 mph. With this file open in Genstat select Create Database Table from Export on the Spread menu. As with reading data using ODBC, you are required to specify a DSN for the database to which you want to connect.



Row	ID	CAR	ZERO_60
1	1	Alfa Romeo Spider	8.4
2	2	Audi A4 Quattro	7.3
3	3	BMW 520i	10.2
4	4	BMW Z3 Roadster	9.5
5	5	Citroen Saxo 1.1 3dr	14.5
6	6	Xantia TurboActiva	8.9
7	7	Daewoo Nexia 1.5	12.3
8	8	Daihatsu Sportrak 1.6	13.2
9	9	Ferrari Berlinetta	4.7
10	10	Fiat Cinquecento	17.8
11	11	Ford Ka	13.8
12	12	Ford Fiesta	14.8

Figure 8.12

Double-clicking on the Car Data DSN in the File Data Source list on the Data Source menu opens the menu in Figure 8.13. Here we have entered the name `NewCarData` for the table that is to be created in Access. By default all the columns are transferred into the table. However, you can select specific columns from the spreadsheet to be transferred by clicking on the Select Columns to be in Table button.

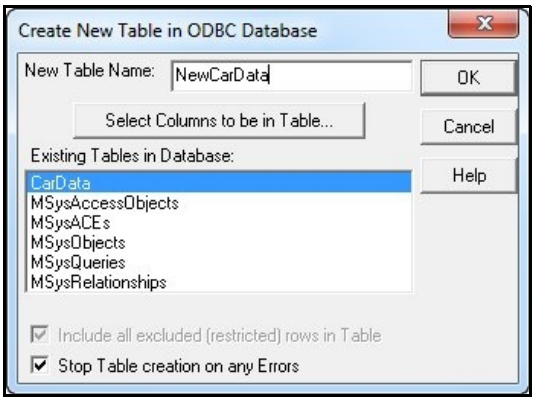


Figure 8.13

Clicking OK adds the new table to the database, and on successful completion of the process a prompt appears as shown in Figure 8.14. When columns are transferred to the new table in the database, the same column names are used as in the spreadsheet.

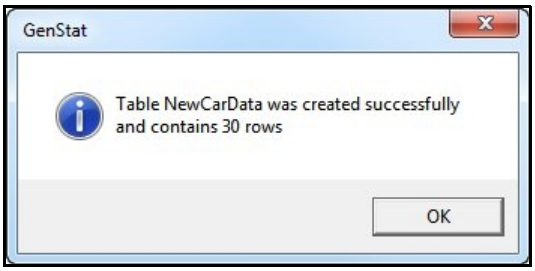


Figure 8.14

Figure 8.15 shows the new table within Access with the same column names as the Genstat spreadsheet.

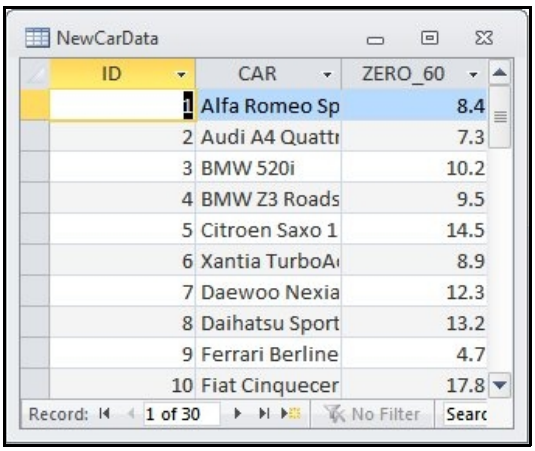


Figure 8.15

The file `Peugeot.gsh` contains additional data for some Peugeot cars, which needs to be added to the database. To add these rows to the new table, we first open the spreadsheet in Genstat (the spreadsheet is shown in Figure 8.16), and then select Insert into Database Table from Export on the Spread menu.

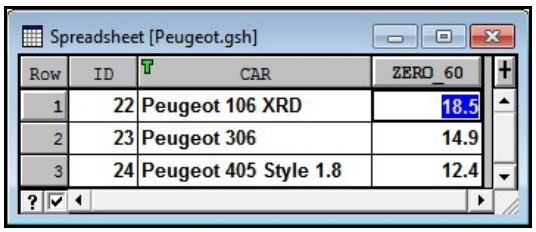


Figure 8.16

Double-clicking the Car Data DSN from the Select Data Source menu opens the menu shown in Figure 8.17. We select the table NewCarData from the Insert into Table list and select the Using Names in Sheet option. You can save the export link into a Genstat ODBC Link file (.GLK), so that you can automatically rerun the insert operation on subsequent spreadsheets without having to go through the menu steps again. We have selected the Save Export Link in GLK file option, and have entered the file name

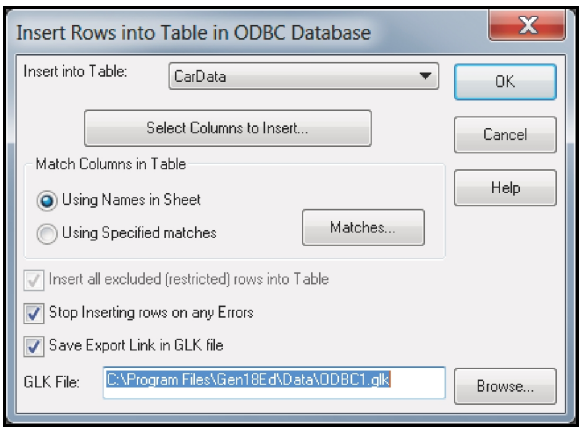


Figure 8.17

```
C:\Program Files (x86)\Gen18Ed\Data\ODBC1.glk
```

to save the export link information.

Figure 8.18

Clicking OK prompts you with confirmation dialog shown in Figure 8.18, and inserts the rows from the spreadsheet into the database as shown in Figure 8.19.

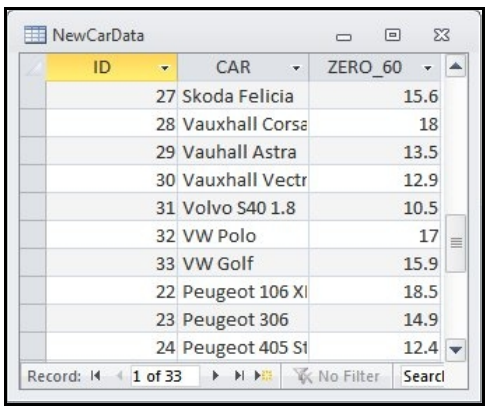


Figure 8.19

The final method of writing to a database is to update existing rows within the database. The file `Ford.gsh` contains data from further testing on Ford cars where the time taken to reach 60 mph has been improved on all models. Opening the file `Ford.gsh` into Genstat, gives the spreadsheet shown in Figure 8.20.

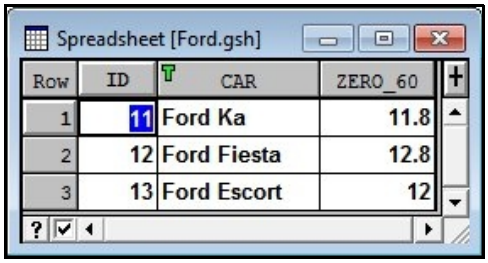


Figure 8.20

We select Merge with existing Database Table from Export on the Spread menu and double-click the Car Data DSN on the Select Data Source menu. This opens the menu shown in Figure 8.21. Select NewCarData from the Merge data into Table list. Each car has an `ID` number that can be used to match them between the spreadsheet and database. So, select the column `ID` from the Matching Sheet Column list and select `ID` from the With Table Column list. This will match the data from the spreadsheet with the database using the column `ID` and replaces the values for the other columns. As with the menu for inserting rows into a database, you can save the export link information in a Genstat ODBC Link file (.GLK) to automatically run the process another time. We have specified this by selecting the Save Export Link in GLK file option, and entered the filename

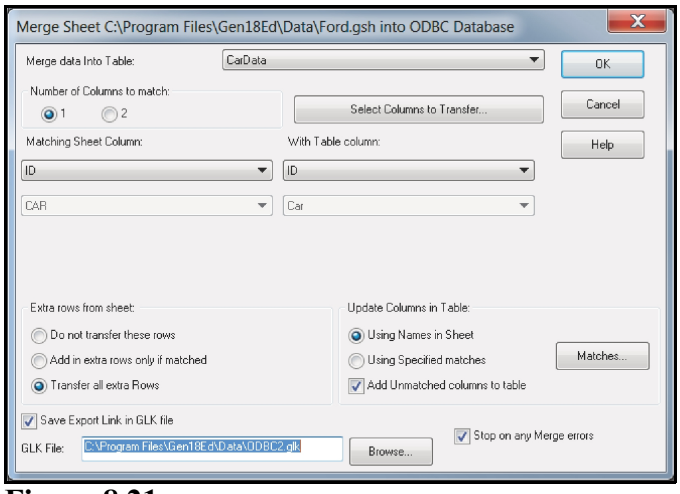


Figure 8.21

```
C:\Program Files (x86)\Gen18Ed\Data\ODBC2.glk
```

in the space provided. A description of the other options on this menu can be found by clicking on the Help button.

Clicking OK prompts you with confirmation dialog shown in Figure 8.22 and replaces the rows in the database using those the Genstat spreadsheet, as shown in Figure 8.23.

Figure 8.22

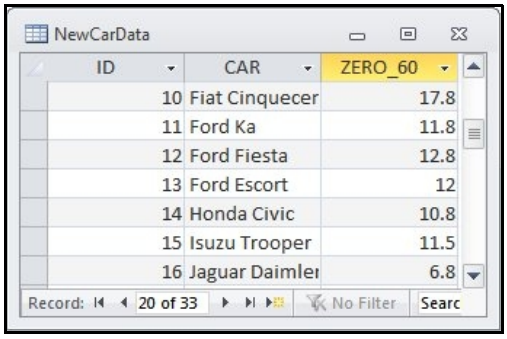


Figure 8.23

To run a Genstat ODBC Link file to automatically insert rows or merge data into a database, select Run OBDC export link from Export on the Spread menu.

This opens the menu shown in Figure 8.24, where you can either run the link using the current spreadsheet, or run the link from a given Genstat spreadsheet file (you will need to specify the location of the file).

In Figure 8.24 we have used the Browse button to select the file

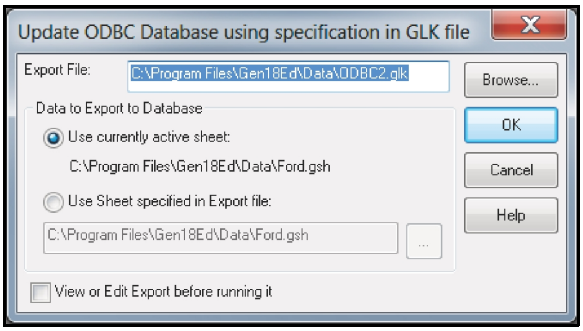


Figure 8.24

```
C:\Program Files (x86)\Gen18Ed\Data\ODBC2.glk
```

and have selected the option to run using the currently active sheet. Clicking OK will rerun the export link for the replacing of rows outlined above and will produce the confirmation dialog and access table as shown in Figures 8.22 and 8.23.

8.1 Exercise

Clear all the data from the Genstat data pool. Using the ODBC Data Query menu from the Spread menu, connect to the car data (`Car.mdb`) and bring in all the columns of data, but only for cars costing £10000 or more. Save the query in a Genstat GDB file.

Sort (and display) the data in ascending order according to their horsepower.

Clear the data from the Genstat data pool. This time using the Open in the File menu, open the GDB file.

9 Other facilities

There are many other facilities for data manipulation using the spreadsheet facilities within Genstat. One useful feature is the ability to set a spreadsheet as an active spreadsheet. If a spreadsheet is set as an active spreadsheet, then only changes made in this spreadsheet will be updated in Genstat; all other spreadsheets will be prevented from updating Genstat, until you remove this setting. Another advantage of specifying an active spreadsheet is that the Spread menu will always be available whether you are in the spreadsheet or within a text window. You can set a spreadsheet as an active spreadsheet by selecting Set as active sheet on the Spread menu. More details on active spreadsheets can be found in the on-line help.

Another method for rearranging data in Genstat is through the Paste Special menu. This is accessed using the Paste Special option on the Edit menu. With this feature you can copy data onto the clipboard from another data source, and then control how the data are to be pasted within the Genstat spreadsheet cells. For example, you can use this to paste a rectangular block of data into a single column, or to paste grouped blocks of data into multiple rows.

You can calculate summary statistics based on just the data within the current spreadsheet. For example, you may want to aggregate data to provide summaries, or perhaps expand a set of factor results to give a row for every factorial combination. This menu is accessed by selecting Summary Stats from Calculate on the Spread menu.

A spreadsheet can have a set of Genstat commands embedded within it. This allows you to provide a statistical analysis along with the spreadsheet. This is explained, with an example, in Section 8.3 of the *Introduction to the Genstat Command Language* (which can be opened by clicking on the Introduction to the Genstat Command Language sub-option of the Genstat Guides option of the Help menu in Genstat's menu bar.

Genstat includes facilities for importing and updating large spreadsheets. Opening large spreadsheet/book files into Genstat can sometimes be slow as millions of data are being loaded into memory. A spreadsheet or book can be viewed within Genstat without loading all the data by opening the file as read only. When a spreadsheet or book file is opened using this mode you can scroll and move around the spreadsheet as normal, but you cannot edit the data on the sheet. To edit the data on the spreadsheet you must remove the read-only status. The `General` tab on the `Spreadsheet Options` menu includes options to allow a fast load of large spreadsheets imported using the Data menu and for updating data from the spreadsheet to the Genstat data core.

10 Commands

Many of the menu options illustrated in this chapter can be equivalently carried out using the command language. However, some of the features such as data verification, copying from the clipboard and bookmarks however, can only be performed in Windows™. Spreadsheets can be loaded and saved from the server. The SPLOAD and IMPORT directives read in data from files, and the EXPORT directive saves data to a file. SPLOAD reads in GSH and GWB files only, but IMPORT reads in these plus files of all formats supported by the spreadsheet (Excel, R etc.). SPLOAD can read in all pages of a GWB, although the SHEET parameter can be used to specify just particular sheets to be read in. The following command reads in just two of the sheets in file Portmatrices.gwb:

```
SPLOAD '%GENDIR%/Data/Portmatrices.gwb'; SHEETNAME=!T('B','C')
```

(note the %GENDIR% stands for the directory where Genstat has been installed). When using IMPORT with spreadsheet files, the SHEETNAME and CELLRANGE parameters can be set to read in just a specified block of data, as you did in Section 1. To read in the data in Exercise 1.1 you would use the command:

```
IMPORT 'Traffic.xls'; SHEETNAME='counts'; CELLRANGE='B3:D43'
```

The EXPORT command will write a single page of data out to a new file, or can be used to add data to an existing file. If the file is an Excel or GWB file, then multiple pages can be added to the file, and if a GSH file then data can be concatenated (columns added on the right of the sheet), appended (rows added to the end of the sheet), or merged (new rows added where the new ids are unique, otherwise data added to existing rows). The following uses EXPORT to create and add data to a GWB file:

```
CALC X1,Y1,Z1,X2,Y2,Z2 = GRNORMAL(3(80,20);3(1,2);3(1,4))
EXPORT [OUTFILE='Test.gwb';METHOD=overwrite] X1,Y1
EXPORT [OUTFILE='Test.gwb';METHOD=concatenate] Z1
EXPORT [OUTFILE='Test.gwb';METHOD=append;GROUPS='Group'] \
  X2,Y2,Z2
EXPORT [OUTFILE='Test.gwb';METHOD=add;SHEETNAME='Group2'] \
 X2,Y2,Z2
```

The first line calculates some random data, and the second stores two variates in a file. The third line adds a new column to the sheet in the file, and the fourth adds new rows to the end of the sheet, as well as a new factor column Group which will have level 1 or the existing rows, and 2 for the added rows. The last line adds a new sheet Group2 to the file.

To filter or restrict data as outlined in Section 3.1 you can use the RESTRICT directive. The VECTOR parameter specifies the data columns that are to be restricted and the CONDITION parameter allows you to set the condition to restrict the data values by. For example, the following shows how to perform the restriction in Figure 3.3:

```
RESTRICT Drench,Lwt2; Condition = ((Lwt2 < 51))
```

to remove a restriction you need to use RESTRICT again but omitting the condition

```
RESTRICT Drench,Lwt2
```

To form a subset of data you can use the SUBSET procedure. In this procedure the condition is supplied using the CONDITION option. The NEWVECTOR parameter allows you to specify a new vector to save the subset (otherwise it overwrites the contents of the original vector). The following example shows how to create a new subset called subLwt2 from the values of Lwt2 using the condition that all the weights are less than 51.

```
SUBSET [CONDITION=(Lwt2 < 51)] Lwt2; NEWVECTOR=subLwt2
```

The sorting of data in Section 4.4 can be performed using the SORT directive. The

index column that defines the sorted order is specified by the INDEX option, and the direction of sorting is specified using the DIRECTION option. The sorting performed in Figure 3.11 can be reproduced by:

```
SORT [INDEX=Lwt2; DIRECTION=ascend] Drench,Rep,Lwt1,Lwt2
```

For a multi-key sort you can supply a list of identifiers for the INDEX option and the data will then be sorted by list order. The data will be sorted by the first item in the index list and then by the second item in the index list, and so on. To reproduce the sort in Figure 3.13 you could use the following.

```
SORT [INDEX=Drench,Lwt1; DIRECTION=ascend] \
  Drench,Rep,Lwt1,Lwt2
```

The stacking and unstacking of data can be reproduced using the STACK and UNSTACK procedures.

For stacking columns together, the source factor is saved using the DATASET option and the columns that are to be stacked are supplied by the parameters V1-V100. The following commands demonstrate how to reproduce the stacked data set in Figure 3.23.

```
STACK [DATASET=Toy] Year_1,CostDog_1,SoldDog_1;\
  V1=Year,CostDog,CostKitten;\ V2=Year,SoldDog,SoldKitten
```

The unstacking of the data in Figure 3.25 can be produced using the command shown below:

```
UNSTACK [DATASET=Year_1] 3(CostDog_1,SoldDog_1);\
  DATASETINDEX = 1998,1999,2000;  \
  UNSTACKEDVECTOR =CostDog_101,CostDog_102,\
  CostDog_103, SoldDog_101,SoldDog_102,SoldDog_103
```

The source factor, Year_1, is supplied using the DATASET option. The DATASETINDEX parameter specifies the levels or labels of the DATASET factor indicating the group whose units are to be stored in the UNSTACKEDVECTOR. In this example we have used the levels for the Year_1: 1998, 1999 and 2000. The data to be stacked is supplied as a list using the UNSTACKEDVECTOR parameter.

For Dynamic Data Exchange the DDEEXPORT procedure can be used for writing data to a DDE server. This can allow you to build up worksheets of results in spreadsheets, such as Excel. Within Excel you can write data to the worksheet cell by cell, or alternatively you can add formula to cells. You can also send macro commands to Excel to open files, add new worksheets, save or close files.

The location within the DDE server is specified using the options SERVER, TOPIC and ITEM. However, for the two common spreadsheets Excel and Quattro Pro for Windows™, these have been broken down into more convenient options called OUTFILE, SHEETNAME, COLUMN and ROW. For Excel and Quattro Pro only the first cell needs to be provided, as Genstat can automatically work out the range given the size of the data. If you want to send commands you can supply these by setting METHOD=command. The following example will open Excel, create a new worksheet and copy the data to the worksheet. Open the columns crop and counts from the file Bacteria.xls, and then run the following program to copy them back into the file on a new sheet.

```
DDEXPORT [METHOD=command]\
  '[OPEN(('C:\\Program Files\\Gen18Ed\\Data\\Bacteria.xls'')]'
DDEXPORT [METHOD=command] '[WORKBOOK.INSERT(1)]'
DDEXPORT [OUTFILE='Bacteria.xls';SHEET='Sheet1';\
  ROW=1; COL=1] crop,counts
DDEXPORT [METHOD=command] '[SAVE()]'
```

The DDE commands used in the example above are a subset of the Excel 4 macro language. The format of the commands is [Function(arg1,arg2,...)]. If there are text strings in the arguments then these must be supplied in double quotes (for example,

"Arg1"). The following list specifies some of the most useful Excel commands that can be used with the DDEEXPORT procedure.

[APP.RESTORE()] Restore the Excel window
[APP.MINIMIZE()] Minimize the Excel window
[APP.ACTIVATE()] Make Excel the application with the focus
[OPEN("filename")] Open a workbook in Excel
[WORKBOOK.INSERT(1)] Insert a new workbook
[WORKBOOK.SELECT("sheetname")] Make the named sheet the current sheet
[WORKBOOK.DELETE()] Delete the current sheet
[SELECT("object")] Select the cells/column/rows specified in object
[SORT(1,"R1C1",1)] Sort the selected cells using key in specified cell
[SAVE()] Save the current workbook
[SAVE.AS("filename",1)] Save the current workbook as a new file
[CLOSE(1)] Close and save the current workbook (0 = close but do not save)

To read data from a database you can use the DBIMPORT procedure. You can supply the name of an existing GDB file containing information on the data to load using the GDBFILE parameter. Alternatively, you can supply a database connection string using the DB parameter with an SQL statement using the SQL parameter. To run the example in Section 4.9 you could supply the saved GDB file as follows:

```
DBIMPORT GDBFILE='C:\\Program Files\\Gen18Ed\\Data\\Query1.gdb'
```

To write tables or data to a database you can use the DBEXPORT procedure. The METHOD option specifies how the data are to be written in the ODBC data source: to create a table use METHOD=create, to add rows to an existing table use METHOD=insert, and to update rows in an existing table use METHOD=merge. In its simplest form, you can just provide a previously saved Genstat ODBC Link file (GLK). The data to be sent can either be specified as a pointer to a set of structures in Genstat or a text giving a Genstat spreadsheet (GSH) file. If you are using an ODBC Link file and this does not specify a Genstat spreadsheet as the data to transfer you will need to specify the data using the DATA parameter. Column names within the ODBC table are assumed to be the same as the Genstat identifiers. If you want to use different names then you can specify COLUMNNAMES and WITH (for matching with MATCH). The COLMERGEMETHOD option controls whether columns from the data not found in the data database table are to be added to the database table. Subsets of columns can be specified using the SUBSET parameter.

If METHOD=merge, the MATCH parameter must be set and five columns at most can be matched. The WITH parameter may be set if the columns in the table do not have the same names as the structures specified by the DATA parameter. The ROWMERGEMETHOD option controls how unmatched rows are handled in a merge: the setting none does not add unmatched rows, the setting matched only adds a row if another with the same matching criteria already exists in the table, and all adds in all unmatched rows into the table. The WARNINGDIALOGS option can be used to control whether warning message boxes are displayed on the Windows™ desktop when errors occur. The option ERRORACTION controls what to do when non-fatal errors occur; you can halt the process or continue. The following example shows how you can run a Genstat ODBC Link file:

```
DBEXPORT [GLKFILE='ODBC1.GLK']
```

The second example will run a Genstat ODBC Link file, but this time data currently stored within Genstat will be used for the merging.

```
DBEXPORT [GLKFILE='ODBC2.GLK'] ID,CAR,ZERO_60
```

The last example demonstrates how you can extract the connection string from a Genstat ODBC Link file, and create a new table in the database using data currently within Genstat.

```
"Read the database connection string from GLK file"
OPEN 'ODBC1.GLK'; CHAN=2; INPUT; WIDTH=600
SKIP [CHAN=2] 1; TEXT [1] DB "Skip ODBC Link ID"
READ [CHAN=2;PRINT=*;LAYOUT=FIXED;FORMAT=!(600);END=*] DB
CLOSE 2; INPUT

"Create the new table in the database"
DBEXPORT [METHOD=create] ID,CAR,ZER0_60; DB=DB;\
   TABLE='NewTable'
```

If you have data stored in Genstat spreadsheet (GSH or GWB) files or foreign format files (e.g. Excel, SAS, R etc.), then you can use the SPCOMBINE procedure to amalgamate the data into a single file. The METHOD option has settings to add, append, concatenate or merge the data into the output file given by the OUTFILE option. If the output file does not exist, the first file will be used as the base for amalgamating the other files into. The file types need not all be the same and the output file can be a range of output formats, as supported by EXPORT. The following code uses this to add the 5 Excel files Grazing 1.xls - Grazing 1.xls into two combined files, either adding the extra data as sheets or rows.

```
"Create a text containing the 5 file names."
TXCONSTRUCT [TEXT=Files] \
          !t(5('%GENDIR%/Data/Grazing ')),!(1...5),\
          !t(5('.xls')); DECIMALS=0

"Add 5 Excel files into a single Excel file with 5 sheets."
SPCOMBINE [OUTFILE='Grazings.xls'; METHOD=add] FILE=#Files; \
          PAGENAME='P1','P2','P3','P4','P5'

"Append 5 Excel files into one Excel file with a factor
 indicating source."
SPCOMBINE [OUTFILE='Grazing.xls'; METHOD=append;\
          GROUPS='Period'] FILE=#Files;\
          GLABEL='1','2','3','4','5'
```

Index