Genstat®

Spreadsheet

www.vsni.co.uk

VSNi

A Guide to the Genstat® Spreadsheet
(22nd Edition)

by David Baird, Darren Murray and Jane Cohen

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 2022, for Genstat *for Windows* 22nd 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.

Genstat Example Data Sets

If installed Genstat into the default folder, the data files used in this Guide will be found in `C:\Program Files\Gen22Ed\Data`. You can import these into Genstat by selecting File | Open then navigating to the file location. Alternatively select File | Open Example Data Sets, then type the name of your required file and click Open data.

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, 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	

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 crops. 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 found in

`C:\Program Files\Gen22Ed\Data`.

In this example we want to open the Excel file containing the data shown in Figure 1.1. To open the file, from the menu bar select File | Open. 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 dropdown 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 dropdown list in the bottom-right corner so that we can see what Excel files are available.

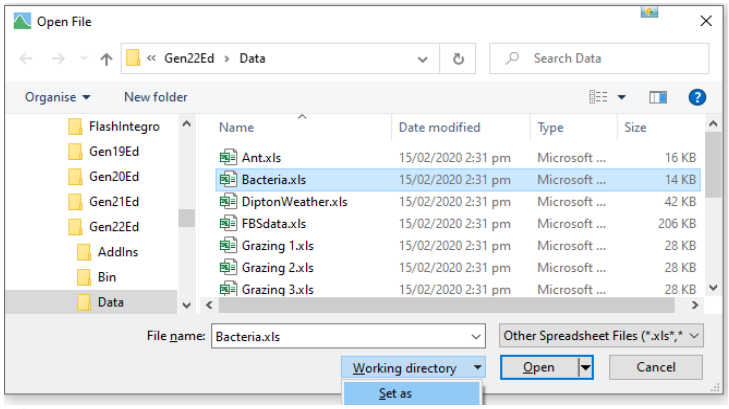


Figure 1.2

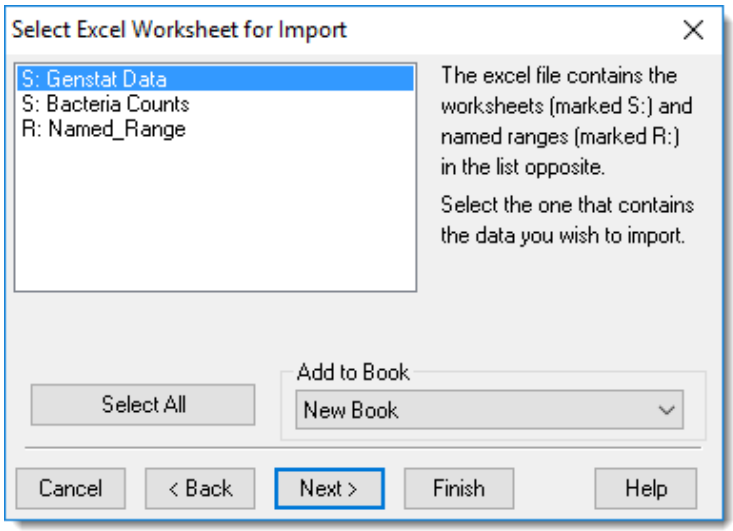


Figure 1.3

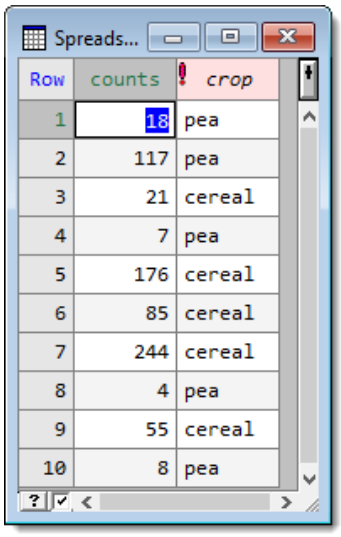


Figure 1.4

Selecting the file `Bacteria.xls` and clicking Open, or double-clicking the filename, displays the menu shown in Figure 1.3. This is the initial menu of the Excel import wizard. 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 dropdown 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.

When you click on the Output window, the data in the spreadsheet are automatically transferred to Genstat's central data pool.
(If the Output window is not displayed click the Window tab on the bottom left of the screen, then double-click Output at the top of the Window View pane.) The Output window displays a brief summary of the data that have been transferred, as shown in Figure 1.5.

```
Data imported from Excel file: C:\Program Files\Gen22Ed\Data\Bacteria.xls
 on: 25-Nov-2021 10:19:46
 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
```

Figure 1.5

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 are in Genstat, by looking in the Data View pane (Figure 1.6).

To see the data, click the Data tab at the bottom left of the screen then move up and open the Vectors | All Vectors folder.

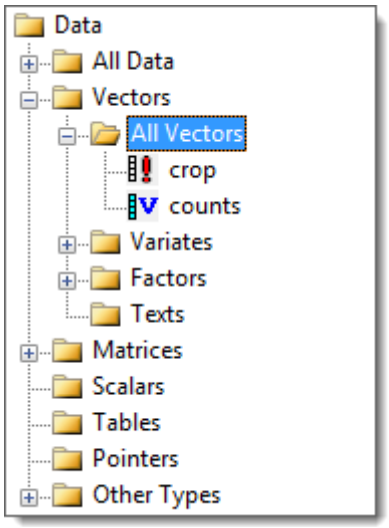


Figure 1.6

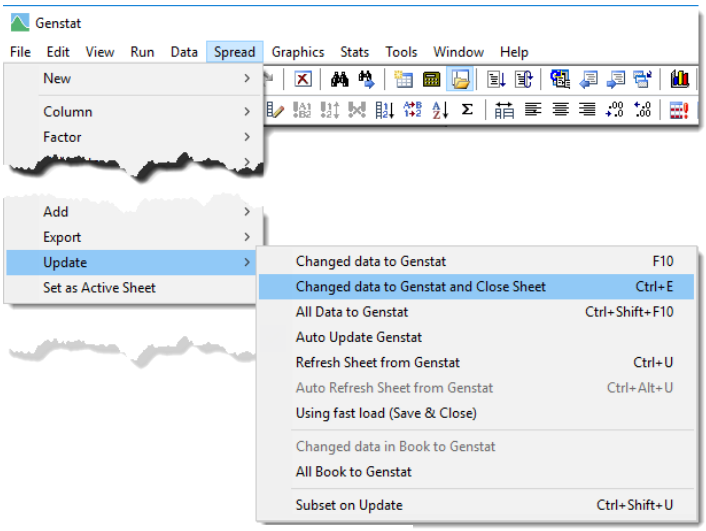


Figure 1.7

Now make the spreadsheet active again: click the Window tab at the bottom left of the screen then double-click the `Bacteria.xls` spreadsheet. This will let you access the options on the Spread menu. You can transfer the data in the Data View explicitly to the data pool by selecting Spread | Update | Changed data to Genstat and Close Sheet (Figure 1.7). Selecting this item updates Genstat then closes the spreadsheet. The standard method of updating the pool uses the Genstat `READ` command.

The item Using fast load (Save & Close) provides a more efficient alternative, using the SPLOAD command, for large spreadsheets in Genstat's native gsh format.

We shall now import some data from another Excel worksheet.

Data are not always stored in a singular rectangular format within a spreadsheet but may have multiple blocks of data entered on a single worksheet. Figure 1.8 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.

From the menu select File | Open and select the file Bacteria.xls. This opens the initial menu of the Excel wizard, as shown earlier in Figure 1.3. A named range containing counts2 and crop2 already exists in the Excel file (R: Named_Range), but we'll ignore this and instead define the same cell range for import using the Excel wizard.

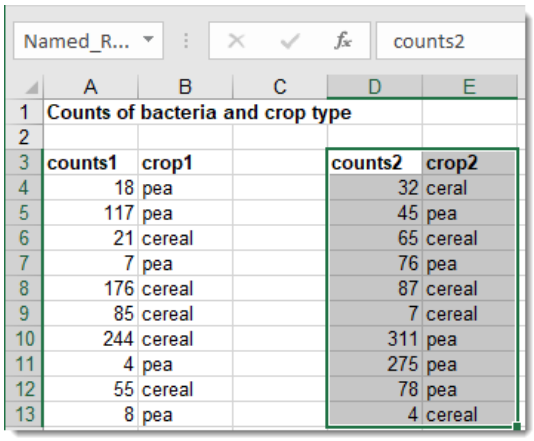


Figure 1.8

Select the worksheet Bacteria Counts and click Next. The second menu in the wizard lets you define the range you want to select. Click the Specified Range radio button and enter the range D3:E13 into the adjacent field as shown in Figure 1.9.

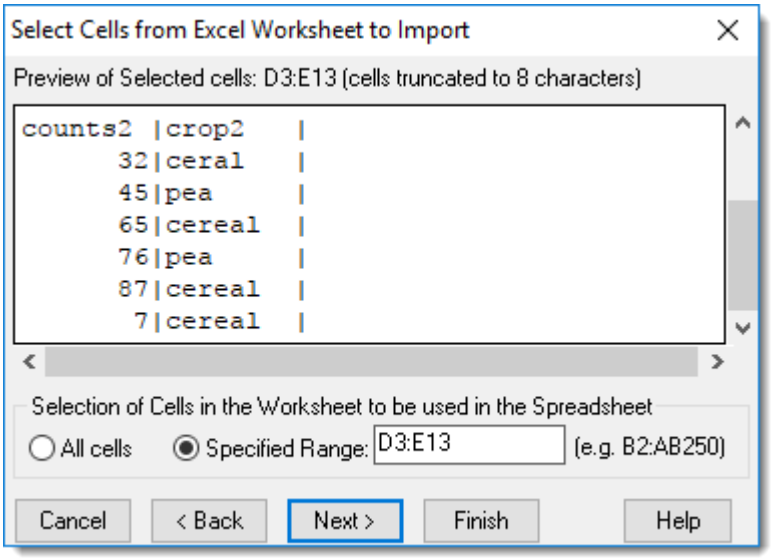


Figure 1.9

When you click Finish Genstat detects that the column `crop2` has repeated labels and displays the menu shown in Figure 1.10. This menu displays all the columns that have repeating values. 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, double-click the name `crop2` in the list (alternatively you can click the Factor button). This changes the prefix from T to F specifying the column will be a factor. Click OK to create a new Genstat spreadsheet.

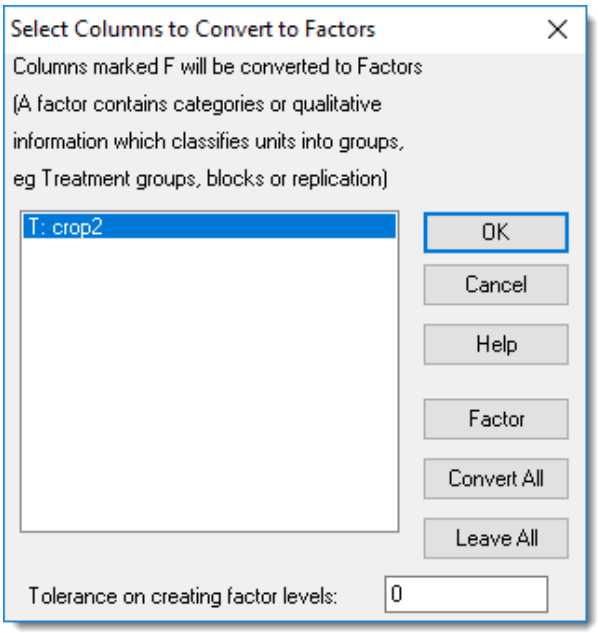


Figure 1.10

If we now click on the Output window, the data in the spreadsheet are transferred to Genstat's central data pool as shown over in Figure 1.11.

```
Data imported from Excel file: C:\Program Files\Gen22Ed\Data\Bacteria.xls
 on: 25-Nov-2021 10:25:35
 taken from sheet "Bacteria Counts", cells D4:E13

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

      Identifier   Values  Missing   Levels
           crop2       10        0        2
```

Figure 1.11

An alternative to defining a cell range in the Excel wizard, is to create a named range for the rectangular block of data directly within Excel then 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 (see previous page) 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 1.9.

The third menu in the wizard (Figure 1.12) lets you choose which of the columns in the worksheet or range to read. By default, they are all read.

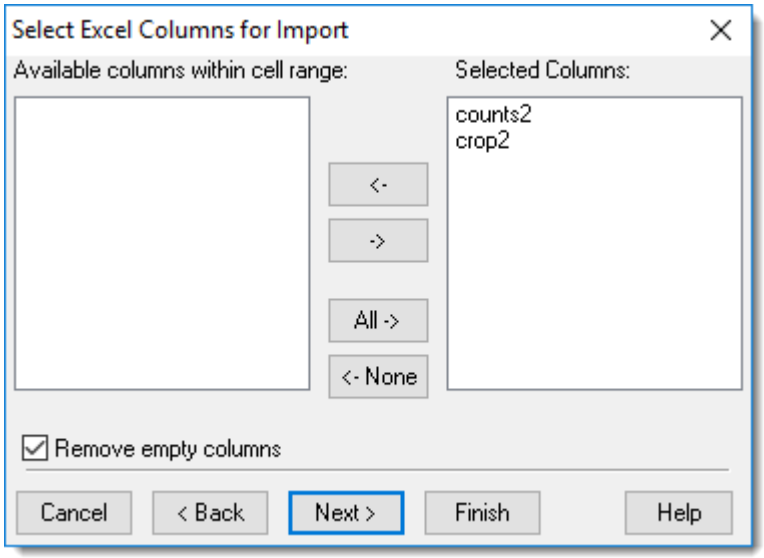


Figure 1.12

When you click Next the final wizard menu appears, shown in Figure 1.13. This menu contains 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.

Instead, we select the Factors tab, and select the checkbox Suggest columns with only a few unique values to be Factors. 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 the menu shown previously in Figure 1.9 offering you the choice to convert them. Clicking the Factor button will perform the conversion and clicking OK will open the spreadsheet.

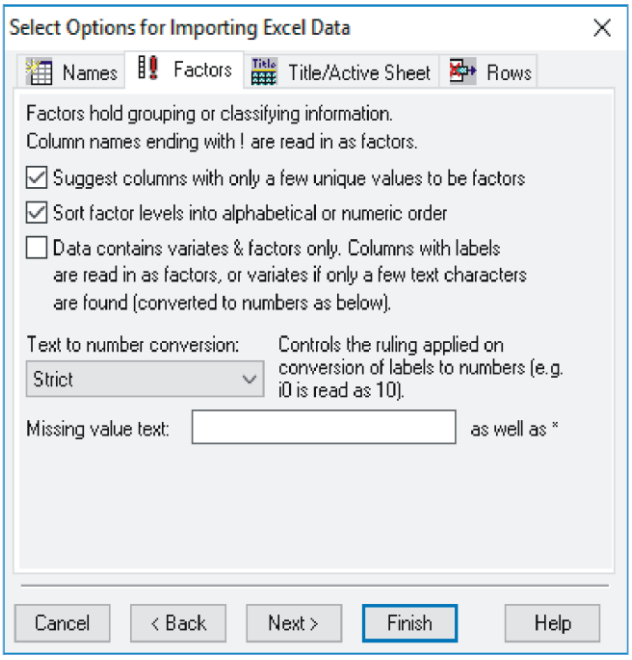


Figure 1.13

An alternative way to input data is to use the facilities within the Spread menu. In this example we will use the clipboard to copy the columns `count1` and `crop1` from the Excel file `Bacteria.xls` 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. Open Excel then open the file `Bacteria.xls`. Select the first rectangle of data including the column names (data range A3:B13).

Copy this selection to the clipboard using Ctrl+C or another method. 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 around the selected cells in Excel.

Now, in Genstat, we create a spreadsheet of the data, by selecting Spread | New | From Clipboard as shown in Figure 1.14.

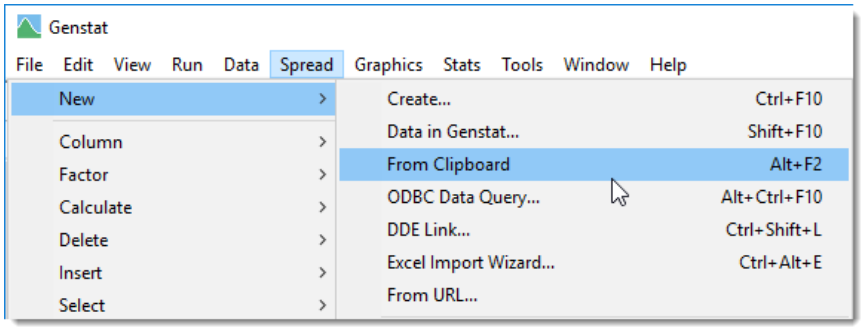


Figure 1.14

The New Spreadsheet from Clipboard menu (Figure 1.15) is displayed. We leave the Suggest columns to be factors box selected and leave Add to Book set to New Book. When you click OK 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 1.10.

Leaving `crop1` as a text and clicking OK produces the spreadsheet shown in Figure 1.16.

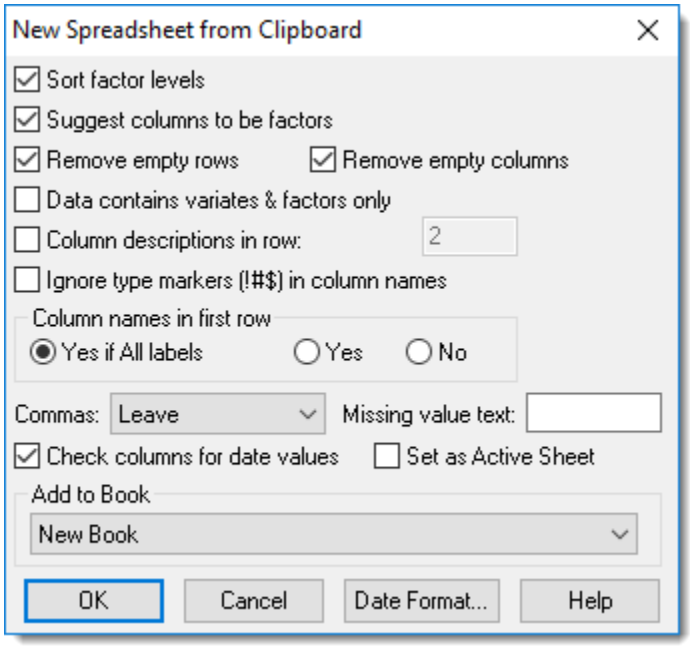


Figure 1.15

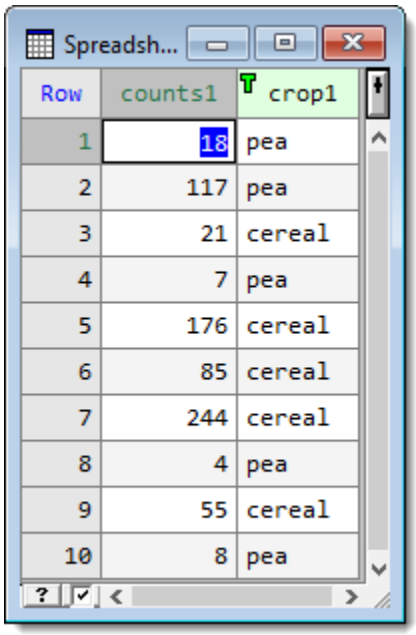


Figure 1.16

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 select Spread | Factor | Convert to, as shown in Figure 1.17.

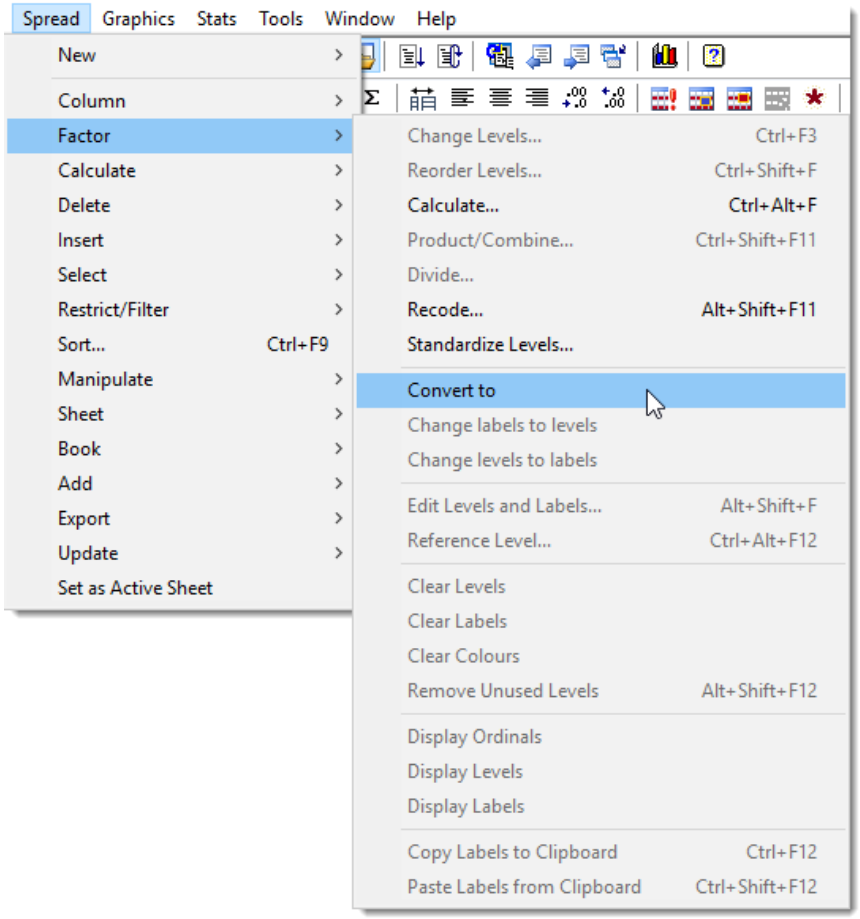


Figure 1.17

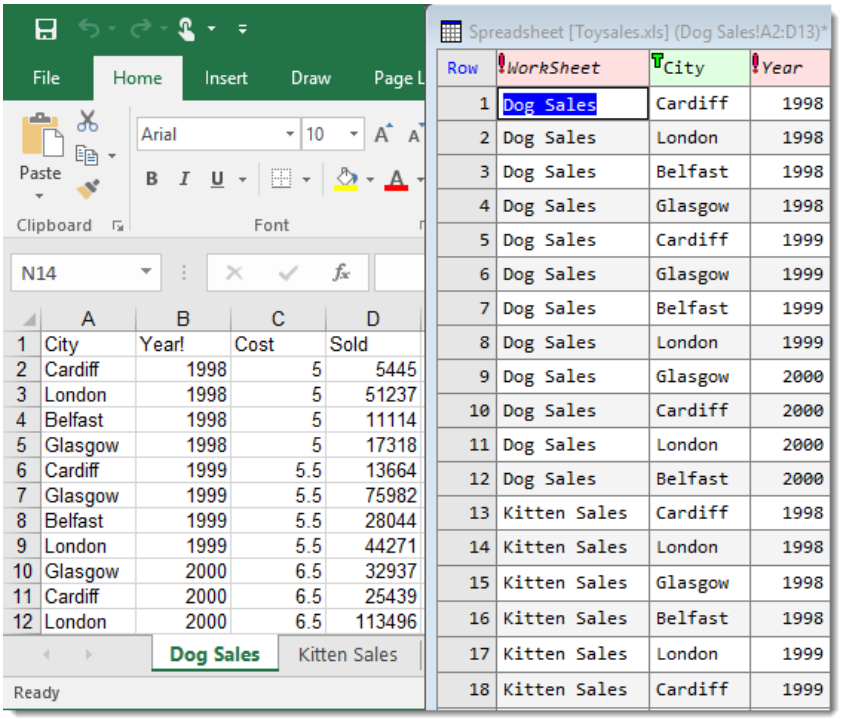
1.1 Exercise

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

1.2 Reading data from multiple sheets or files

Sometimes data from separate categories or trials will have been entered on separate pages within an Excel file or be held 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.

In our next example, all the data are on separate sheets in one Excel file, so we'll use Genstat's Append menu to combine them into one spreadsheet.



	A	B	C	D
1	City	Year!	Cost	Sold
2	Cardiff	1998	5	5445
3	London	1998	5	51237
4	Belfast	1998	5	11114
5	Glasgow	1998	5	17318
6	Cardiff	1999	5.5	13664
7	Glasgow	1999	5.5	75982
8	Belfast	1999	5.5	28044
9	London	1999	5.5	44271
10	Glasgow	2000	6.5	32937
11	Cardiff	2000	6.5	25439
12	London	2000	6.5	113496

Row	WorkSheet	City	Year
1	Dog Sales	Cardiff	1998
2	Dog Sales	London	1998
3	Dog Sales	Belfast	1998
4	Dog Sales	Glasgow	1998
5	Dog Sales	Cardiff	1999
6	Dog Sales	Glasgow	1999
7	Dog Sales	Belfast	1999
8	Dog Sales	London	1999
9	Dog Sales	Glasgow	2000
10	Dog Sales	Cardiff	2000
11	Dog Sales	London	2000
12	Dog Sales	Belfast	2000
13	Kitten Sales	Cardiff	1998
14	Kitten Sales	London	1998
15	Kitten Sales	Glasgow	1998
16	Kitten Sales	Belfast	1998
17	Kitten Sales	London	1999
18	Kitten Sales	Cardiff	1999

Figure 1.18

The left half of Figure 1.18 shows the file `Toysales.xls` in Excel. Two of the tabs contain data from sales of toy dogs (Dog Sales) and toy kittens (Kitten Sales). The right side of Figure 1.18 shows the Genstat spreadsheet we'll create by appending these two data sets from Excel. The data from Kitten Sales are placed directly under Dog Sales.

From the menu select Spread | New | Append Multiple Excel Spreadsheets. Locate the file `Toysales.xls` and double-click to open it.

After you have selected the file, you will get the dialog in Figure 1.19, 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 Ctrl or Shift key, or else by using the Select All button if all sheets are to be appended. 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 in this dialog is how the columns are to be matched between the two sheets. The Match Columns by section 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 sheet's 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.

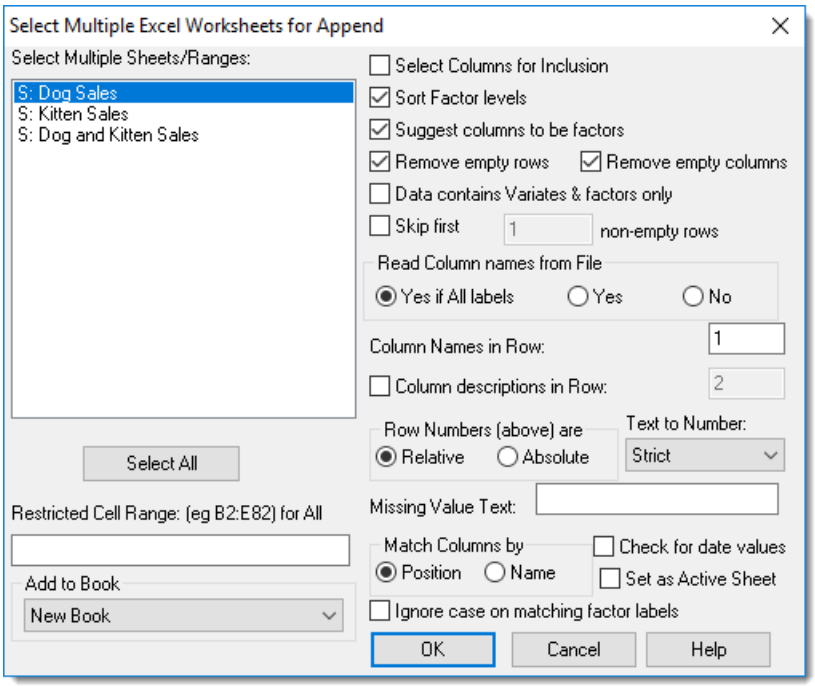


Figure 1.19

When you click OK the dialog in Figure 1.20 will display. This detects that some columns look like factors as they have a few unique values that repeat. Click OK to close the dialog without making any changes. (Alternatively, you could double-click an entry to make it a factor.)

The resulting spreadsheet is shown in Figure 1.21. This contains the new factor column Worksheet, which gives the name of the worksheet that each row came from.

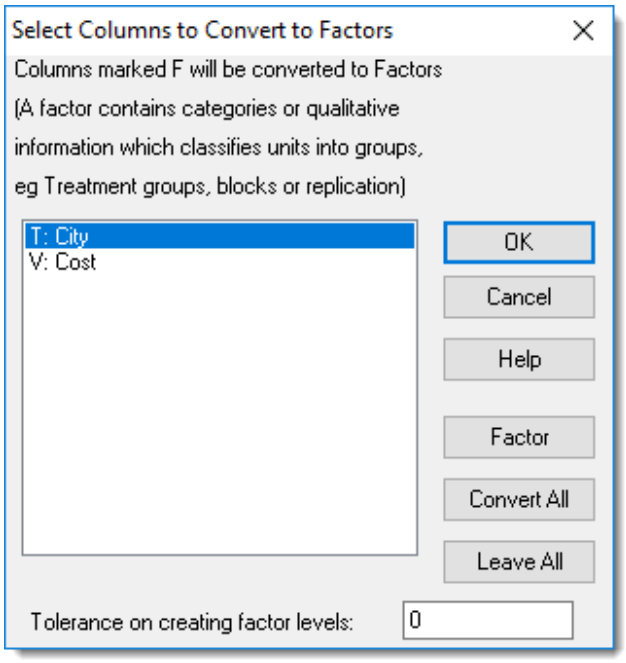


Figure 1.20

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

Row	WorkSheet	City	Year	Cost	Sold
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 1.21

1.3 Appending data from multiple files

If the data you wish to append are in multiple files or on pages in Genstat workbook files (.gwb) these can be appended into a single file. We'll demonstrate this by appending 5 Excel files into a single spreadsheet. The 5 Excel files, `Grazing 1.xls` - `Grazing 5.xls` 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).

Select Spread | New | Append Multiple Files. In the Append Multiple Files dialog (Figure 1.22) click the Browse button ... and navigate to `C:\Program Files\Gen22Ed\Data`. (You could use the Browse button multiple times if the files were in different directories).

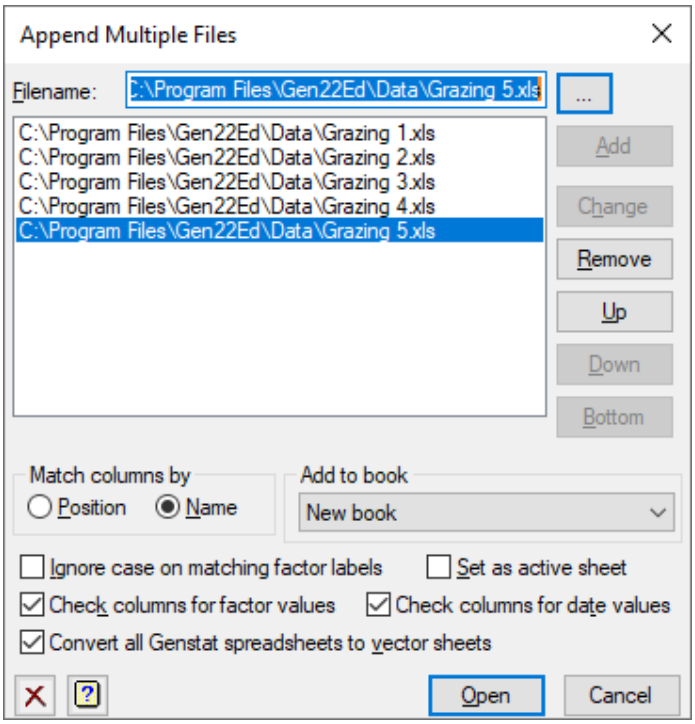


Figure 1.22

In the Open File dialog (Figure 1.23) select Other Spreadsheet Files (*.xls...,*.ods) from the dropdown list so that only non-Genstat spreadsheets are listed. Import the Excel files by clicking `Grazing1.xls` then hold the Shift key and click `Grazing5.xls`.

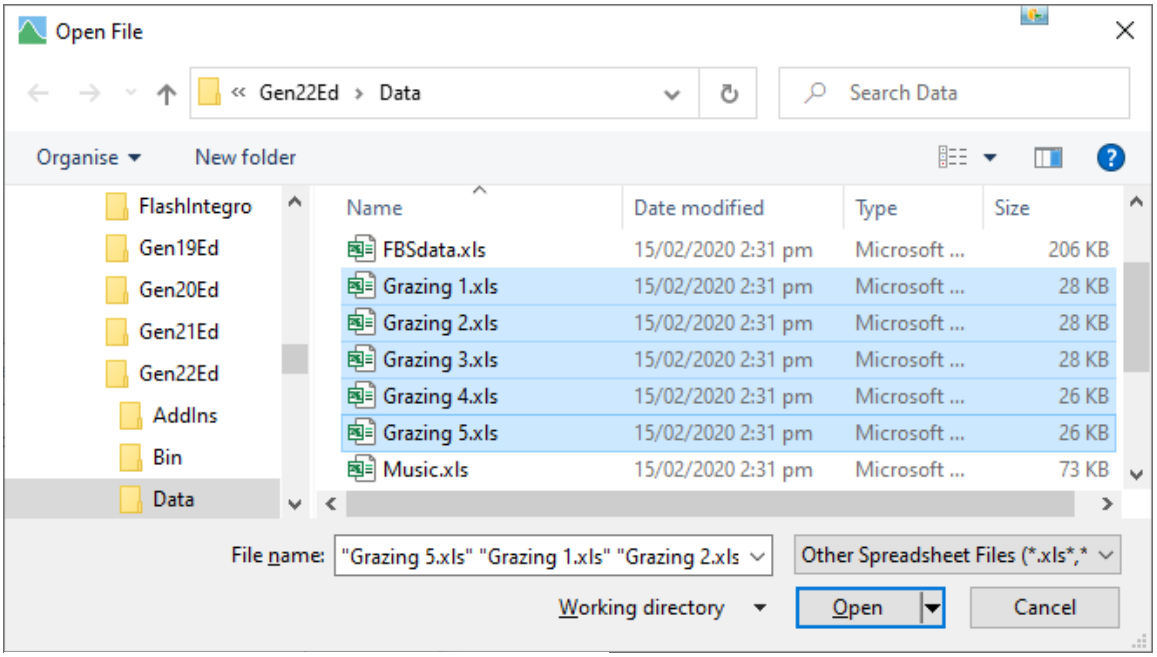


Figure 1.23

Clicking Open displays the dialog shown previously in Figure 1.22, which allows you to browse to your files and specify how the columns are matched (by name or by column position). 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. 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 by selecting a file then using the Up, Down and Bottom buttons. Leave the settings at their defaults then click Open. This will open each of the selected files in turn. As they are Excel files, you will be prompted for the options for reading in an Excel file 5 times, once for each file.

Figure 1.24 shows the Excel import options dialog for the first file `Grazing 1.xls`. We do not need to change any of the options to import the data, so we'll click Finish. Four identical dialogs will display in turn and you can just click Finish for each one.

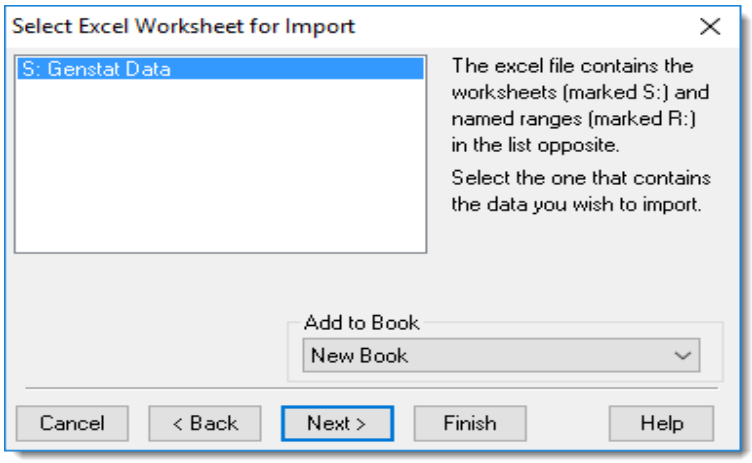
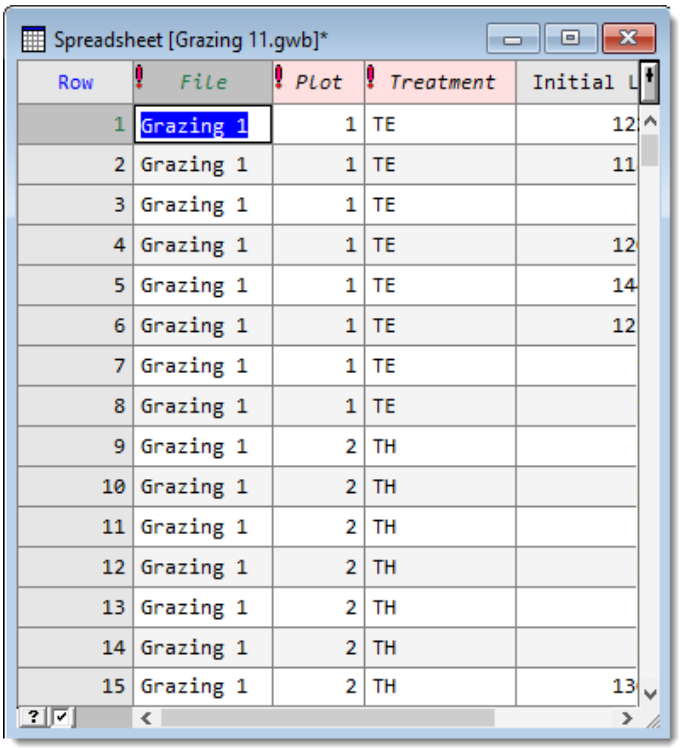


Figure 1.24

After the final Excel import dialog has been closed the resulting spreadsheet will be opened, as shown in Figure 1.25. 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.



Row	File	Plot	Treatment	Initial L
1	Grazing 1	1	TE	12
2	Grazing 1	1	TE	11
3	Grazing 1	1	TE	
4	Grazing 1	1	TE	12
5	Grazing 1	1	TE	14
6	Grazing 1	1	TE	12
7	Grazing 1	1	TE	
8	Grazing 1	1	TE	
9	Grazing 1	2	TH	
10	Grazing 1	2	TH	
11	Grazing 1	2	TH	
12	Grazing 1	2	TH	
13	Grazing 1	2	TH	
14	Grazing 1	2	TH	
15	Grazing 1	2	TH	13

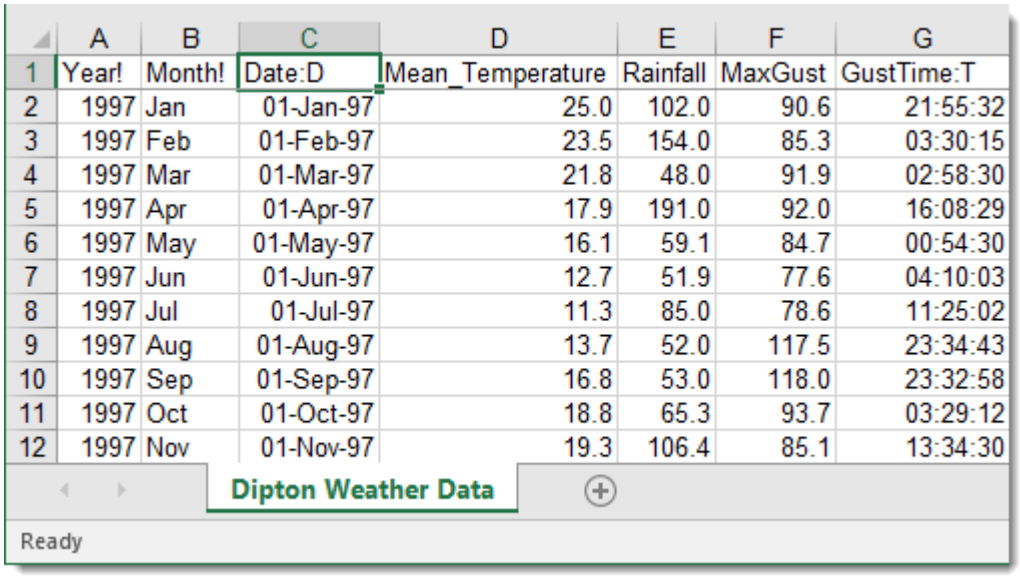
Figure 1.25

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 a suffix of `:D` and a time column should have a suffix of `:T`.

In 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 import Excel data 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 1.26 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.

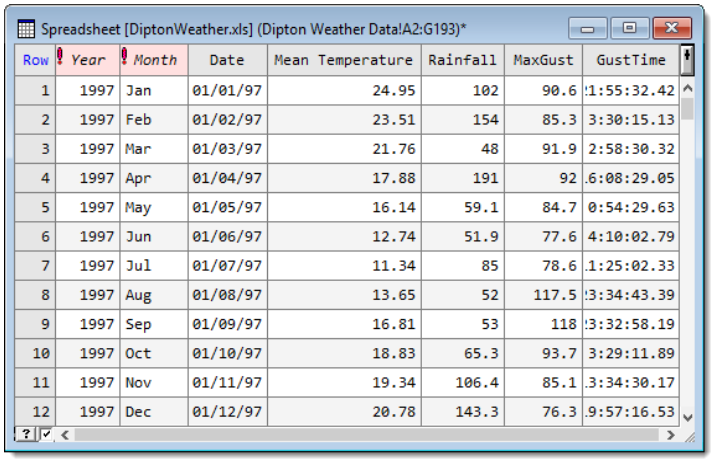


	A	B	C	D	E	F	G
1	Year!	Month!	Date:D	Mean_Temperature	Rainfall	MaxGust	GustTime:T
2	1997	Jan	01-Jan-97	25.0	102.0	90.6	21:55:32
3	1997	Feb	01-Feb-97	23.5	154.0	85.3	03:30:15
4	1997	Mar	01-Mar-97	21.8	48.0	91.9	02:58:30
5	1997	Apr	01-Apr-97	17.9	191.0	92.0	16:08:29
6	1997	May	01-May-97	16.1	59.1	84.7	00:54:30
7	1997	Jun	01-Jun-97	12.7	51.9	77.6	04:10:03
8	1997	Jul	01-Jul-97	11.3	85.0	78.6	11:25:02
9	1997	Aug	01-Aug-97	13.7	52.0	117.5	23:34:43
10	1997	Sep	01-Sep-97	16.8	53.0	118.0	23:32:58
11	1997	Oct	01-Oct-97	18.8	65.3	93.7	03:29:12
12	1997	Nov	01-Nov-97	19.3	106.4	85.1	13:34:30

Figure 1.26

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.

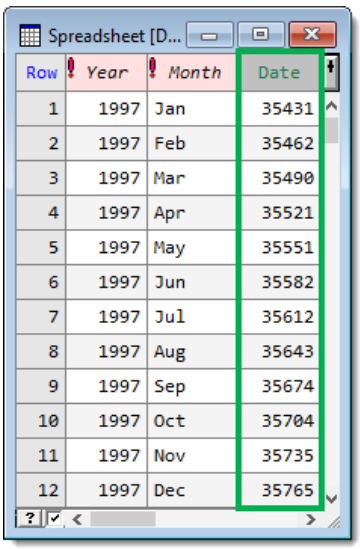
Use either File | Open or File | Open Example Data Sets to load the data into a Genstat spreadsheet. This gives the spreadsheet shown in Figure 1.27. Note how `Date` and `GustTime` are displayed in date format. If the `:D` suffix was missing from the column name for `Date`, then the column would be displayed as in Figure 1.28, 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	01/01/97	24.95	102	90.6	1:55:32.42
2	1997	Feb	01/02/97	23.51	154	85.3	3:30:15.13
3	1997	Mar	01/03/97	21.76	48	91.9	2:58:30.32
4	1997	Apr	01/04/97	17.88	191	92	6:08:29.05
5	1997	May	01/05/97	16.14	59.1	84.7	0:54:29.63
6	1997	Jun	01/06/97	12.74	51.9	77.6	4:10:02.79
7	1997	Jul	01/07/97	11.34	85	78.6	1:25:02.33
8	1997	Aug	01/08/97	13.65	52	117.5	3:34:43.39
9	1997	Sep	01/09/97	16.81	53	118	3:32:58.19
10	1997	Oct	01/10/97	18.83	65.3	93.7	3:29:11.89
11	1997	Nov	01/11/97	19.34	106.4	85.1	3:34:30.17
12	1997	Dec	01/12/97	20.78	143.3	76.3	9:57:16.53

Figure 1.27



Spreadsheet [D...

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

Figure 1.28

You can change the format used to display the date or time using the Spread | Column | Attributes/Format menu (shown in Figure 1.29). For a numerical column, the Numerical Format will be set to Date. To change the date format, click the Date Format button to open the dialog shown in Figure 1.30. You can select a date or time format from the dropdown list. If the wrong base date has been selected, this can be changed using the Start date from option. There are 49 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 2 or 4 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 1.31.

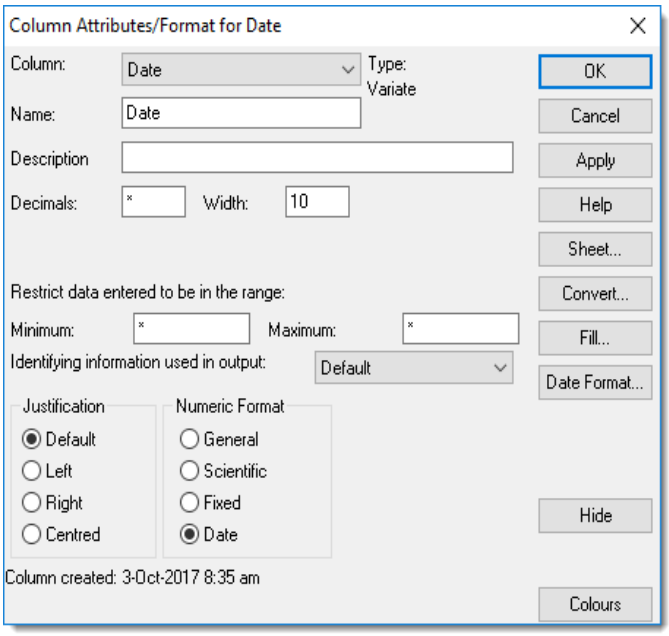


Figure 1.29

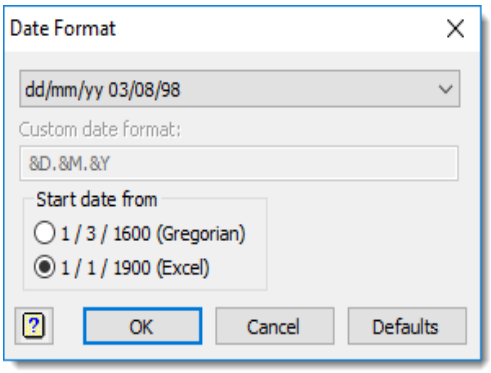


Figure 1.30

If for some reason you needed to change base date for a column, you can use the Spread | Calculate | Rebase Dates menu.

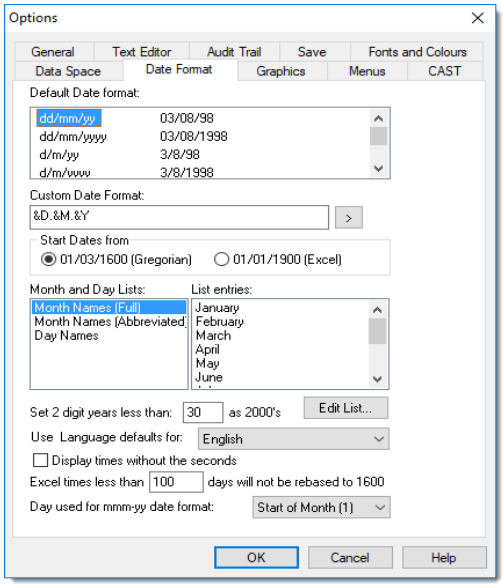


Figure 1.31

If this is used on the column `Date`, it would give the prompt shown in Figure 1.32. Normally this is not necessary as Genstat keeps track of the 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.

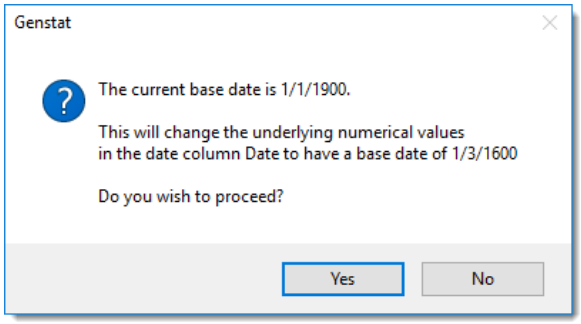


Figure 1.32

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, select Spread | New | 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.

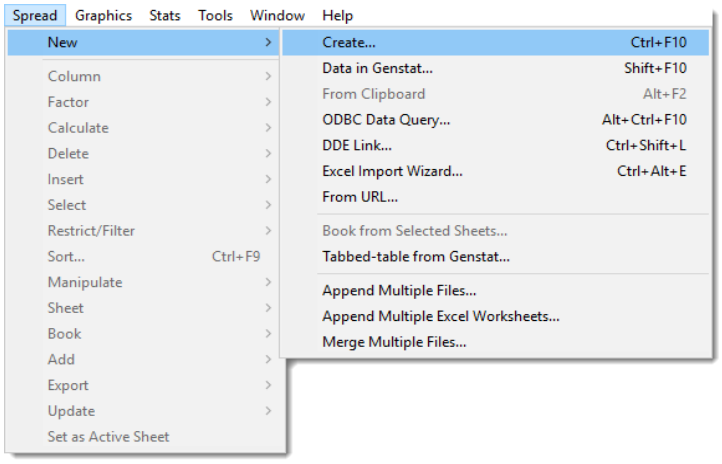


Figure 2.1

When a spreadsheet is created, it can either be opened within a new 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 12 rows and 4 columns. It does not matter if you do not know the number of rows and columns needed initially for entering your data, as 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.

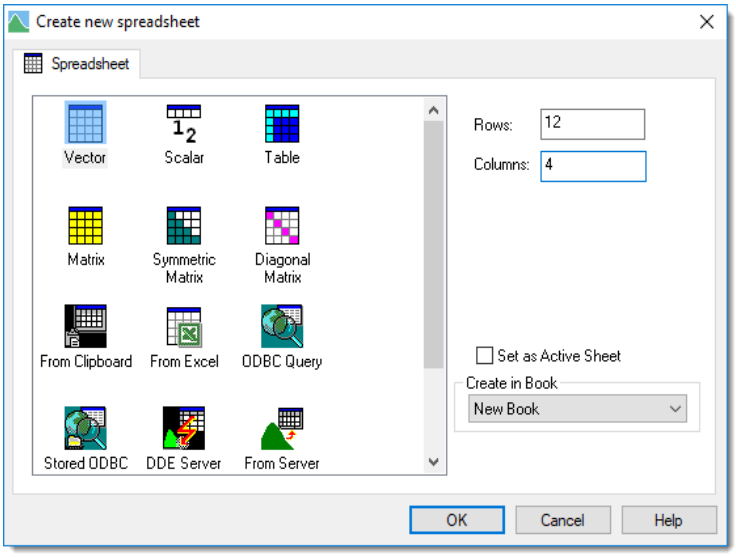


Figure 2.2

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 and displayed in the Data pane on the left 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.

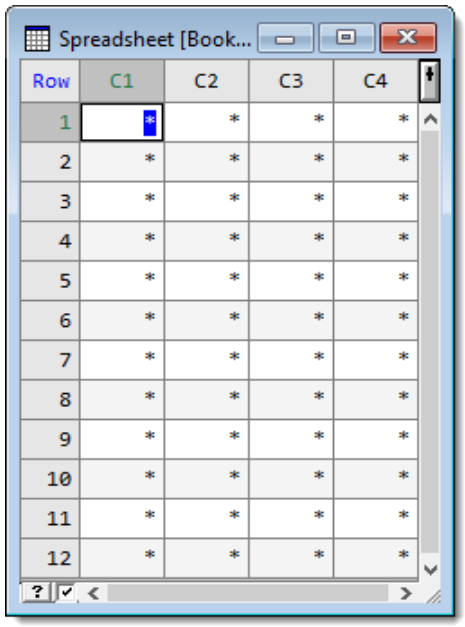


Figure 2.3

To rename the columns select Spread | Column | Rename, which opens the menu shown in Figure 2.4. Put the cursor in column C1 and rename it as Drench, then click OK.

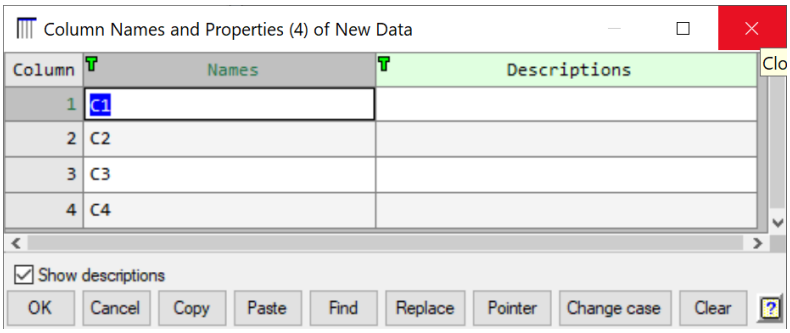


Figure 2.5

An alternative way to rename a column is to right-click the column heading and select Rename. This opens the dialog shown in Figure 2.5. Do this for column C2 and type the new name Rep then click OK.

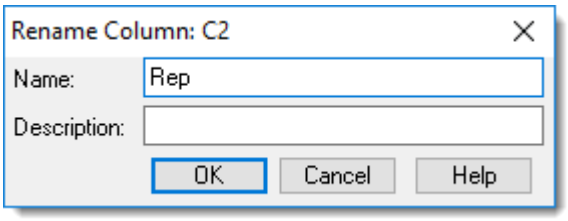


Figure 2.4

Use one of these methods now to change C3 to Lwt1 and C4 to Lwt2.

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. This produces the menu shown in Figure 2.6. Select Convert to Factor to open the dialog shown in Figure 2.7. 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 3 groups: Control, Drenched once and Drenched twice, so we have entered 3 in the Number of Levels field. We now want to change the labels to represent the 3 groups.

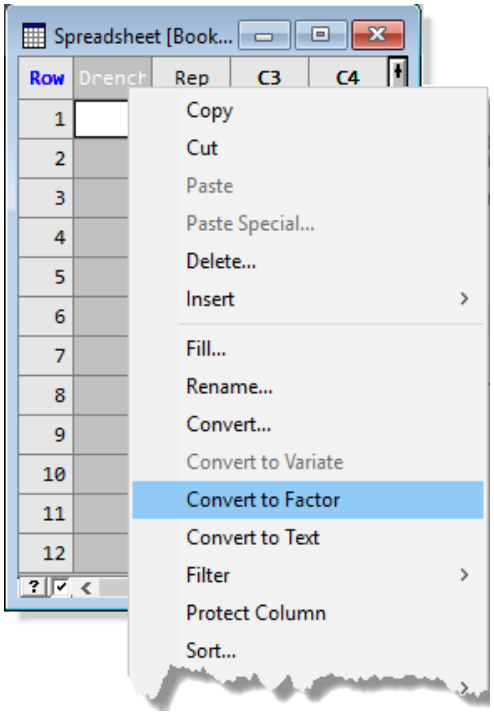


Figure 2.6

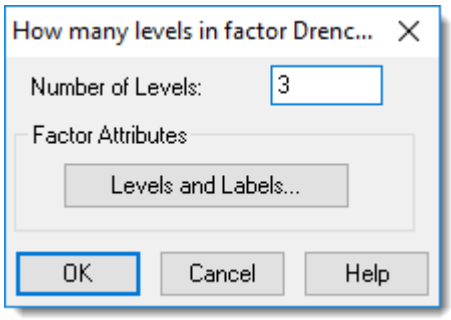


Figure 2.7

Click the Levels and Labels button to open the dialog in Figure 2.8. In the Labels field enter `Control` for group 1 and press Enter or the down arrow to apply this label. For group 2 label this `Once` and finally for group 3 label this group as `Twice`. We can also apply different coloured backgrounds for each group by clicking the colour wheel icon in the Colour column. This opens the colour selection dialog in Figure 2.9. Choose a colour for your selected group by clicking a colour box then click OK. Colour the other 2 groups in the same manner.

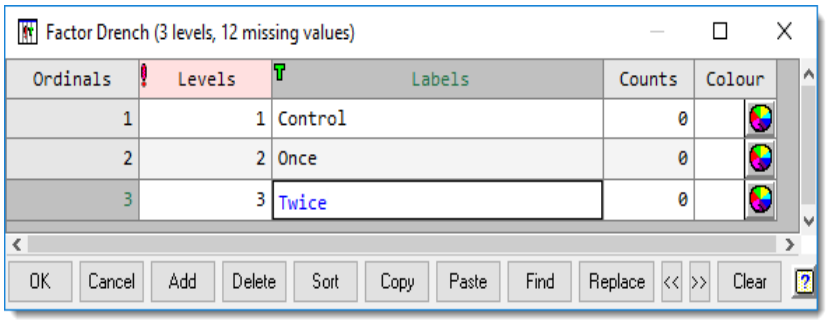


Figure 2.8

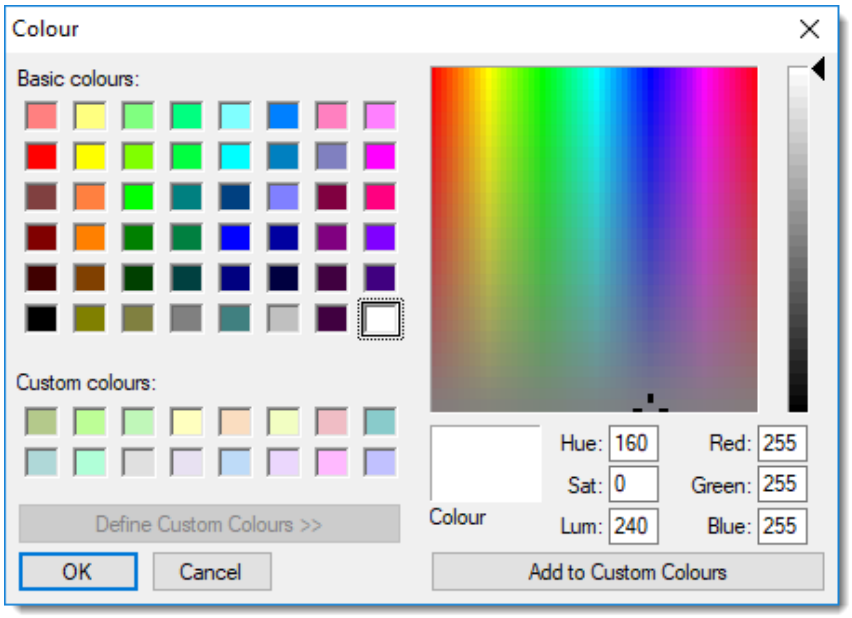


Figure 2.9

Clicking OK again 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 2.10).

You can now enter the label names by typing directly into a cell, or by double-clicking on a cell and selecting the appropriate label from the list, as shown in Figure 2.11. 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.

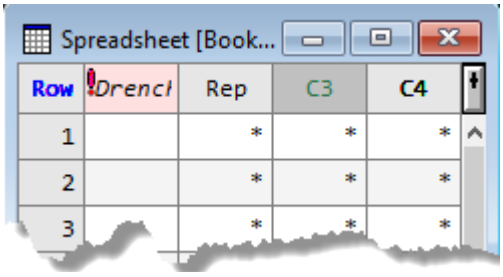


Figure 2.10

Figure 2.11

Note that if 2 labels begin with the same character 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. You can type upper or lower case unless two items use the same labels and are only differentiated by case (e.g., `a` and `A` are both labels of the factor). The cell backgrounds will fill with your selected colours as you enter each label. Figure 2.12 shows the column complete with the new factor labels.

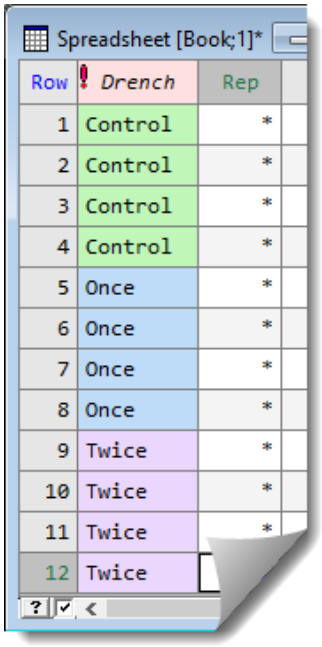


Figure 2.12

We now enter the data into the columns `Lwt1` and `Lwt2` using Figure 2.13 as our guide. 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 you make a mistake, you can double-click a cell to edit its value.

The column `Rep` contains patterned data with the values 1,2,3 and 4 repeated 3 times. We'll use Genstat's Fill menu to automatically fill this column with patterned data. Select Spread | Calculate | Fill to open the dialog shown in Figure 2.14. Select the column `Rep` from the dropdown list then enter 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 2.13.

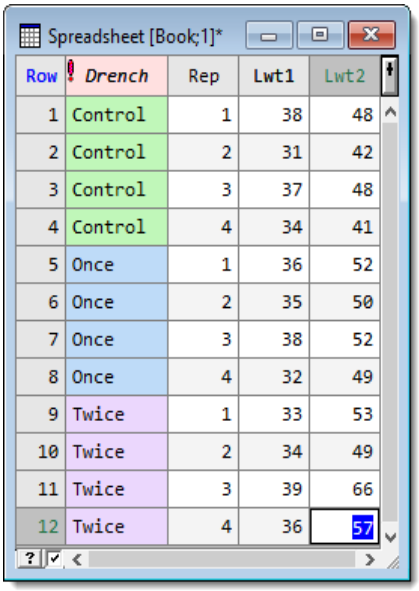


Figure 2.13

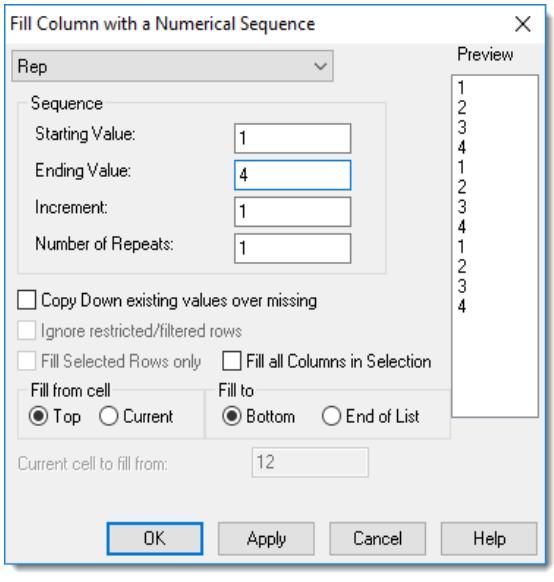


Figure 2.14

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 Spread | Sheet | Verify, which opens the dialog shown in Figure 2.15. To choose columns to verify either double-click them or select the column names and click Verify. Columns selected for verification will be prefixed by 'V:'. In Figure 2.15 we have double-clicked on the names `Lwt1` and `Lwt2` to specify that they are to be verified.

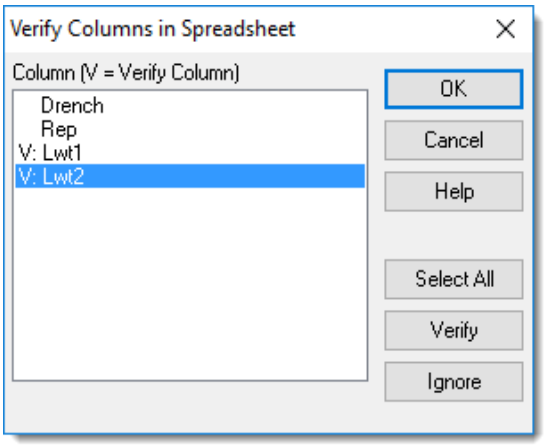


Figure 2.15

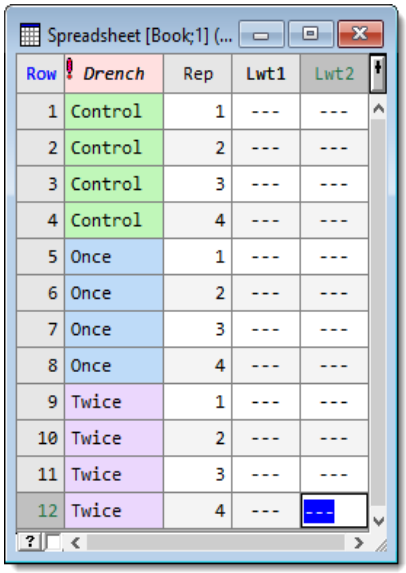


Figure 2.16

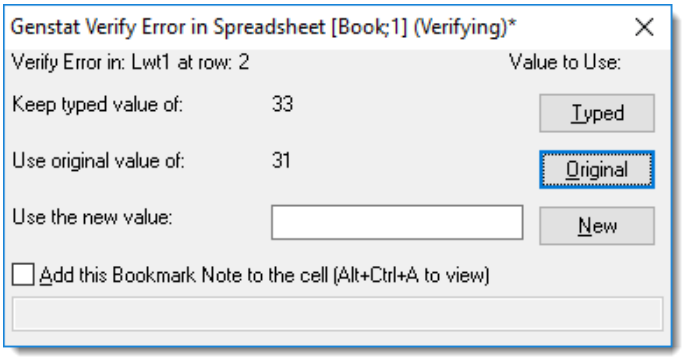


Figure 2.17

Clicking OK changes the columns `Lwt1` and `Lwt2` in the spreadsheet to display three minus (−) characters in place of the values (see Figure 2.16). To verify the data, we now re-enter the values in these cells. 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 dialog shown in Figure 2.17.

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 the Typed button to register this value. Clicking the Typed button creates a new spreadsheet containing a record of the mismatch in the data entry (Figure 2.18). This spreadsheet will appear underneath the sheet you're working with; you will need to move your current spreadsheet to one side with the mouse to see the new one. Each row within the new spreadsheet contains

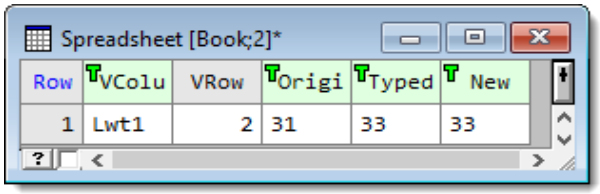


Figure 2.18

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`. On entering the last value of the verification in row 12 of the column `Lwt2` the prompt in Figure 2.19 will display.

This prompt allows you to set the verified columns as read-only to protect them from any further changes. Clicking Yes 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.

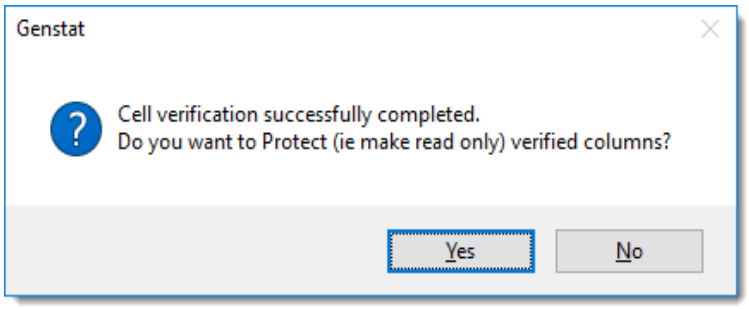


Figure 2.19

You can set or remove the protection for a column at any time. To remove the column protection on `Lwt1` and `Lwt2` select Spread | Column | Protection. This opens the dialog shown in Figure 2.20.

The columns within the spreadsheet are listed, with protected columns prefixed 'P:'. 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 Unprotect will also remove protection. You can protect a column in a similar way by double-clicking the name in the list or clicking the Protect button. Clicking OK returns you to the spreadsheet and removes the blue background from the column titles.

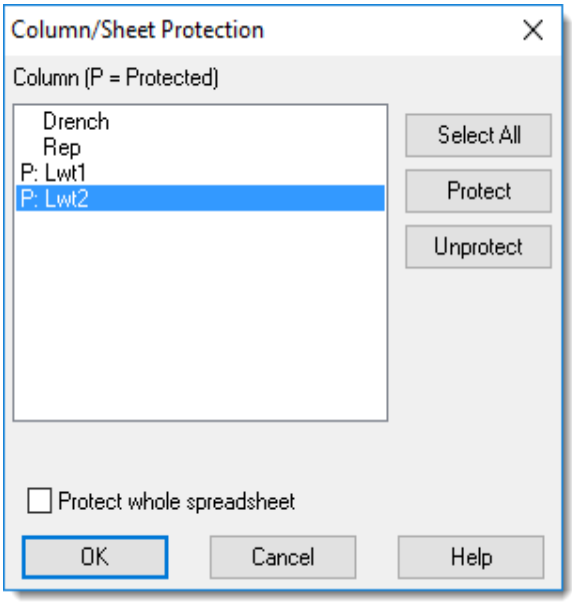


Figure 2.20

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 we created 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 `Drench.gsh`.

Select Spread | Sheet | Compare to open the dialog shown in Figure 2.21. The Data Source option identifies where the data that you wish to compare are located. Our comparison data are in a .gsh file so select File. Click Browse then navigate to the file location as shown in Figure 2.21. The remaining options on the menu control how the comparison is to be made.

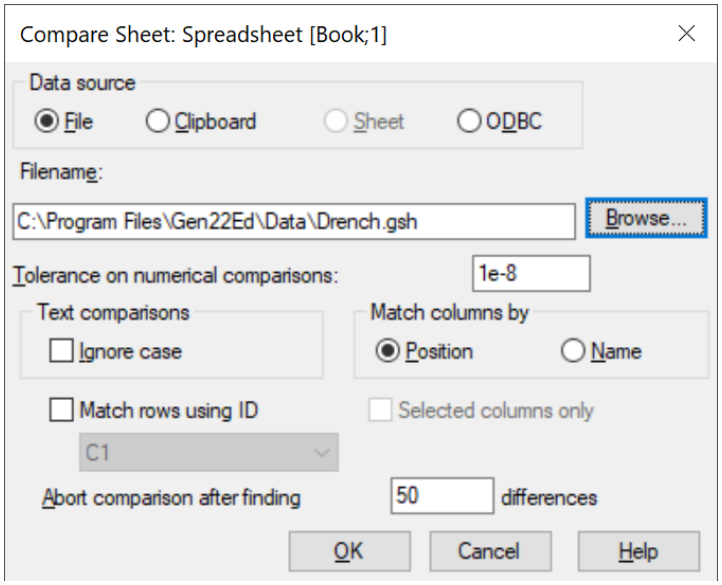


Figure 2.21

Leaving the default settings and clicking on OK produces a dialog (Figure 2.22) to warn that the sheets are different, and prints a report in the Output Window, as shown below in Figure 2.23. 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*.

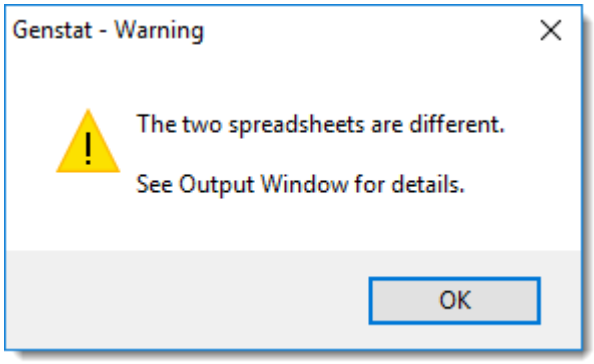


Figure 2.22

```
"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. "
```

Figure 2.23

2.3 Inserting and deleting rows or columns

Columns and rows can be deleted using the Delete options on the Spread menu. (If your spreadsheet is hidden by the Output window, re-display it by clicking the Data tab at the bottom left of the screen, then move the cursor up and double-click `Book;1`.) To delete the column `Rep` click anywhere on the column and select Spread | Delete | Current Column.

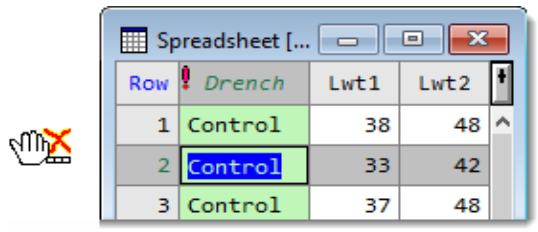


Figure 2.24

Rows can be deleted by clicking on the row number and dragging the row outside the spreadsheet. Figure 2.24 illustrates this being done with `Row 2`. You can select and drag multiple rows (or columns) for deletion in a similar fashion.

New columns and rows can be inserted using the facilities available within the options on the Spread | Insert menu. To insert a new row at the bottom of the spreadsheet, click on the last row of the spreadsheet and select Spread | Insert | Row After Current Row. This will add a new row as shown in Figure 2.25. New values default to missing values (represented by asterisks for numbers or empty cells for labels or strings).

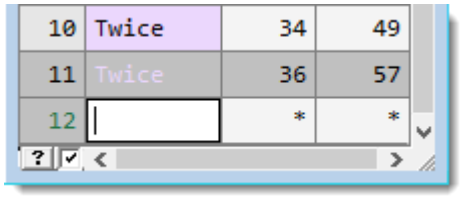


Figure 2.25

To insert a new column, click in the `Drench` column and select Spread | Insert | Column after Current Column to open the dialog shown in Figure 2.26. Here you can choose what type of data structure the new column will be, give the column a name and set an initial value for each cell.

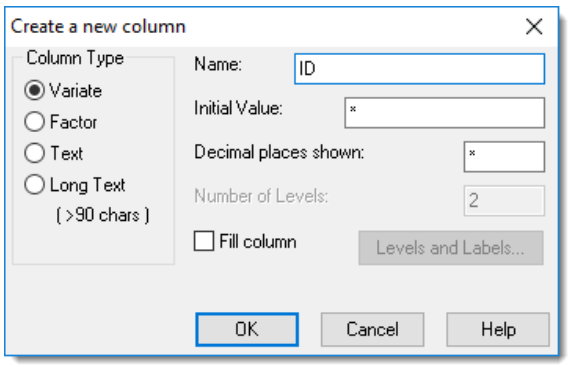


Figure 2.26

Select Variate from the Column Type, enter the name `ID` and click OK to produce a new column, initialized with missing values, as shown in Figure 2.27.

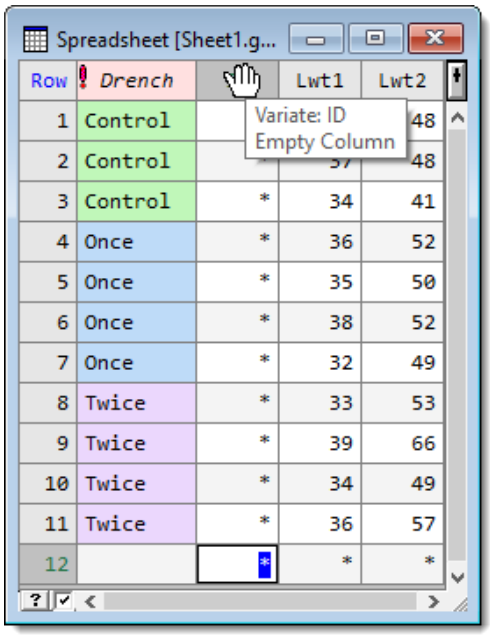


Figure 2.27

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 Spread | Column | Duplicate.

This opens the dialog shown in Figure 2.28 where we have selected the column `Lwt1` and entered a new name for the duplicate column, `vLwt1` 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 2.29.

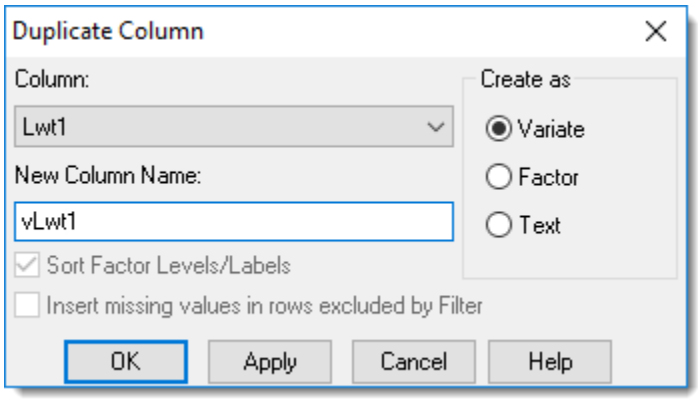
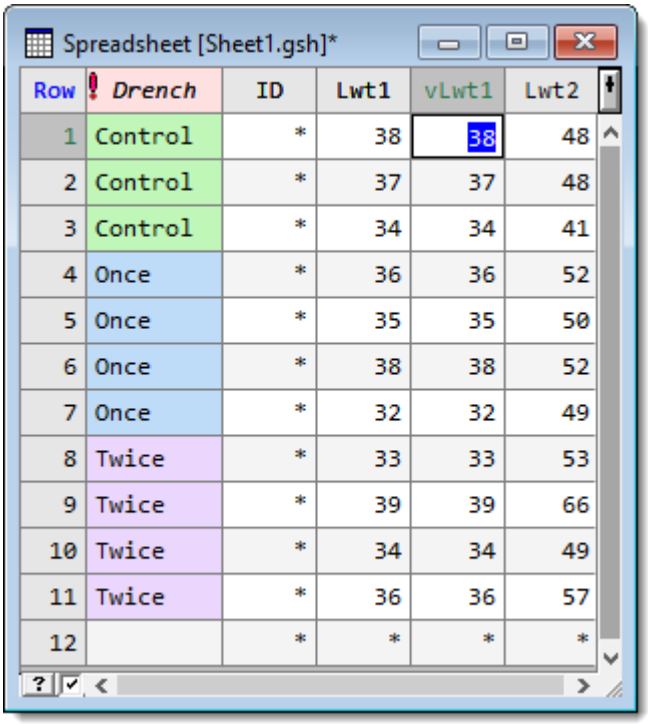


Figure 2.28



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	*	39	39	66
10	Twice	*	34	34	49
11	Twice	*	36	36	57
12		*	*	*	*

Figure 2.29

2.4 Exercise

The following data are from an experiment assessing the durability of four different types of carpet: 4 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 Spread | Calculate | Fill 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, and d. Using the Verify menu from the Sheet option on the Spread menu, check that you have entered the data correctly. The comparison 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 by selecting from the main menu Data | Clear All Data.

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. If you have already closed the Drench.gsh spreadsheet, re-open it. Click outside the spreadsheet to load the data into the central data pool. We'll form a new spreadsheet in a new book containing only the columns `Drench` and `Lwt2`.

From the menu select Spread | New | Data in Genstat to generate the Load Spreadsheet dialog in Figure 3.1. In this dialog we select `Drench` and `Lwt2` then click 

to move them to the Data to Load field. Select New Book from the Load in book list the click Load.

The resulting spreadsheet is shown in Figure 3.2.

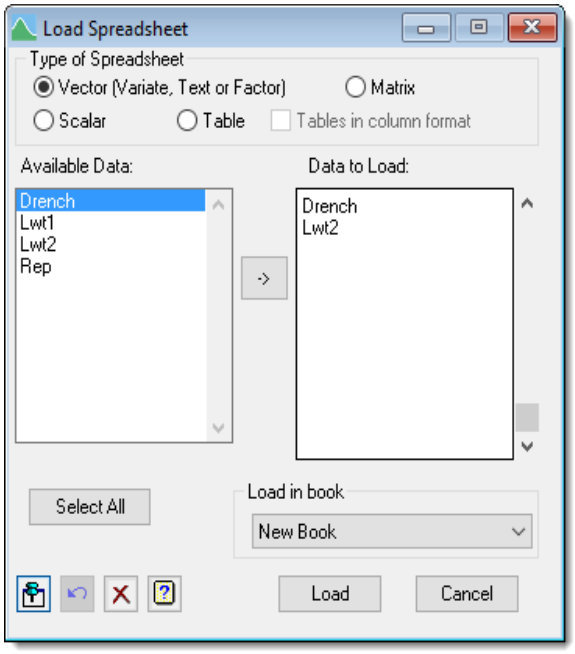


Figure 3.1

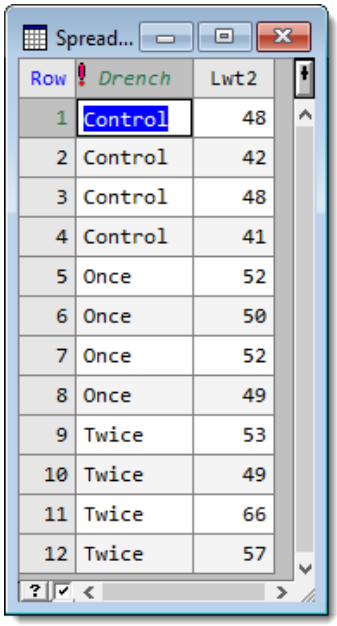


Figure 3.2

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

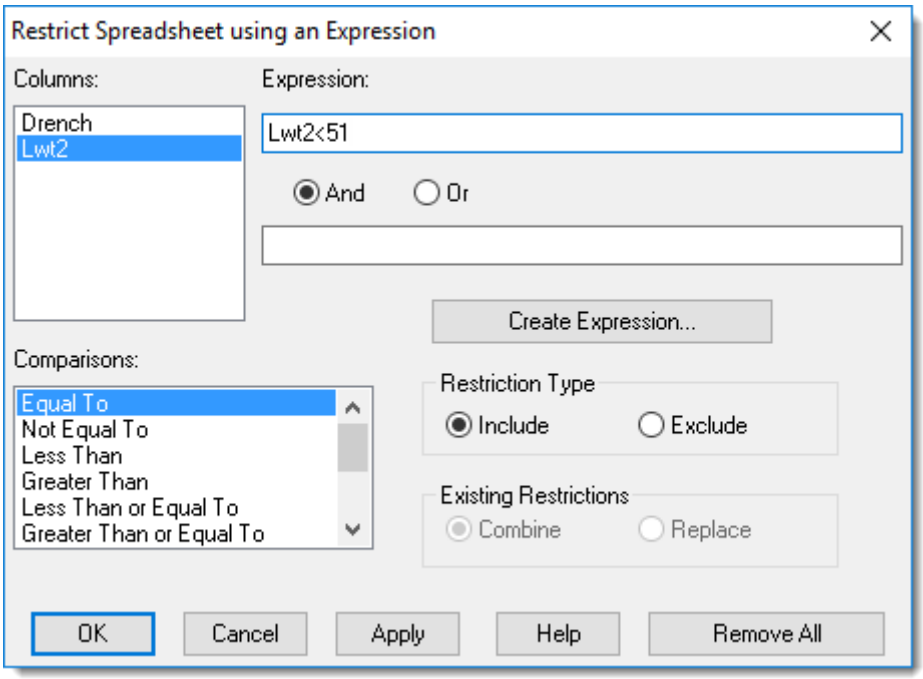


Figure 3.3

In our example we want to *include* all the units within the restriction where the units in `Lwt2` are lower than 51. To create the expression for this restriction we double-click `Lwt2` in the Columns list to move it into the first Expression field. We then double-click the Less Than option from the Comparison list which puts a '<' symbol into the Expression field and then type '51'. We select the Include option as the Restriction Type then click OK.

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.)

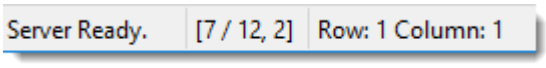


Figure 3.4

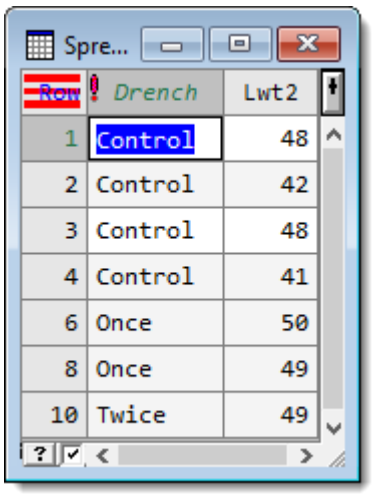


Figure 3.5

This is illustrated in Figure 3.6, where we use the PRINT directive, to print Lwt1 and Lwt2.

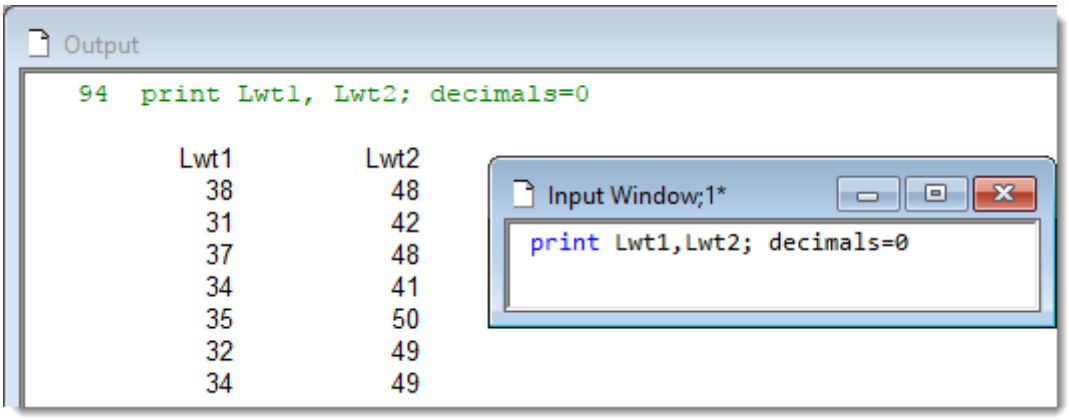


Figure 3.6

Notice in Figure 3.6 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). To use the PRINT directive, open a new text window by clicking the button at the top left of the screen shown in Figure 3.7. Type the text shown in Figure 3.6 then from the menu select Run | Submit Line.

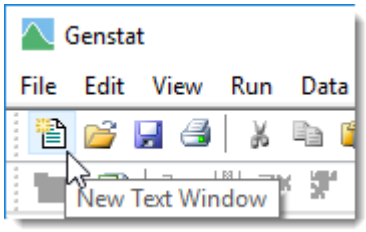


Figure 3.7

If your spreadsheet is hidden, re-display it but clicking the Window tab at the bottom left of the screen then double-click Book;2 in the Window view. The restricted units are not discarded and can be viewed in the spreadsheet in an alternative colour. To do this select Spread | Restrict/Filter | Display Excluded Rows. This will display all the rows in the spreadsheet, but with the restricted out (excluded) rows shown in red (the default colour); see Figure 3.8.

You can also toggle the display of the restricted rows by clicking the '+' button in the top-right corner of the spreadsheet.

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.

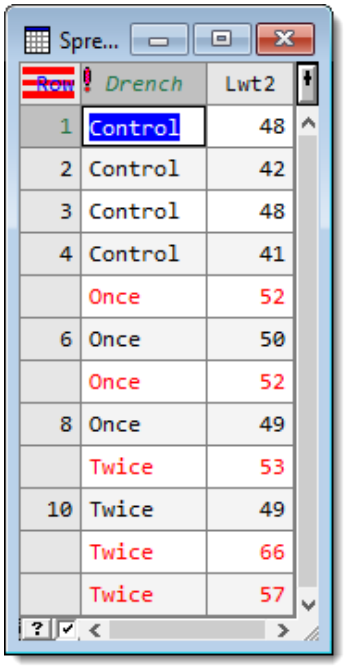


Figure 3.8

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 dialog, as shown in Figure 3.9. To open the dialog, select Spread | Restrict/Filter | To Groups (factor levels). This 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 3.10.

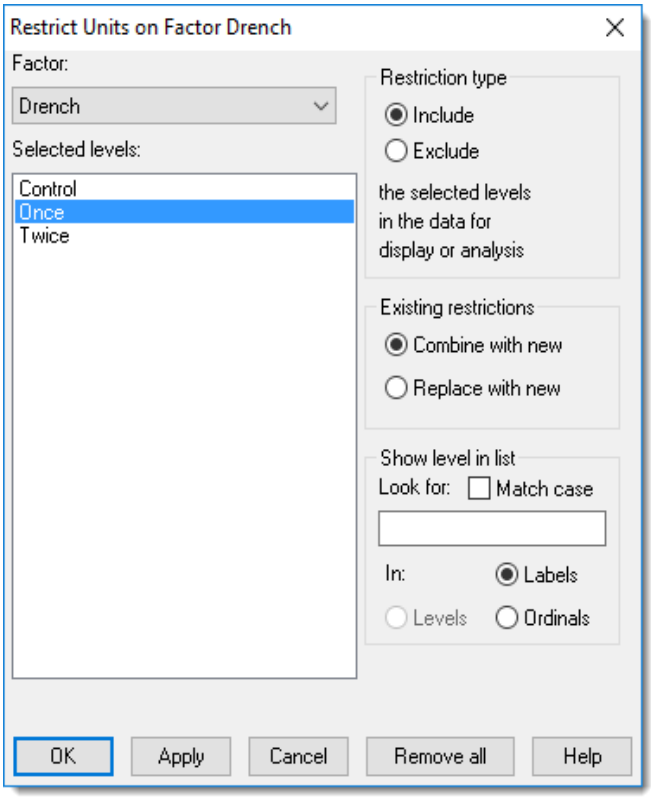


Figure 3.9

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 dialog you can create a restriction by combining the two logical conditions into a single condition using both the expression boxes.

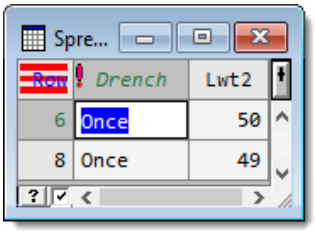


Figure 3.10

Figure 3.11 shows how to do this for our example. First, we remove the current restriction. From the menu select Spread | Restrict/Filter | By Logical Expression then click 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 field. Then, in the second field 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 Inclusion in the list of Comparisons.

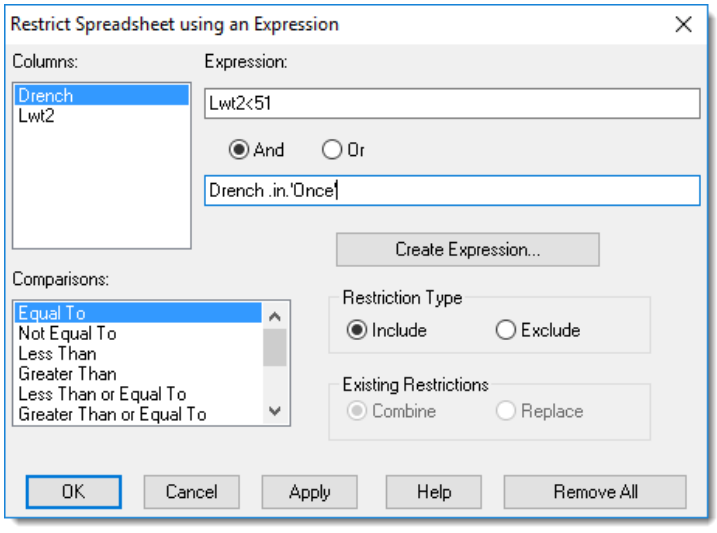


Figure 3.11

To combine these two conditions, we have selected the And option between the Expression fields; 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 OK produces the same spreadsheet as shown in 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 Spread | Restrict/Filter | Remove All or by clicking the clear restrictions button in the toolbar 

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. To open this, select Data | Subset. You can also define the restriction by specifying the rows in the spreadsheet explicitly. The rows are selecting using the Spread | Select 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 in months where 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 By Logical 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. Now select File | Open to locate and re-open `Drench.gsh` shown in Figure 3.12.

Spreadsheet [Dr...

Row	Drench	Rep	Lwt1	Lwt2
1	Control	1	38	48
2	Control	2	31	42
3	Control	3	37	48
4	Control	4	34	41
5	Once	1	36	52
6	Once	2	35	50
7	Once	3	38	52
8	Once	4	32	49
9	Twice	1	33	53
10	Twice	2	34	49
11	Twice	3	39	66
12	Twice	4	36	57

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 Spread | Sort; this opens the dialog shown in Figure 3.13.

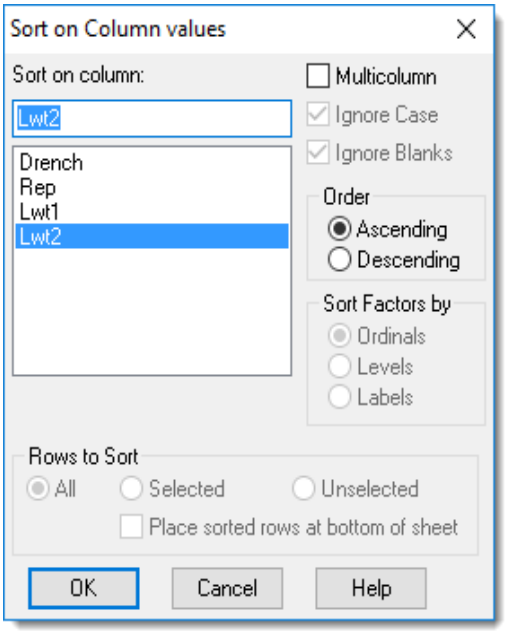
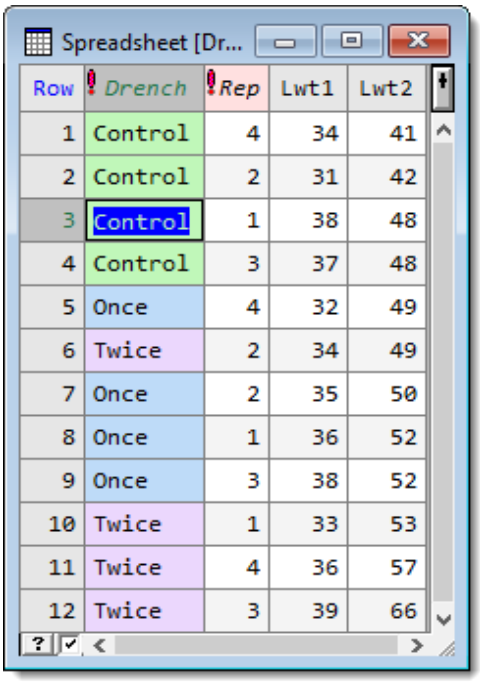


Figure 3.13

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.



Row	Drench	Rep	Lwt1	Lwt2
1	Control	4	34	41
2	Control	2	31	42
3	Control	1	38	48
4	Control	3	37	48
5	Once	4	32	49
6	Twice	2	34	49
7	Once	2	35	50
8	Once	1	36	52
9	Once	3	38	52
10	Twice	1	33	53
11	Twice	4	36	57
12	Twice	3	39	66

Figure 3.14

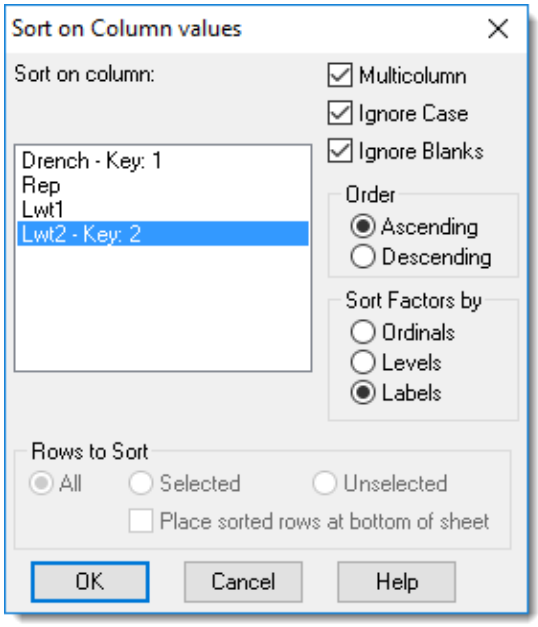


Figure 3.15

Select Spread | Sort again then select `Drench` as the first column that we are going to 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 OK produces the spreadsheet shown in 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 Ctrl key down and click on the second row of the selection, and so on (keeping the Ctrl key 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.

Spreadsheet [Dr...

Row	Drench	Rep	Lwt1	Lwt2
1	Control	2	31	42
2	Control	4	34	41
3	Control	3	37	48
4	Control	1	38	48
5	Once	4	32	49
6	Once	2	35	50
7	Once	1	36	52
8	Once	3	38	52
9	Twice	1	33	53
10	Twice	2	34	49
11	Twice	4	36	57
12	Twice	3	39	66

Figure 3.16

Spreadsheet [Dr...

Row	Drench	Rep	Lwt1	Lwt2
1	Control	2	31	42
2	Control	4	34	41
3	Control	3	37	48
4	Control	1	38	48
5	Once	4	32	49
6	Once	2	35	50
7	Once	1	36	52
8	Once	3	38	52
9	Twice	1	33	53
10	Twice	2	34	49
11	Twice	4	36	57
12	Twice	3	39	66

Figure 3.17

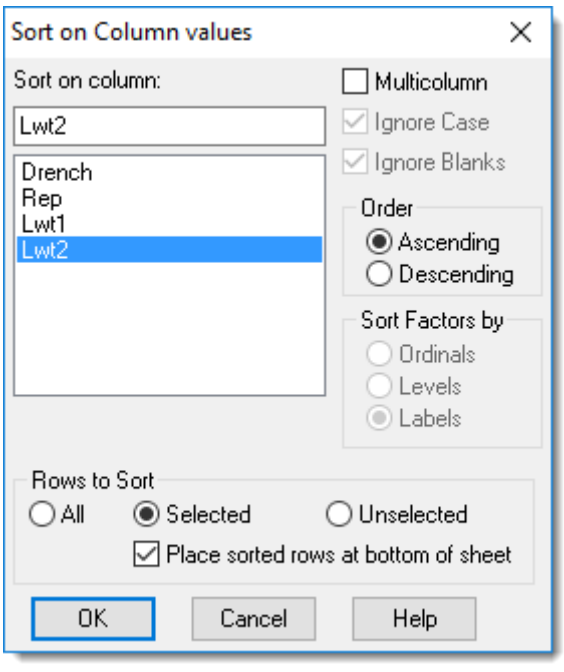


Figure 3.18

De-select Multicolumn to clear your previous sort then select Lwt2 from the Sort on column list. Now select 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 Place sorted rows at bottom of sheet.

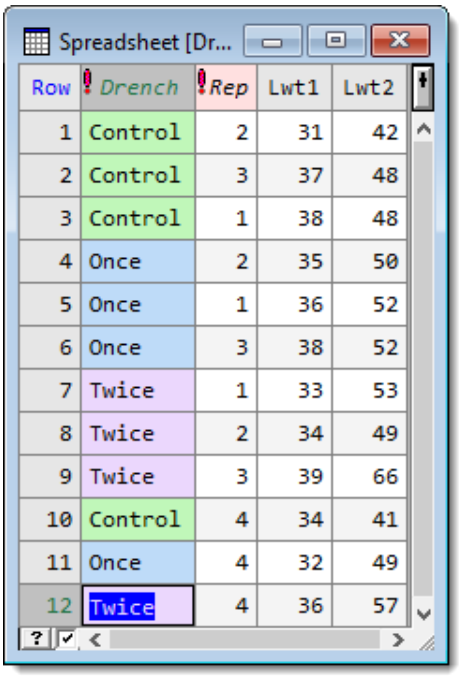
Figure 3.19 shows the results of this process on our example spreadsheet. Other facilities for sorting data are provided by the Genstat SORT directive.

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 by selecting Data | Clear All Data.



Row	Drench	Rep	Lwt1	Lwt2
1	Control	2	31	42
2	Control	3	37	48
3	Control	1	38	48
4	Once	2	35	50
5	Once	1	36	52
6	Once	3	38	52
7	Twice	1	33	53
8	Twice	2	34	49
9	Twice	3	39	66
10	Control	4	34	41
11	Once	4	32	49
12	Twice	4	36	57

Figure 3.19

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.

Select File | Open then locate and open Toysales.xls. 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 dialog (Figure 3.20), select New Book in the Add to Book list, and click Finish.

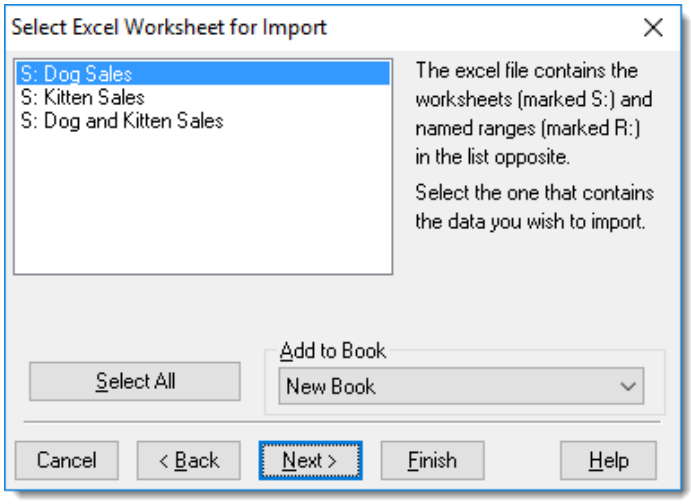
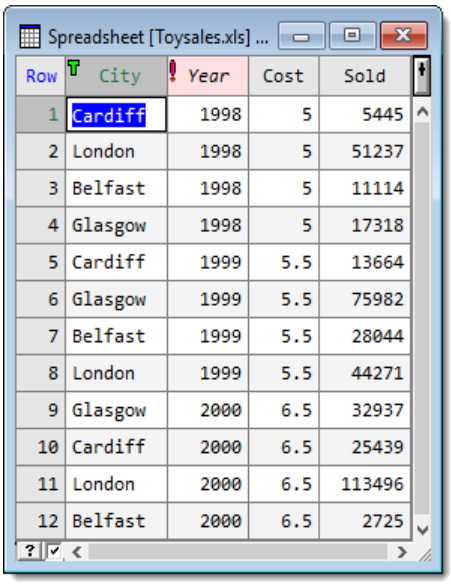


Figure 3.20

The resulting spreadsheet is shown in Figure 3.21.



Row	City	Year	Cost	Sold
1	Cardiff	1998	5	5445
2	London	1998	5	51237
3	Belfast	1998	5	11114
4	Glasgow	1998	5	17318
5	Cardiff	1999	5.5	13664
6	Glasgow	1999	5.5	75982
7	Belfast	1999	5.5	28044
8	London	1999	5.5	44271
9	Glasgow	2000	6.5	32937
10	Cardiff	2000	6.5	25439
11	London	2000	6.5	113496
12	Belfast	2000	6.5	2725

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). To open this, select Spread | Manipulate | Append. We select File as our data source and use the Browse button to select the file `Toysales.xls`. 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 field. 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 fields. In Figure 3.22 we have entered `Kitten` to label the appended data and `Dog` to represent the original data.

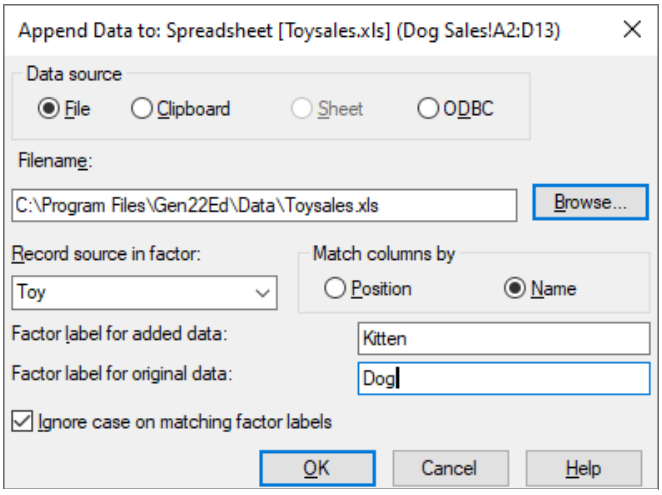
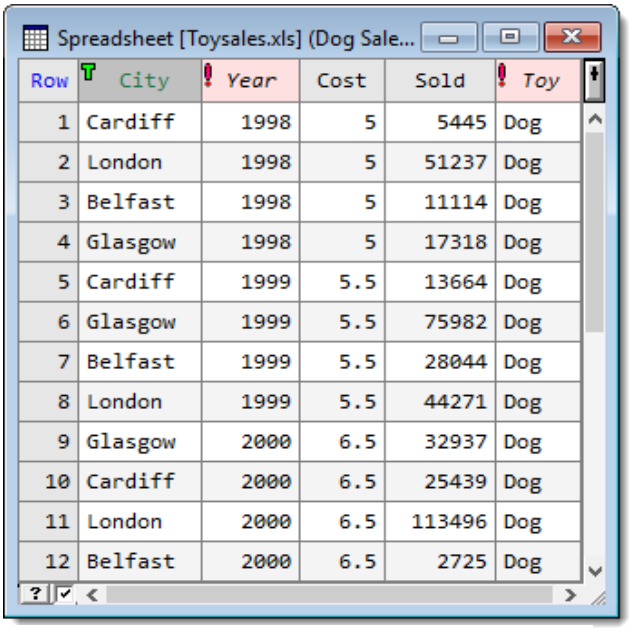


Figure 3.22

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

Figure 3.23

3.5.1 Appending data from multiple worksheets

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 Spread | New | Append Multiple Excel Worksheets. Select `Toysales.xls` as before then click Open. This opens the dialog 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 dialog hold down the Ctrl key then select both the Kitten Sales and Dog Sales worksheets in the Select Multiple Sheets/Ranges list. On the right side, deselect Suggest columns to be factors as we want to import the data 'as is'. 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.

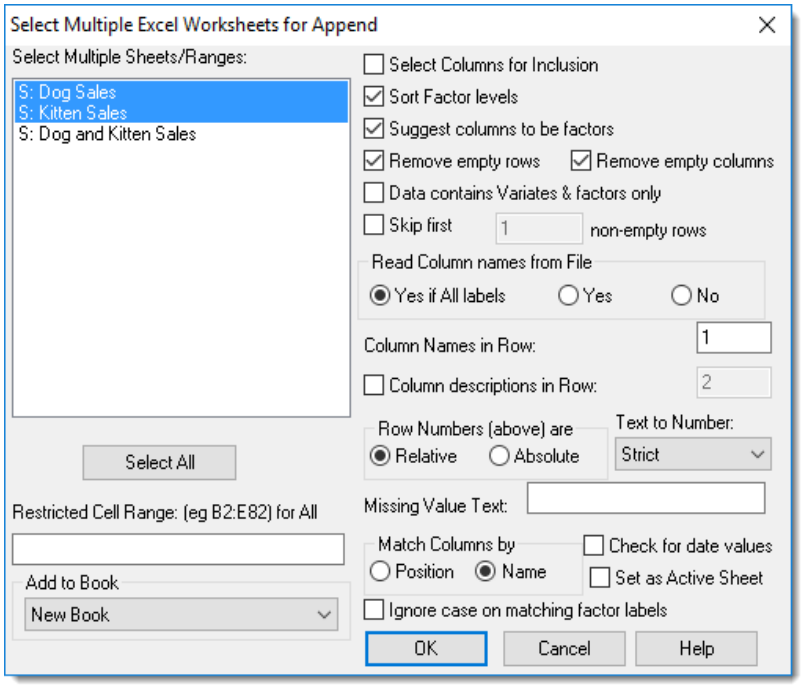


Figure 3.24

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.

3.5.2 Stacking data

We now close this sheet and use the File | Open menu to re-open `Toysales.xls` and this time import the third worksheet, Dog and Kitten Sales.

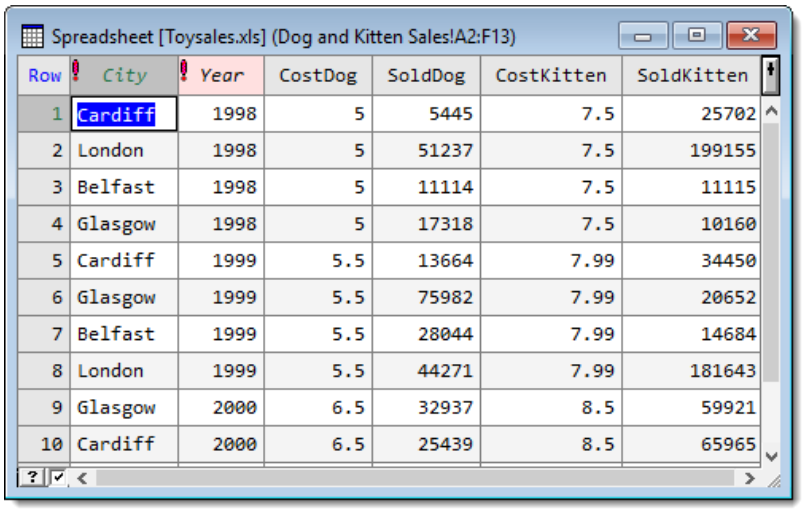
Genstat has two menus that enable you to easily stack or unstack your columns of data within a spreadsheet. We will first look at how you can stack columns together.

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

Figure 3.25

The data in the worksheet Dog and Kitten Sales are shown in Figure 3.26. There are six columns; the city (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

Figure 3.26

To stack the columns, select Spread | Manipulate | Stack, which produces the dialog 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 field. We enter the factor name `Toy` into the Record column source in Factor field; this creates a new column containing a factor where each level will represent a column that has been stacked. Click inside the Stack Columns list to give this field the focus, then hold down the Ctrl key and in Available Data select `CostDog` and `CostKitten`. Click -> to copy these to the Stack Columns list.

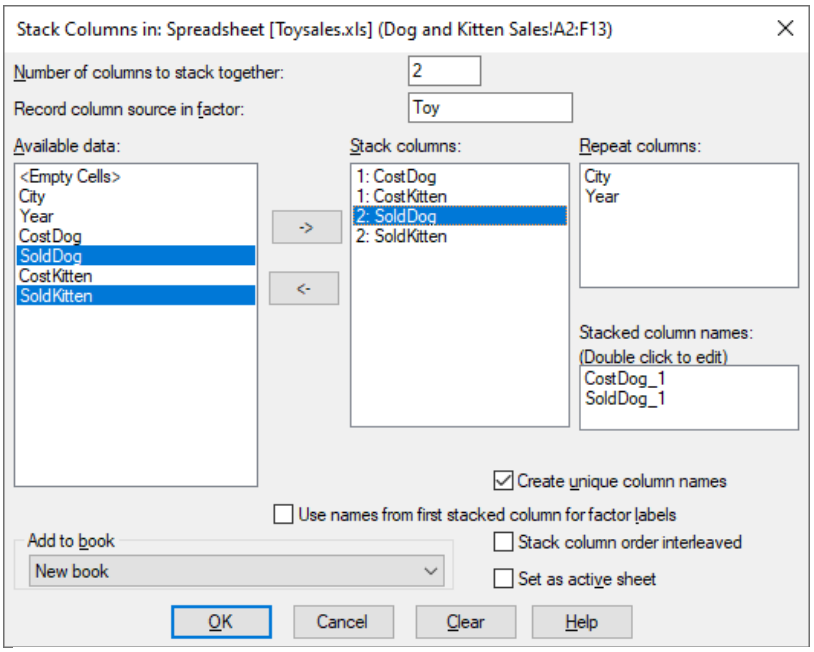


Figure 3.27

The names are prefixed with a 1, which indicates these columns will be placed in the first stacked column. Move `SoldDog` and `SoldKitten` to the Stack Columns list in the same manner. 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 field and then double-clicking their names in the Available data field.

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.

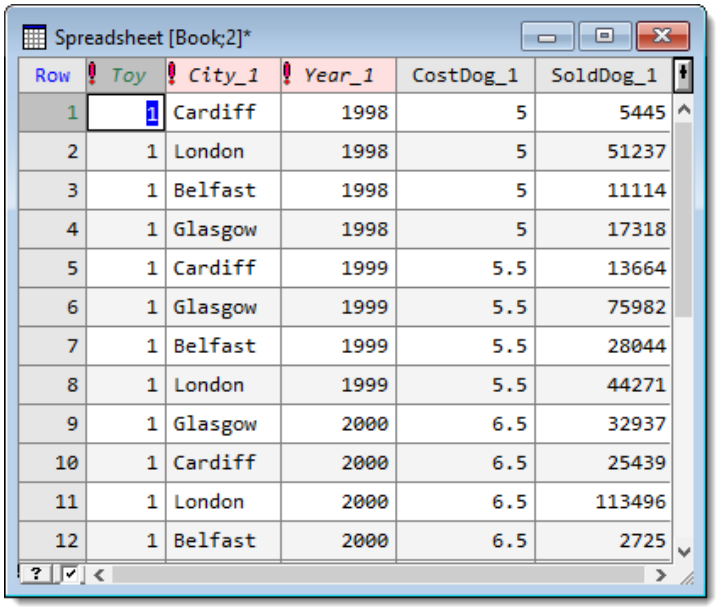


Figure 3.28

The new spreadsheet creates default names, with suffixes "_1" for the repeated and stacked columns. Note that the underscores will not be visible if the option Display spaces for underscores within column names is selected within the Tools | Spreadsheet Options, Appearance tab. If you want to rename the columns, select Spread | Column | Rename.

3.5.3 Unstacking data

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

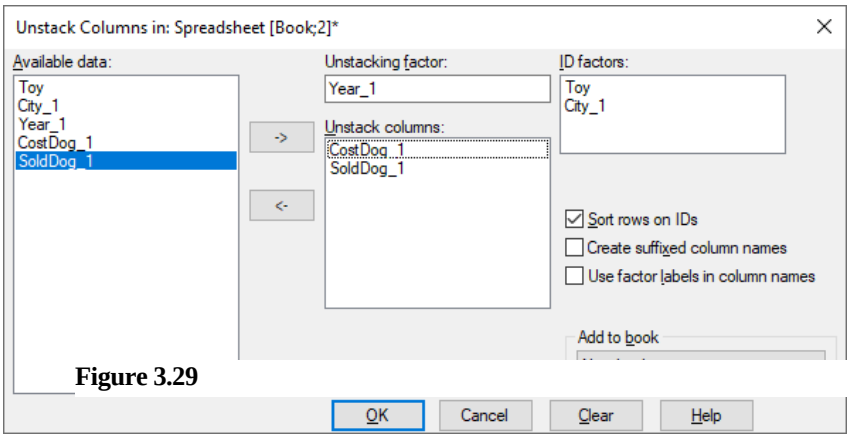
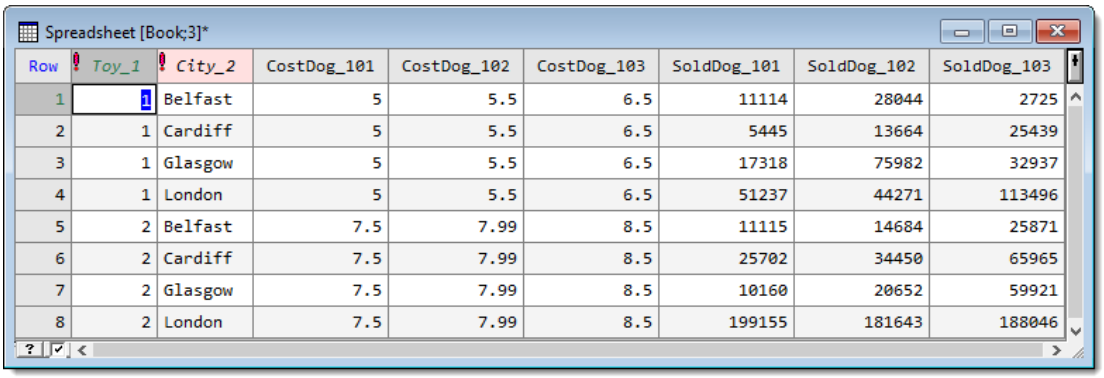


Figure 3.29

The Unstack dialog (Figure 3.29) is opened by selecting Spread | Manipulate | Unstack. This dialog 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 field. The columns will be 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  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.) Move Toy and City_1 into

the ID Factors field then click OK.

This 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 you can change using Spread | Column | Rename). Now close any open spreadsheets and clear the data from the central data pool by selecting Data | Clear All Data.


Spreadsheet [Book;3]*

Row	Toy_1	City_2	CostDog_101	CostDog_102	CostDog_103	SoldDog_101	SoldDog_102	SoldDog_103
1	1	Belfast	5	5.5	6.5	11114	28044	2725
2	1	Cardiff	5	5.5	6.5	5445	13664	25439
3	1	Glasgow	5	5.5	6.5	17318	75982	32937
4	1	London	5	5.5	6.5	51237	44271	113496
5	2	Belfast	7.5	7.99	8.5	11115	14684	25871
6	2	Cardiff	7.5	7.99	8.5	25702	34450	65965
7	2	Glasgow	7.5	7.99	8.5	10160	20652	59921
8	2	London	7.5	7.99	8.5	199155	181643	188046

Figure 3.30

3.5.4 Merging data

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, we will match together 2 sets of data where the data has been stored in different files. The files `Health1.gsh` and `Health2.gsh` contain data carried out on university students. The file `Health1.gsh` contains measurements of their height, weight, age and gender, while 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 index to merge the spreadsheets. To merge the data both files need to be open within spreadsheets in Genstat. Open them using File | Open to display the spreadsheets in Figure 3.31.

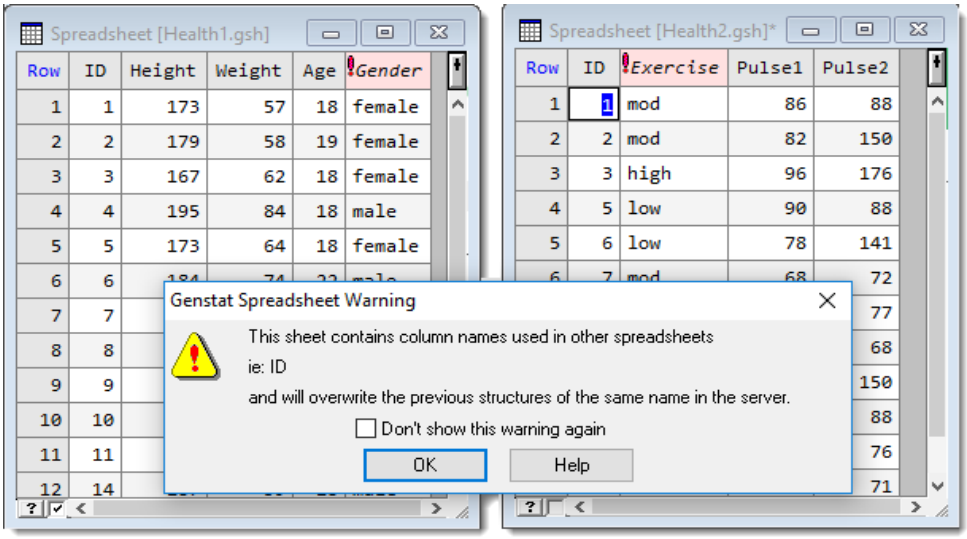


Figure 3.31

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

Clicking in the spreadsheet `Health1.gsh` and then selecting Spread | Manipulate | Merge opens the dialog shown in Figure 3.32.

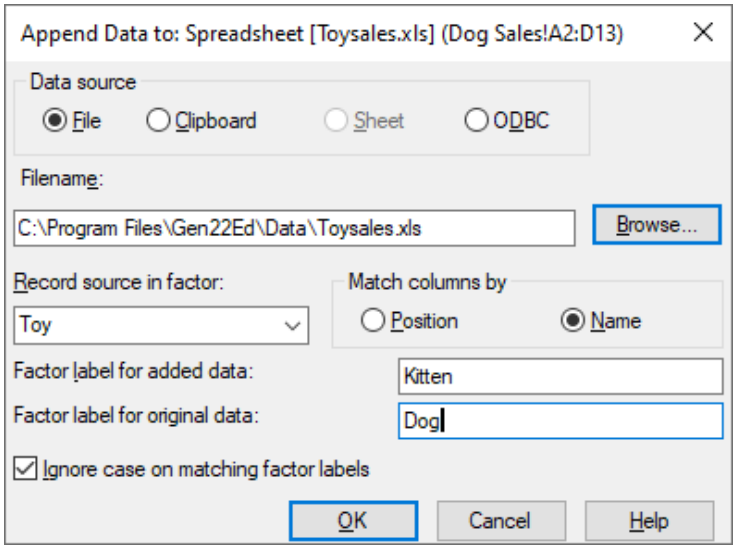


Figure 3.32

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 all settings at their defaults then clicking OK produces the merged spreadsheet shown in Figure 3.33.

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 3.33

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.

An alternative method is available for merging data from multiple files where the files can be in different file formats. Select Spread | New | Merge Multiple Files to open the dialog shown in Figure 3.34. Click on the Browse button ... then locate and select `Health1.gsh` and `Health2.gsh` and click OK.

This places the names of the two files in the Filename list. We wish to merge the spreadsheets using each student's id, so we have entered `ID` into the Identifying Columns for Merge list. Clicking Open will produce a spreadsheet identical to the spreadsheet shown in Figure 3.33.

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

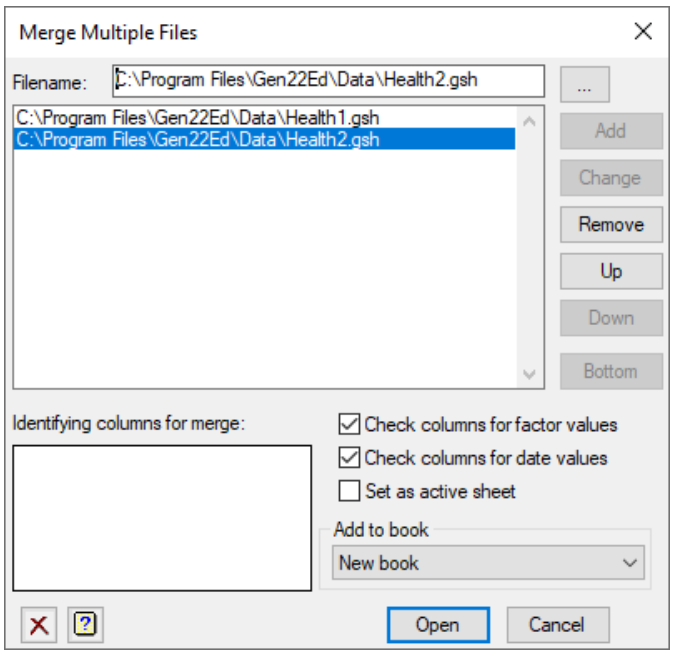


Figure 3.34

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.

Use the Graphics wizard to create a scatter plot of the mean number of florets against the temperature. To do this, from the menu select Graphics | 2D Scatter Plot and use the wizard to plot the graph. Now redraw the graph, but this time enter the source factor (created from the stack) into the Groups field to highlight the two different groups.

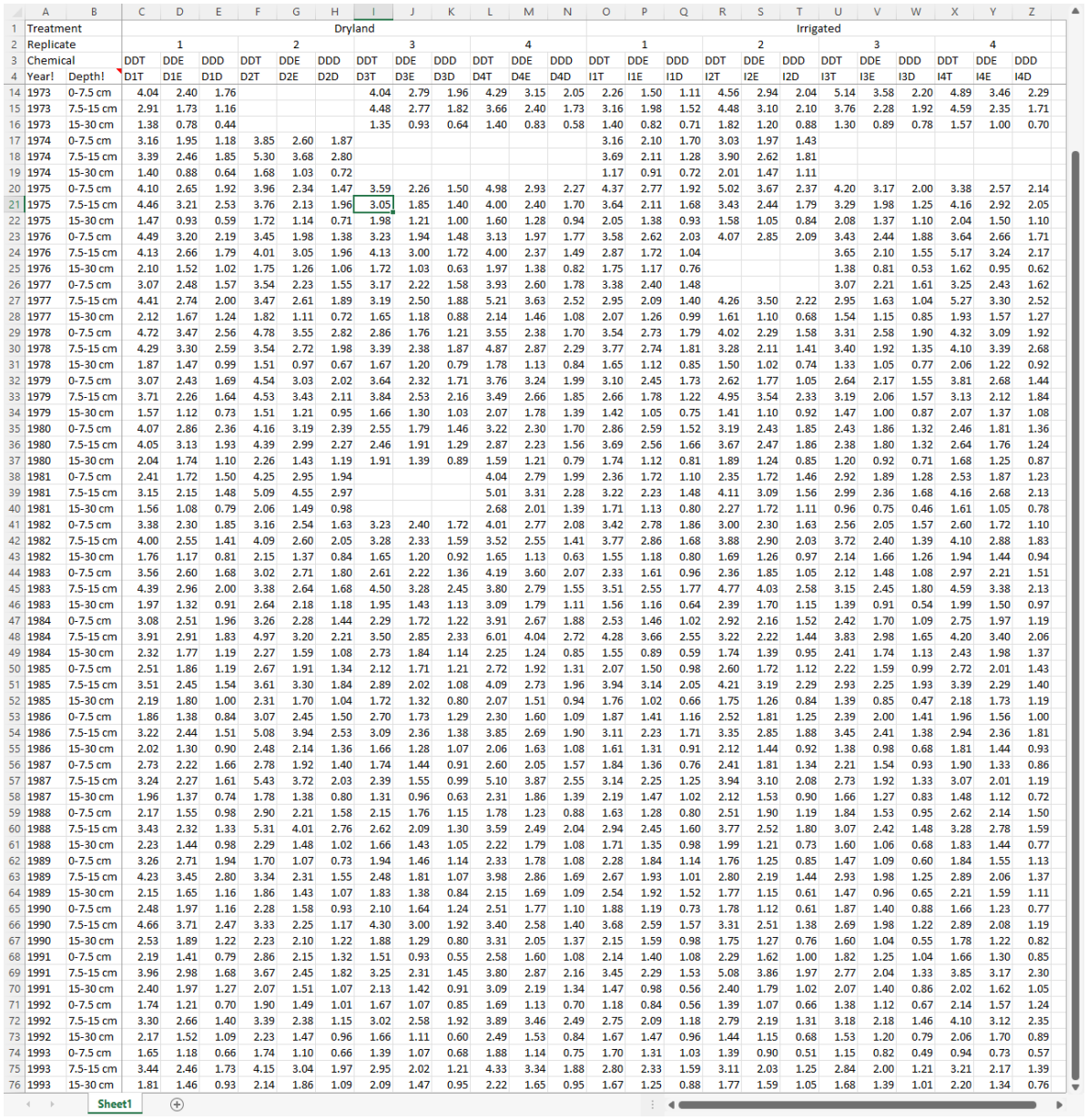
3.7 Reshaping Spreadsheets

When you have information in rows that should be in columns and vice-versa the Spread | Manipulate | Reshape menu lets you swap information between rows and columns. This may be needed when the data have been laid out in a non-standard format for visual reasons rather than for simplicity of analysis. The Excel file

`C:\Program Files\Gen22Ed\Examples\DDT Soil Samples 1970-93.xlsx`

is an example of such a file. (Note you will need to go up from the `Data` folder and into the `Examples` folder to open this file.) The content of the file is shown in Figure 3.35. The columns are the plots in the trial (two irrigation levels with 4 replicates) and the chemicals measured on each plot. The chemicals are the pesticide DDT and its two break-down products DDE and DDD. This pesticide is long-lasting in soil, and can contaminate products such as meat and milk when animals ingest the soil. The trial has measured the

levels of the chemicals in the soil over 24 years and at 3 depths. These form the rows of the dataset. The data were probably entered in this format so that a new set of rows could be added each year. To analyse the data, we need to keep the chemicals in rows, but put the 8 plots (irrigation × replicate) into columns.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Treatment		Dryland												Irrigated											
2	Replicate		1			2			3			4			1			2			3			4		
3	Chemical		DDT	DDE	DDD	DDT	DDE	DDD	DDT	DDE	DDD	DDT	DDE	DDD	DDT	DDE	DDD	DDT	DDE	DDD	DDT	DDE	DDD	DDT	DDE	DDD
4	Year!	Depth!	D1T	D1E	D1D	D2T	D2E	D2D	D3T	D3E	D3D	D4T	D4E	D4D	I1T	I1E	I1D	I2T	I2E	I2D	I3T	I3E	I3D	I4T	I4E	I4D
14	1973	0-7.5 cm	4.04	2.40	1.76				4.04	2.79	1.96	4.29	3.15	2.05	2.26	1.50	1.11	4.56	2.94	2.04	5.14	3.58	2.20	4.89	3.46	2.29
15	1973	7.5-15 cm	2.91	1.73	1.16				4.48	2.77	1.82	3.66	2.40	1.73	3.16	1.98	1.52	4.48	3.10	2.10	3.76	2.28	1.92	4.59	2.35	1.71
16	1973	15-30 cm	1.38	0.78	0.44				1.35	0.93	0.64	1.40	0.83	0.58	1.40	0.82	0.71	1.82	1.20	0.88	1.30	0.89	0.78	1.57	1.00	0.70
17	1974	0-7.5 cm	3.16	1.95	1.18	3.85	2.60	1.87							3.16	2.10	1.70	3.03	1.97	1.43						
18	1974	7.5-15 cm	3.39	2.46	1.85	5.30	3.68	2.80							3.69	2.11	1.28	3.90	2.62	1.81						
19	1974	15-30 cm	1.40	0.88	0.64	1.68	1.03	0.72							1.17	0.91	0.72	2.01	1.47	1.11						
20	1975	0-7.5 cm	4.10	2.65	1.92	3.96	2.34	1.47	3.59	2.26	1.50	4.98	2.93	2.27	4.37	2.77	1.92	5.02	3.67	2.37	4.20	3.17	2.00	3.38	2.57	2.14
21	1975	7.5-15 cm	4.46	3.21	2.53	3.76	2.13	1.96	3.05	1.85	1.40	4.00	2.40	1.70	3.64	2.11	1.68	3.43	2.44	1.79	3.29	1.98	1.25	4.16	2.92	2.05
22	1975	15-30 cm	1.47	0.93	0.59	1.72	1.14	0.71	1.98	1.21	1.00	1.60	1.28	0.94	2.05	1.38	0.93	1.58	1.05	0.84	2.08	1.37	1.10	2.04	1.50	1.10
23	1976	0-7.5 cm	4.49	3.20	2.19	3.45	1.98	1.38	3.23	1.94	1.48	3.13	1.97	1.77	3.58	2.62	2.03	4.07	2.85	2.09	3.43	2.44	1.88	3.64	2.66	1.71
24	1976	7.5-15 cm	4.13	2.66	1.79	4.01	3.05	1.96	4.13	3.00	1.72	4.00	2.37	1.49	2.87	1.72	1.04				3.65	2.10	1.55	5.17	3.24	2.17
25	1976	15-30 cm	2.10	1.52	1.02	1.75	1.26	1.06	1.72	1.03	0.63	1.97	1.38	0.82	1.75	1.17	0.76				1.38	0.81	0.53	1.62	0.95	0.62
26	1977	0-7.5 cm	3.07	2.48	1.57	3.54	2.23	1.55	3.17	2.22	1.58	3.93	2.60	1.78	3.38	2.40	1.48				3.07	2.21	1.61	3.25	2.43	1.62
27	1977	7.5-15 cm	4.41	2.74	2.00	3.47	2.61	1.89	3.19	2.50	1.88	5.21	3.63	2.52	2.95	2.09	1.40	4.26	3.50	2.22	2.95	1.63	1.04	5.27	3.30	2.52
28	1977	15-30 cm	2.12	1.67	1.24	1.82	1.11	0.72	1.65	1.18	0.88	2.14	1.46	1.08	2.07	1.26	0.99	1.61	1.10	0.68	1.54	1.15	0.85	1.93	1.57	1.27
29	1978	0-7.5 cm	4.72	3.47	2.56	4.78	3.55	2.82	2.86	1.76	1.21	3.55	2.38	1.70	3.54	2.73	1.79	4.02	2.29	1.58	3.31	2.58	1.90	4.32	3.09	1.92
30	1978	7.5-15 cm	4.29	3.30	2.59	3.54	2.72	1.98	3.39	2.38	1.87	4.87	2.87	2.29	3.77	2.74	1.81	3.28	2.11	1.41	3.40	1.92	1.35	4.10	3.39	2.68
31	1978	15-30 cm	1.87	1.47	0.99	1.51	0.97	0.67	1.67	1.20	0.79	1.78	1.13	0.84	1.65	1.12	0.85	1.50	1.02	0.74	1.33	1.05	0.77	2.06	1.22	0.92
32	1979	0-7.5 cm	3.07	2.43	1.69	4.54	3.03	2.02	3.64	2.32	1.71	3.76	3.24	1.99	3.10	2.45	1.73	2.62	1.77	1.05	2.64	2.17	1.55	3.81	2.68	1.44
33	1979	7.5-15 cm	3.71	2.26	1.64	4.53	3.43	2.11	3.84	2.53	2.16	3.49	2.66	1.85	2.66	1.78	1.22	4.95	3.54	2.33	3.19	2.06	1.57	3.13	2.12	1.84
34	1979	15-30 cm	1.57	1.12	0.73	1.51	1.21	0.95	1.66	1.30	1.03	2.07	1.78	1.39	1.42	1.05	0.75	1.41	1.10	0.92	1.47	1.00	0.87	2.07	1.37	1.08
35	1980	0-7.5 cm	4.07	2.86	2.36	4.16	3.19	2.39	2.55	1.79	1.46	3.22	2.30	1.70	2.86	2.59	1.52	3.19	2.43	1.85	2.43	1.86	1.32	2.46	1.81	1.36
36	1980	7.5-15 cm	4.05	3.13	1.93	4.39	2.99	2.27	2.46	1.91	1.29	2.87	2.23	1.56	3.69	2.56	1.66	3.67	2.47	1.86	2.38	1.80	1.32	2.64	1.76	1.24
37	1980	15-30 cm	2.04	1.74	1.10	2.26	1.43	1.19	1.91	1.39	0.89	1.59	1.21	0.79	1.74	1.12	0.81	1.89	1.24	0.85	1.20	0.92	0.71	1.68	1.25	0.87
38	1981	0-7.5 cm	2.41	1.72	1.50	4.25	2.95	1.94				4.04	2.79	1.99	2.36	1.72	1.10	2.35	1.72	1.46	2.92	1.89	1.28	2.53	1.87	1.23
39	1981	7.5-15 cm	3.15	2.15	1.48	5.09	4.55	2.97				5.01	3.31	2.28	3.22	2.23	1.48	4.11	3.09	1.56	2.99	2.36	1.68	4.16	2.68	2.13
40	1981	15-30 cm	1.56	1.08	0.79	2.06	1.49	0.98				2.68	2.01	1.39	1.71	1.13	0.80	2.27	1.72	1.11	0.96	0.75	0.46	1.61	1.05	0.78
41	1982	0-7.5 cm	3.38	2.30	1.85	3.16	2.54	1.63	3.23	2.40	1.72	4.01	2.77	2.08	3.42	2.78	1.86	3.00	2.30	1.63	2.56	2.05	1.57	2.60	1.72	1.10
42	1982	7.5-15 cm	4.00	2.55	1.41	4.09	2.60	2.05	3.28	2.33	1.59	3.52	2.55	1.41	3.77	2.86	1.68	3.88	2.90	2.03	3.72	2.40	1.39	4.10	2.88	1.83
43	1982	15-30 cm	1.76	1.17	0.81	2.15	1.37	0.84	1.65	1.20	0.92	1.65	1.13	0.63	1.55	1.18	0.80	1.69	1.26	0.97	2.14	1.66	1.26	1.94	1.44	0.94
44	1983	0-7.5 cm	3.56	2.60	1.68	3.02	2.71	1.80	2.61	2.22	1.36	4.19	3.60	2.07	2.33	1.61	0.96	2.36	1.85	1.05	2.12	1.48	1.08	2.97	2.21	1.51
45	1983	7.5-15 cm	4.39	2.96	2.00	3.38	2.64	1.68	4.50	3.28	2.45	3.80	2.79	1.55	3.51	2.55	1.77	4.77	4.03	2.58	3.15	2.45	1.80	4.59	3.38	2.13
46	1983	15-30 cm	1.97	1.32	0.91	2.64	2.18	1.18	1.95	1.43	1.13	3.09	1.79	1.11	1.56	1.16	0.64	2.39	1.70	1.15	1.39	0.91	0.54	1.99	1.50	0.97
47	1984	0-7.5 cm	3.08	2.51	1.96	3.26	2.28	1.44	2.29	1.72	1.22	3.91	2.67	1.88	2.53	1.46	1.02	2.92	2.16	1.52	2.42	1.70	1.09	2.75	1.97	1.19
48	1984	7.5-15 cm	3.91	2.91	1.83	4.97	3.20	2.21	3.50	2.85	2.33	6.01	4.04	2.72	4.28	3.66	2.55	3.22	2.22	1.44	3.83	2.98	1.65	4.20	3.40	2.06
49	1984	15-30 cm	2.32	1.77	1.19	2.27	1.59	1.08	2.73	1.84	1.14	2.25	1.24	0.85	1.55	0.89	0.59	1.74	1.39	0.95	2.41	1.74	1.13	2.43	1.98	1.37
50	1985	0-7.5 cm	2.51	1.86	1.19	2.67	1.91	1.34	2.12	1.71	1.21	2.72	1.92	1.31	2.07	1.50	0.98	2.60	1.72	1.12	2.22	1.59	0.99	2.72	2.01	1.43
51	1985	7.5-15 cm	3.51	2.45	1.54	3.61	3.30	1.84	2.89	2.02	1.08	4.09	2.73	1.96	3.94	3.14	2.05	4.21	3.19	2.29	2.93	2.25	1.93	3.39	2.29	1.40
52	1985	15-30 cm	2.19	1.80	1.00	2.31	1.70	1.04	1.72	1.32	0.80	2.07	1.51	0.94	1.76	1.02	0.66	1.75	1.26	0.84	1.39	0.85	0.47	2.18	1.73	1.19
53	1986	0-7.5 cm	1.86	1.38	0.84	3.07	2.45	1.50	2.70	1.73	1.29	2.30	1.60	1.09	1.87	1.41	1.16	2.52	1.81	1.25	2.39	2.00	1.41	1.96	1.56	1.00
54	1986	7.5-15 cm	3.22	2.44	1.51	5.08	3.94	2.53	3.09	2.36	1.38	3.85	2.69	1.90	3.11	2.23	1.71	3.35	2.85	1.88	3.45	2.41	1.38	2.94	2.36	1.81
55	1986	15-30 cm	2.02	1.30	0.90	2.48	2.14	1.36	1.66	1.28	1.07	2.06	1.63	1.08	1.61	1.31	0.91	2.12	1.44	0.92	1.38	0.98	0.68	1.81	1.44	0.93
56	1987	0-7.5 cm	2.73	2.22	1.66	2.78	1.92	1.40	1.74	1.44	0.91	2.60	2.05	1.57	1.84	1.36	0.76	2.41	1.81	1.34	2.21	1.54	0.93	1.90	1.33	0.86
57	1987	7.5-15 cm	3.24	2.27	1.61	5.43	3.72	2.03	2.39	1.55	0.99	5.10	3.87	2.55	3.14	2.25	1.25	3.94	3.10	2.08	2.73	1.92	1.33	3.07	2.01	1.19
58	1987	15-30 cm	1.96	1.37	0.74	1.78	1.38	0.80	1.31	0.96	0.63	2.31	1.86	1.39	2.19	1.47	1.02	2.12	1.53	0.90	1.66	1.27	0.83	1.48	1.12	0.72
59	1988	0-7.5 cm	2.17	1.55	0.98	2.90	2.21	1.58	2.15	1.76	1.15	1.78	1.23	0.88	1.63	1.28	0.80	2.51	1.90	1.19	1.84	1.53	0.95	2.62	2.14	1.50
60	1988	7.5-15 cm	3.43	2.32	1.33	5.31	4.01	2.76	2.62	2.09	1.30	3.59	2.49	2.04	2.94	2.45	1.60	3.77	2.52	1.80	3.07	2.42	1.48	3.28	2.78	1.59
61	1988	15-30 cm	2.23	1.44	0.98	2.29	1.48	1.02	1.66	1.43	1.05	2.22	1.79	1.08	1.71	1.35	0.98	1.99	1.21	0.73	1.60	1.06	0.68	1.83	1.44	0.77
62	1989	0-7.5 cm	3.26	2.71	1.94	1.70	1.07	0.73	1.94	1.46	1.14	2.33	1.78	1.08	2.28	1.84	1.14	1.76	1.25	0.85	1.47	1.09	0.60	1.84	1.55	1.13
63	1989	7.5-15 cm	4.23	3.45	2.80	3.34	2.31	1.55	2.48	1.81	1.07	3.98	2.86	1.69	2.67	1.93	1.01	2.80	2.19	1.44	2.93	1.98	1.25	2.89	2.06	1.37
64	1989	15-30 cm	2.15	1.65	1.16	1.86	1.43	1.07	1.83	1.38	0.84	2.15	1.69	1.09	2.54	1.92	1.52	1.77	1.15	0.61	1.47	0.96	0.65	2.21	1.59	1.11
65	1990	0-7.5 cm	2.48	1.97	1.16	2.28	1.58	0.93	2.10	1.64	1.24	2.51	1.77	1.10	1.88	1.19	0.73	1.78	1.12	0.61	1.87	1.40	0.88	1.66	1.23	0.77
66	1990	7.5-15 cm	4.66	3.71	2.47	3.33	2.25	1.17	4.30	3.00	1.92	3.40	2.58	1.40	3.68	2.59	1.57	3.31	2.51	1.38	2.69	1.98	1.22	2.89	2.08	1.19
67	1990	15-30 cm	2.53	1.89	1.22	2.23	2.10	1.22	1.88	1.29	0.80	3.31	2.05	1.37	2.15	1.59	0.98	1.75	1.27	0.76	1.60	1.04	0.55	1.78	1.22	0.82
68	1991	0-7.5 cm	2.19	1.41	0.79	2.86	2.15	1.32	1.51	0.93	0.55	2.58	1.60	1.08	2.14	1.40	1.08	2.29	1.62	1.00	1.82	1.25	1.04	1.66	1.30	0.85
69	1991	7.5-15 cm	3.96	2.98	1.68	3.67	2.45	1.82	3.25	2.31	1.45	3.80	2.87	2.16	3.45	2.29	1.53	5.08	3.86	1.97	2.77	2.04	1.33	3.85	3.17	2.30
70	1991	15-30 cm	2.40	1.97	1.27	2.07	1.51	1.07	2.13	1.42	0.91	3.09	2.19	1.34	1.47	0.98	0.56	2.40	1.79	1.02	2.07	1.40	0.86	2.02	1.62	1.05
71	1992	0-7.5 cm	1.74	1.21	0.70	1.90	1.49	1.01	1.67	1.07	0.85	1.69	1.13	0.70	1.18	0.84	0.56	1.39	1.07	0.66	1.38	1.12	0.67	2.14	1.57	1.24
72	1992	7.5-15 cm	3.30	2.66	1.40	3.39	2.38	1.15	3.02	2.58	1.92	3.89	3.46	2.49	2.75	2.09	1.18	2.79	2.19	1.31	3.18	2.18	1.46	4.10	3.12	2.35
73	1992	15-30 cm	2.17	1.52	1.09	2.23	1.47	0.96	1.66	1.11	0.60	2.49	1.53	0.84	1.67	1.47	0.96	1.44	1.15	0.68	1.53	1.20	0.79	2.06	1.70	0.89
74	1993	0-7.5 cm	1.65	1.18	0.66	1.74	1.10	0.66	1.39	1.07	0.68	1.88	1.14	0.75	1.70	1.31	1.03	1.39	0.90	0.51	1.15	0.82	0.49	0.94	0.73	0.57
75	1993	7.5-15 cm	3.44	2.46	1.73	4.15	3.04	1.97	2.95	2.02	1.21	4.33	3.34	1.88	2.80	2.33	1.59	3.11	2.03	1.25	2.84	2.00	1.21	3.21	2.17	1.39
76	1993	15-30 cm	1.81	1.46	0.93	2.14	1.86	1.09	2.09	1.47	0.95	2.22	1.65	0.95	1.67	1.25	0.88	1.77	1.59	1.05	1.68	1.39	1.01	2.20	1.34	0.76

Sheet1

Figure 3.35

To open the data into a spreadsheet, use the File | Open menu, and select the file above. Then click Next in the Excel wizard, and set the cell range to A4 to avoid the first three rows of column information as show in Figure 3.36. Then click Finish.

Next open the Spread | Manipulate | Reshape menu, and fill in the menu as shown in Figure 3.37. All 24 columns `D1T…I4D` are to be set as the Data columns. `Year` and `Depth` are the row factors. The column factors need to be created, so click the Create column factors button to give the spreadsheet shown in Figure 3.38.

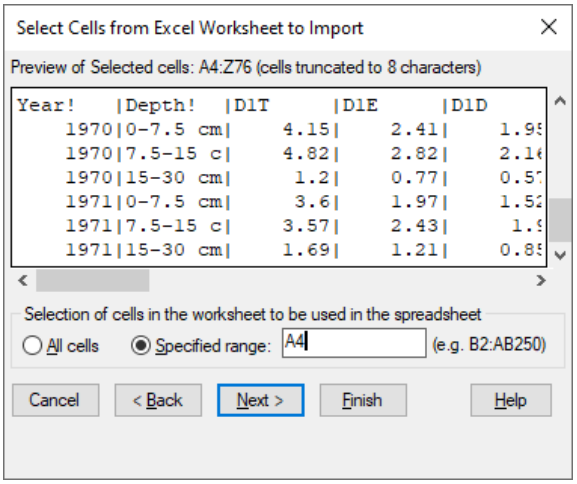


Figure 3.36

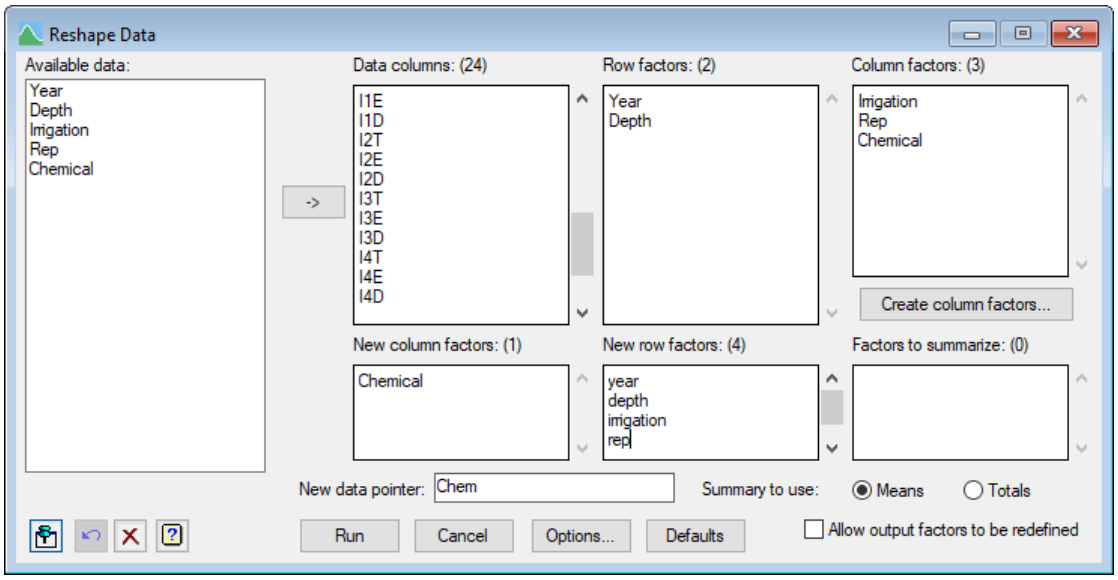


Figure 3.37

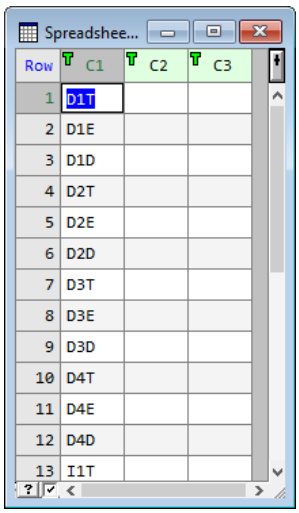


Figure 3.38

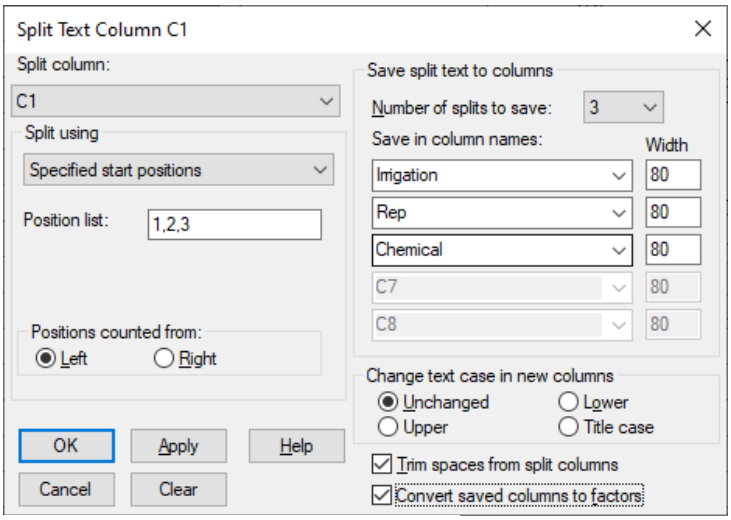
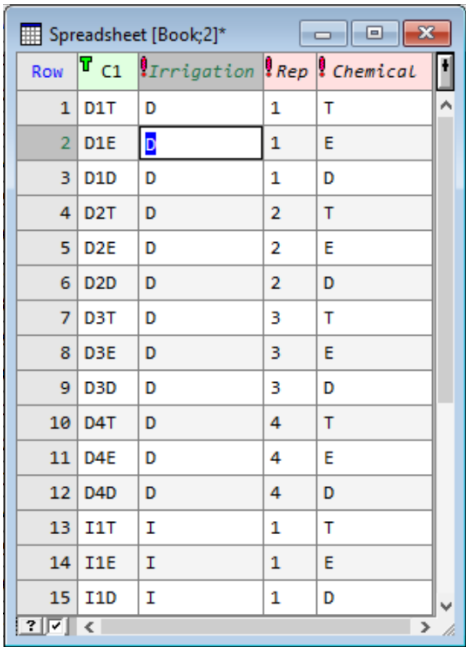


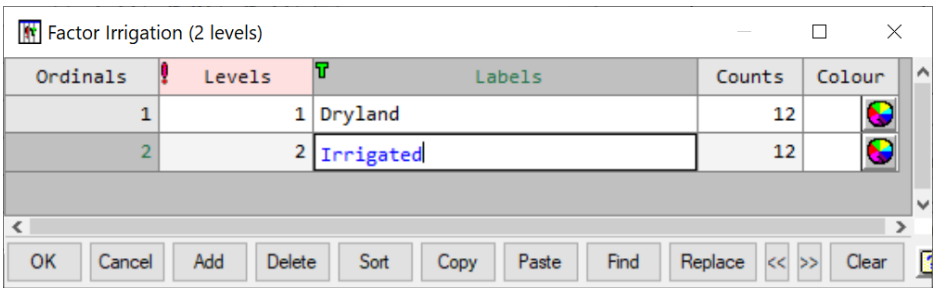
Figure 3.39

The information on the irrigation, replicate and chemical is encoded in the name. So we can use the Spread | Calculate | Text Split to split this into 3 factors for `Irrigation`, `Rep` and `Chemical` as shown in Figure 3.39 on the previous page. This forms the spreadsheet in Figure 3.40 (after deleting the unneeded columns `C2` and `C3)`. We can now use the Spread | Factor | Edit levels and labels menu, as shown in Figure 3.41, to expand the one letter codes to the longer descriptions `Dryland`, `Irrigated` for `Irrigation` and `DDD`, `DDE`, `DDT` for `Chemical`. Note that rather than opening the Edit levels and labels menu separately for each column, you can click the >> button twice to go from editing the `Irrigation` column to editing the `Chemical` column. This gives the spreadsheet shown in Figure 3.42 on the next page. Then, back in the Reshape menu (see Figure 3.37), you can put `Irrigation`, `Rep` and `Chemical` as the Column factors. The New column factors entry is `Chemical`, as the 3 chemicals are to be the resulting data columns. You now need to type in names for the new row columns holding the `Year`, `Depth`, `Irrigation` and `Rep` information. Type these into the New row factors list. There should be one column name for each row or column factor that is not in the New column factors list, and these must be in the order specified firstly in Row factors and then in Column factors. Here they have been entered in lower case, so that they do not conflict with the original column names. Finally, put the name `Chem` for the 3 data columns in the New data pointer field, and click Run. The resulting spreadsheet is shown in Figure 3.43 on the next page.



Row	C1	Irrigation	Rep	Chemical
1	D1T	D	1	T
2	D1E	D	1	E
3	D1D	D	1	D
4	D2T	D	2	T
5	D2E	D	2	E
6	D2D	D	2	D
7	D3T	D	3	T
8	D3E	D	3	E
9	D3D	D	3	D
10	D4T	D	4	T
11	D4E	D	4	E
12	D4D	D	4	D
13	I1T	I	1	T
14	I1E	I	1	E
15	I1D	I	1	D

Figure 3.40



Ordinals	Levels	Labels	Counts	Colour
1	1	Dryland	12	
2	2	Irrigated	12	

OK Cancel Add Delete Sort Copy Paste Find Replace << >> Clear

Figure 3.41

Spreadsheet [Book;2]*

Row	C1	Irrigation	Rep	Chemical
1	D1T	Dryland	1	DDT
2	D1E	Dryland	1	DDE
3	D1D	Dryland	1	DDD
4	D2T	Dryland	2	DDT
5	D2E	Dryland	2	DDE
6	D2D	Dryland	2	DDD
7	D3T	Dryland	3	DDT
8	D3E	Dryland	3	DDE
9	D3D	Dryland	3	DDD
10	D4T	Dryland	4	DDT
11	D4E	Dryland	4	DDE
12	D4D	Dryland	4	DDD
13	I1T	Irrigated	1	DDT
14	I1E	Irrigated	1	DDE
15	I1D	Irrigated	1	DDD

Figure 3.42

Spreadsheet [Loaded Data;1]*

Row	year	depth	irrigation	rep	Chem['DDD']	Chem['DDE']	Chem['DDT']
1	1970	0-7.5 cm	Dryland	1	1.95	2.41	4.15
2	1970	7.5-15 cm	Dryland	1	2.16	2.82	4.82
3	1970	15-30 cm	Dryland	1	0.57	0.77	1.20
4	1971	0-7.5 cm	Dryland	1	1.52	1.97	3.60
5	1971	7.5-15 cm	Dryland	1	1.90	2.43	3.57
6	1971	15-30 cm	Dryland	1	0.85	1.21	1.69
7	1972	0-7.5 cm	Dryland	1	1.53	2.18	3.63
8	1972	7.5-15 cm	Dryland	1	1.03	1.54	3.15
9	1972	15-30 cm	Dryland	1	0.65	0.84	1.37
10	1973	0-7.5 cm	Dryland	1	1.76	2.40	4.04
11	1973	7.5-15 cm	Dryland	1	1.16	1.73	2.91
12	1973	15-30 cm	Dryland	1	0.44	0.78	1.38
13	1974	0-7.5 cm	Dryland	1	1.18	1.95	3.16
14	1974	7.5-15 cm	Dryland	1	1.85	2.46	3.39
15	1974	15-30 cm	Dryland	1	0.64	0.88	1.40
16	1975	0-7.5 cm	Dryland	1	1.92	2.65	4.10
17	1975	7.5-15 cm	Dryland	1	2.53	3.21	4.46
18	1975	15-30 cm	Dryland	1	0.59	0.93	1.47
19	1976	0-7.5 cm	Dryland	1	2.19	3.20	4.49

Figure 3.43

4 Calculations and summaries

Once you have put your observed data into 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

Open the file `Grazing.gsh` then from the menu select Spread | Calculate | Column. The Calculate dialog, shown in Figure 4.1, allows you to enter numerical expressions that that will calculate a new column using data from the other columns. We'll use this to 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. The new column is added to the end of the spreadsheet, with the background column name in yellow.

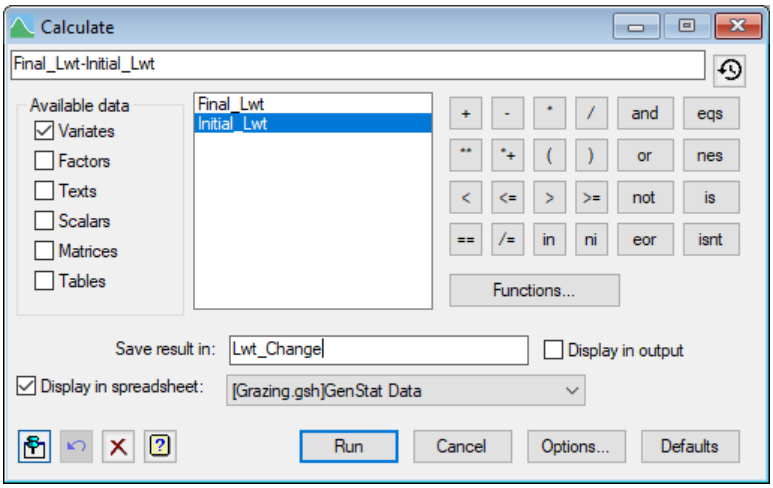


Figure 4.1

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

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).

Spreadsheet [Grazing.gsh]

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
5	1	1	TE	144.5	173.5	29
6	1	1	TE	121.5	153	31.5
7	1	1	TE	126	152	26
8	1	1	TE	122	155.5	33.5

Figure 4.2

You can edit the calculation in the Column Attributes dialog, 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:

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

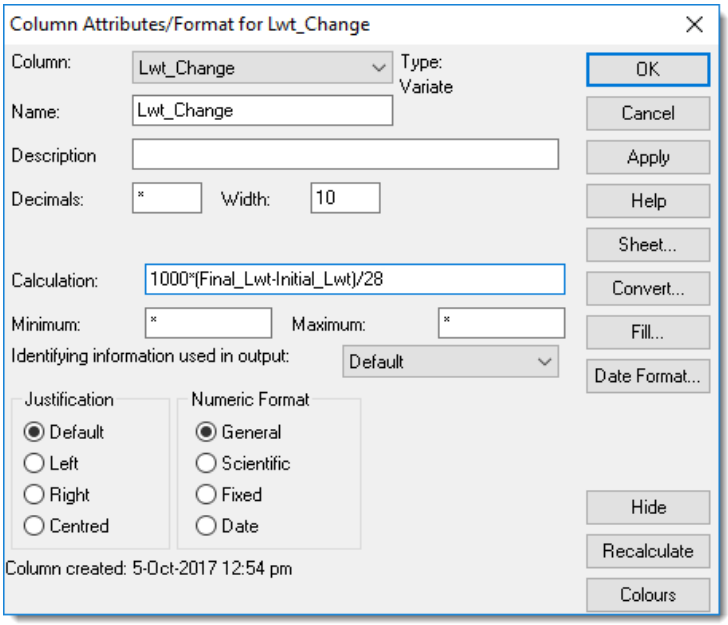


Figure 4.3

If you change the expression, then when you click OK, you will get a prompt to update the columns values as shown in Figure 4.4.

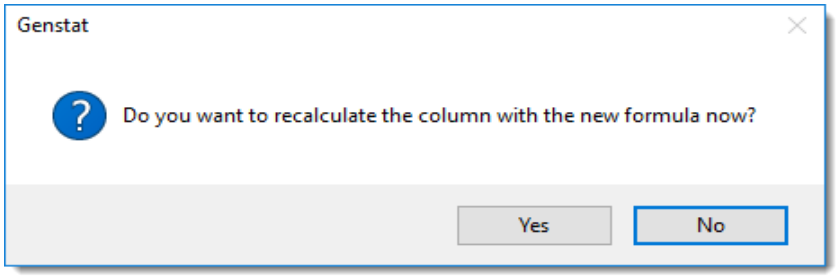


Figure 4.4

Also, if you edit any values in columns used in the expression then update the server (for example, by clicking outside the spreadsheet), you will be prompted to update the calculated column's values with the dialog shown in Figure 4.5.

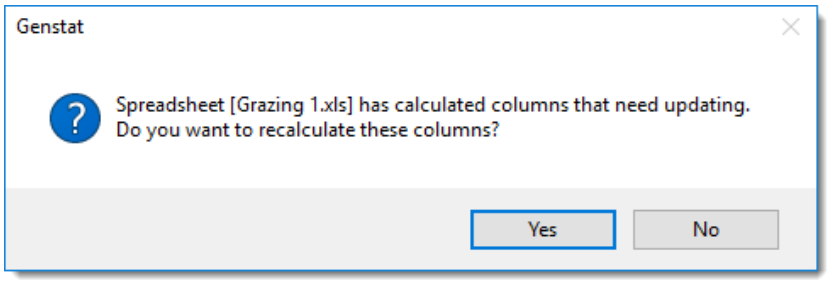


Figure 4.5

You can turn off this prompt if desired by selecting Tools | Spreadsheet Options and clicking the General tab. The option Prompt to update calculated columns controls whether the prompt appears (deselect this option to turn off the prompt).

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 you press Enter or move outside the cell the calculation will be evaluated and the numerical result put into the cell (as in Figure 4.7).

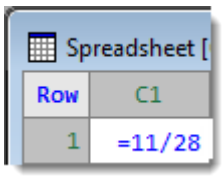


Figure 4.6

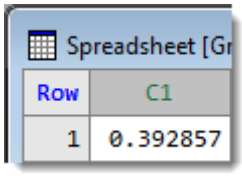
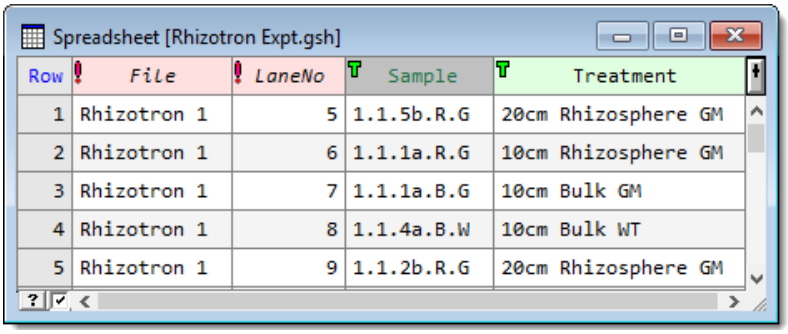


Figure 4.7

4.2 Creating text columns

There are two menus that allow you to create new text columns from other columns. Spread | Calculate | Text Split takes sections of text from an existing column and splits it into new columns. Spread | Calculate | Combine Text combines text from two or more 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, 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).

Open the spreadsheet file `Rhizotron Expt.gsh`. This file 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.



Row	File	LaneNo	Sample	Treatment
1	Rhizotron 1	5	1.1.5b.R.G	20cm Rhizosphere GM
2	Rhizotron 1	6	1.1.1a.R.G	10cm Rhizosphere GM
3	Rhizotron 1	7	1.1.1a.B.G	10cm Bulk GM
4	Rhizotron 1	8	1.1.4a.B.W	10cm Bulk WT
5	Rhizotron 1	9	1.1.2b.R.G	20cm Rhizosphere GM

Figure 4.8

Click anywhere within the `Treatment` column. This column contains text giving the depth, source and plant type (GM = genetically modified, and WT = wild type). We're going to split this column into 3 parts. From the menu select Spread | Calculate | Text Split. The sections of text are separated by spaces, so we use this as the option for Split using. We specify that the Number of splits to save is 3, 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.

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

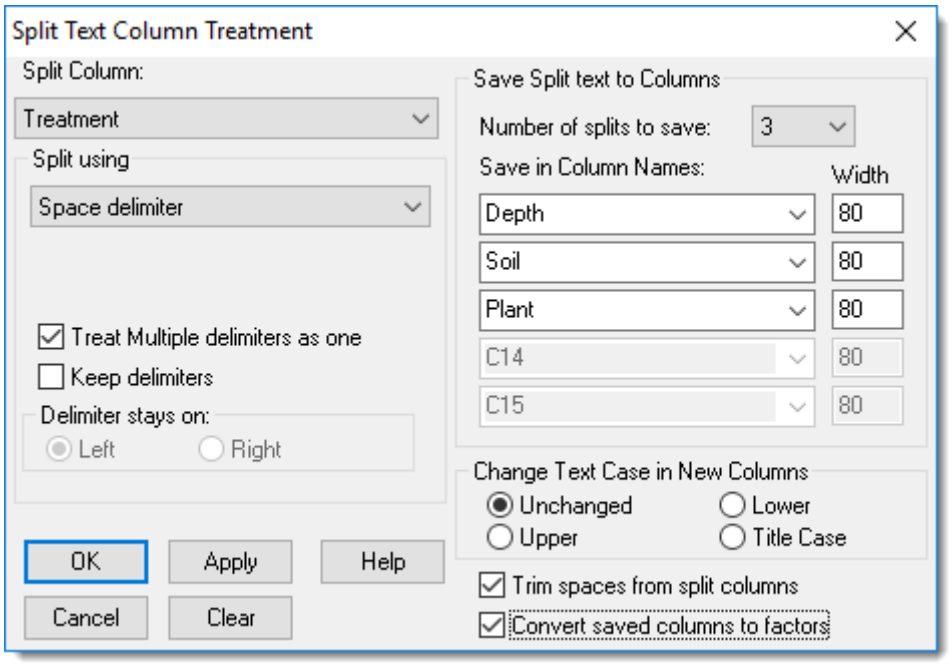
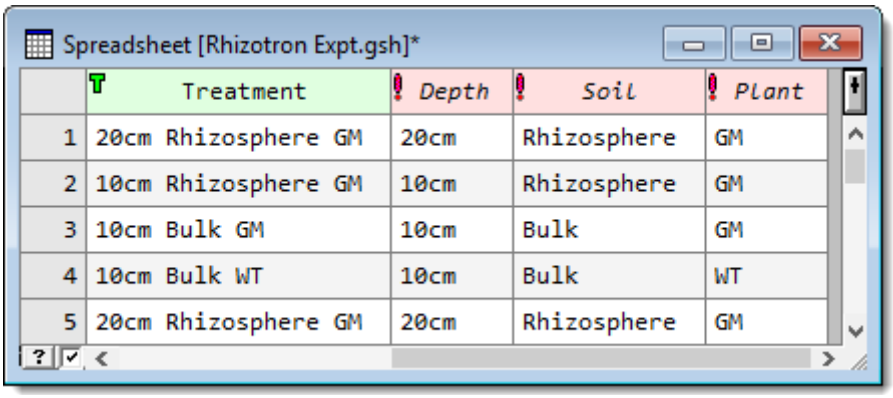


Figure 4.9



	Treatment	Depth	Soil	Plant
1	20cm Rhizosphere GM	20cm	Rhizosphere	GM
2	10cm Rhizosphere GM	10cm	Rhizosphere	GM
3	10cm Bulk GM	10cm	Bulk	GM
4	10cm Bulk WT	10cm	Bulk	WT
5	20cm Rhizosphere GM	20cm	Rhizosphere	GM

Figure 4.10

The Combine Text menu (Figure 4.11) can be used to combine several 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. Select Spread | Calculate | Combine Text. We set the Number of columns to combine as 2 and select the two columns `File` and `LaneNo`.The column `File` contains the text 'Rhizotron', which we don't want to appear in our new ID column, so we'll remove this. Set the Start position as 11 for the `File` column; this will remove the first 10 characters ('Rhizotron' plus the leading space). 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 only one character there. We will get the same result if we use the start as position 10 with length 2, because 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 OK creates the new column `ID` in the spreadsheet as shown in Figure 4.12.

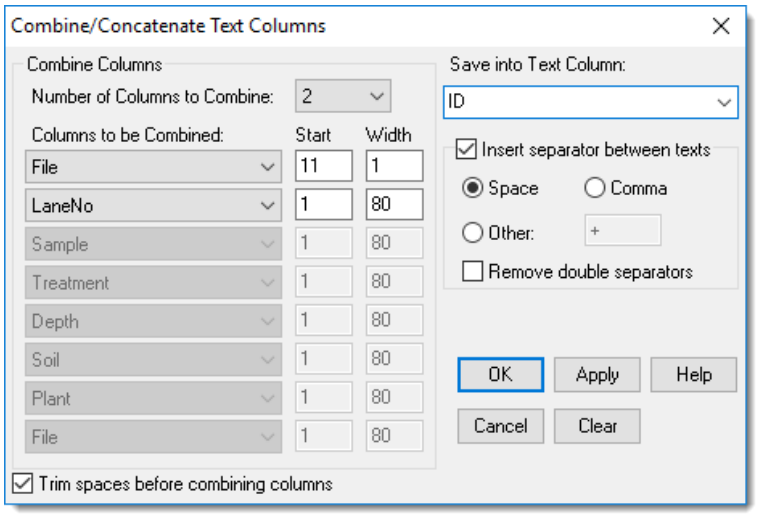


Figure 4.11

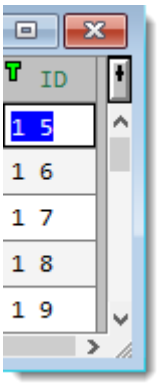
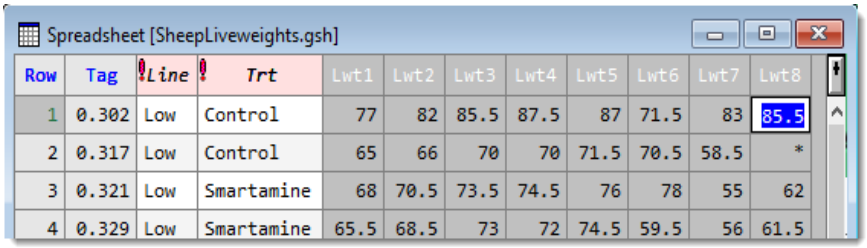


Figure 4.12

4.3 Summaries across rows

The Row Summaries 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`. 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



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
4	0.329	Low	Smartamine	65.5	68.5	73	72	74.5	59.5	56	61.5

Figure 4.13

Figure 4.13).

Now select Spread | Calculate | Row summaries menu (Figure 4.14. Choose the Row summary to calculate as Mean/Average and enter the Save In column name as `MeanLwt`.

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.

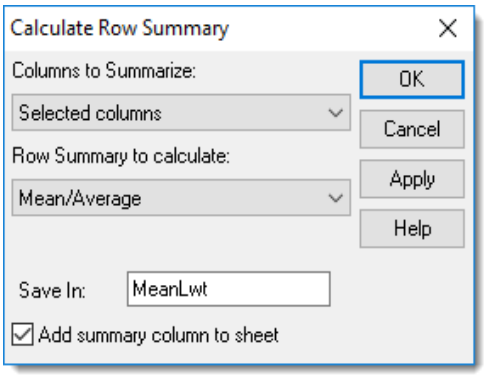
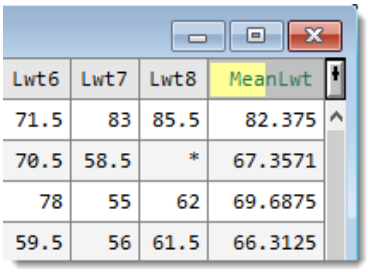


Figure 4.14



Lwt6	Lwt7	Lwt8	MeanLwt
71.5	83	85.5	82.375
70.5	58.5	*	67.3571
78	55	62	69.6875
59.5	56	61.5	66.3125

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.

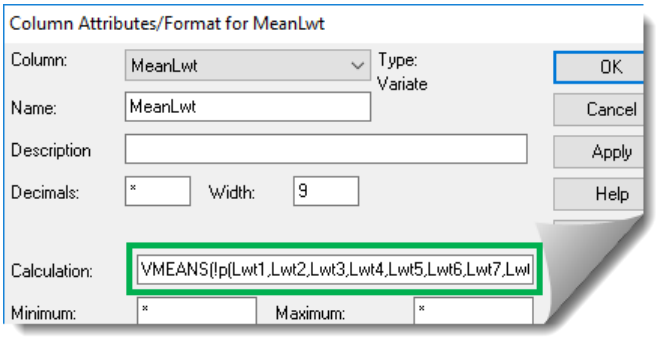


Figure 4.16

You can create a permanent pointer to the columns Lwt1-8 using the Pointers menu. Select the 8 columns as previously and then select Spread | Sheet | Pointers to open the dialog show in Figure 4.17.

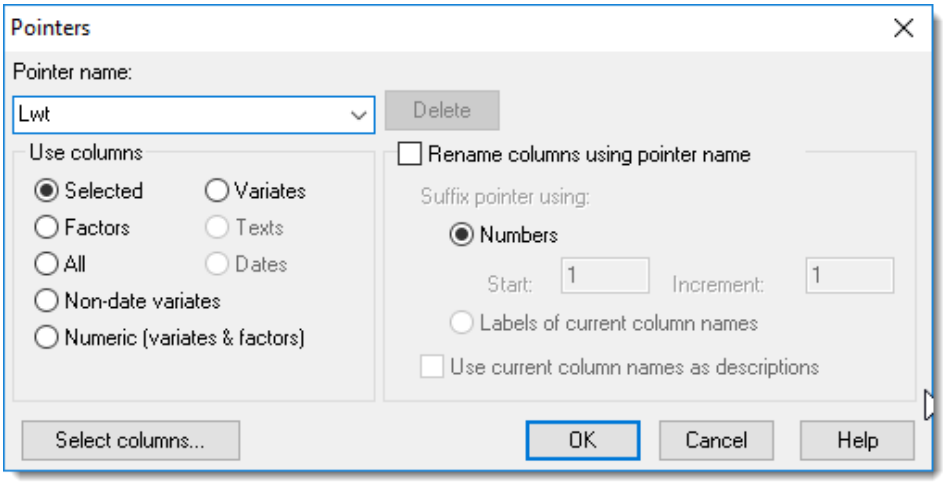


Figure 4.17

You need to give a name to the pointer (here it's called Lwt. Now if you select Spread | Calculate | Row Summaries the new pointer will appear in the dropdown list as shown in Figure 4.18.

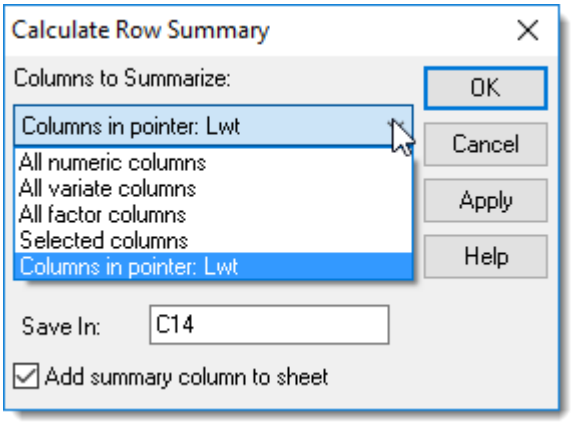
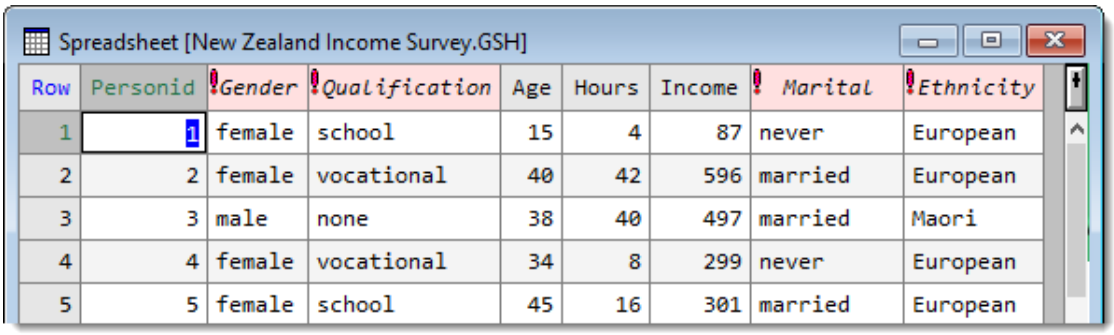


Figure 4.18

4.4 Forming Factors

If you want to form groups from a variate by allocating its values to different ranges, you can do this using the Code to Groups submenu. Open the spreadsheet New Zealand Income Survey.gsh, which contains 200 randomly selected cases from a survey on personal income for New Zealanders; see Figure 4.19.



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

We want to categorize the variate Age into 3 groups for people aged < 18, 18-40 and 40+. Place the cursor into a cell in the Age column and select Spread | Calculate | Code to Groups. This opens the dialog in Figure 4.20. Change the Groups Column Name to Age_Group and change the Number of Groups to 3.

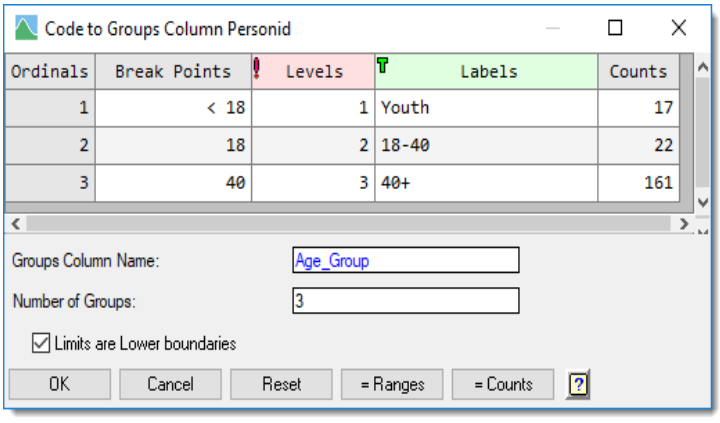


Figure 4.20

In the Break Points column ignore the first cell and enter the values 18 and 40 in cells 2 and 3. If you do not edit the labels, they will change to reflect the break points you have entered. In the Labels column, change the label to Youth for the <18 group and 40+ for the >=40 group as shown in Figure 4.20.

If you untick the item Links are Lower boundaries you put in the upper bound of each group, rather than the lower bound, starting from cell 1. Clicking OK produces creates the Age_Group column in Figure 4.21.

If you want to re-code individual items in a factor, variate or text to a new value use the Recode option. This lists all the unique items in your selected column and then for each item you can specify a new value. 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.

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

Figure 4.21

Place the cursor in the factor column Qualification and select Spread | Calculate | Recode to open the dialog in Figure 4.22. We'll recode the factor column Qualification by combining the 4 levels none, school, vocational and degree into just two levels - none/school and post school. From the menu select *Spread | Calculate | Recode*. Enter new values as shown in Figure 4.22 and enter the Recoded Column Name as Training. If Recode to Numeric is ticked the new values will be numeric. If Code as a Factor 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.

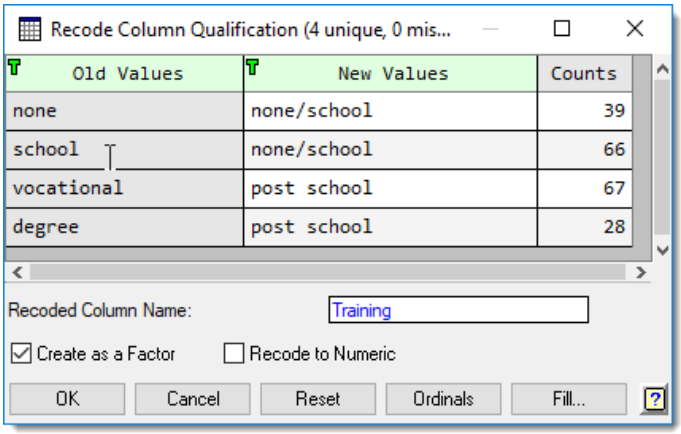


Figure 4.22

Figure 4.23 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

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

Figure 4.23

4.23. If we want to create a variate from the `Qualification` column, we can use the Recode menu again. Put the cursor in the `Qualification` column then select Spread | Calculate | Recode. Untick Create as Factor and tick Recode to Numeric. Enter new values and the column name `Qual_Score` as shown in Figure 4.24. Click OK to add the spreadsheet column shown in Figure 4.25

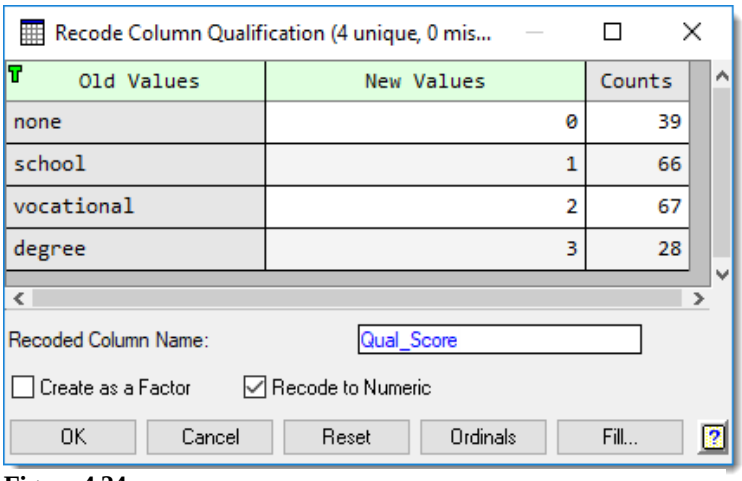
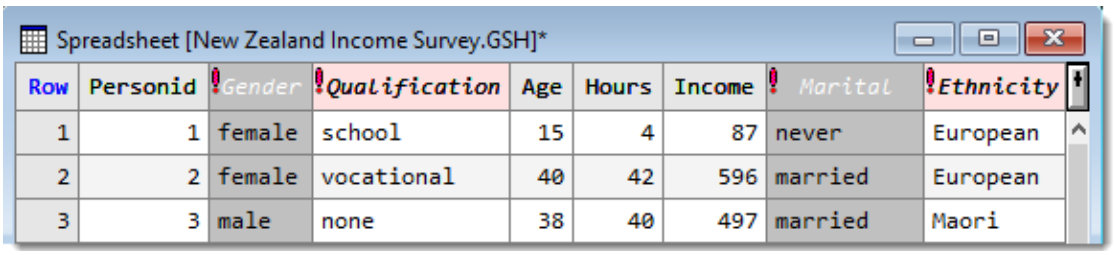


Figure 4.24

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

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 can use the Product/Combine 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. To create a factor that indexes both `Gender` and `Marital` status, select these as in Figure 4.26 then Spread | Factor | Product/Combine menu to get the dialog shown in Figure 4.27.



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

Figure 4.26

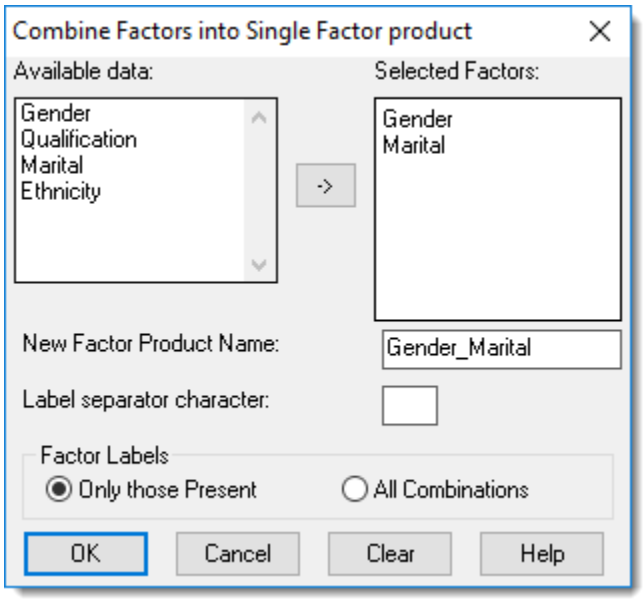


Figure 4.27

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

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.

The opposite action to combining factors is to divide a factor. Genstat 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, put the cursor in this column then select Spread | Factor | Divide to open the dialog shown in Figure 4.29. We enter the names for the two new factors as `Sex` and `Marital_Status`. The first factor `Sex` has 2 levels and shows we have to change the default value for Num Levels to 2 from the 4 that is entered by default. The Num Levels value for `Marital_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.

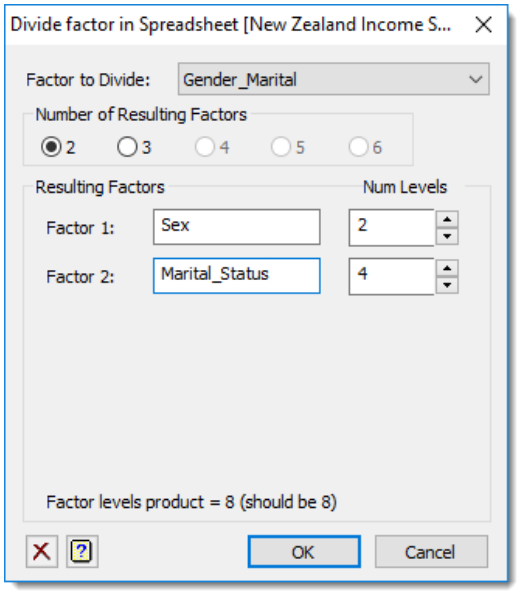


Figure 4.29

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

Figure 4.30

The factors do not have labels, so these would need to be manually entered Spread | Factor | Edit Levels and Labels. This dialog is shown for editing the factor column `Sex` in Figure 4.31. Now close any open spreadsheets and clear the data by selecting Data | Clear All Data.

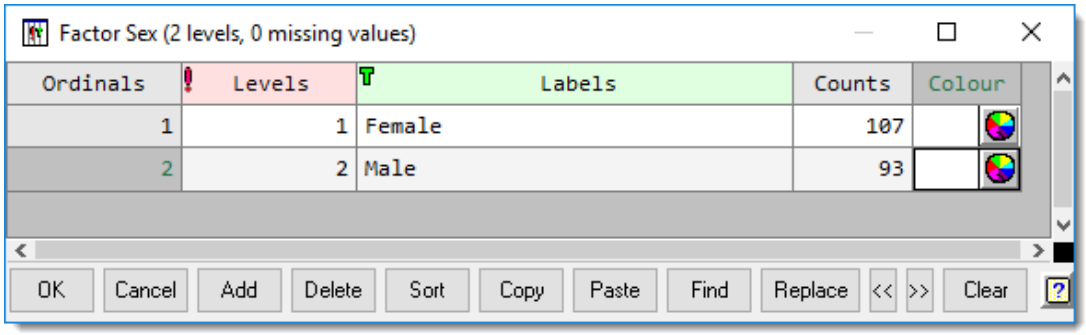


Figure 4.31

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. You can do this using the Standardize Levels menu. To illustrate this menu, we will open the two Excel files (`Grazing 1.xls` and `Grazing 2.xls`). Select File | Open, then select both files and open them. In Genstat the import Excel dialog will appear so click Finish twice to import the spreadsheets without making changes. If the Genstat Spreadsheet Warning dialog appears "This sheet contains column names used in other spreadsheets", click OK to close it. Note the labels for `Treatment` in `Grazing 2.xls` are in lower case.

Now select Spread | Factor | Standardize Levels to open the dialog in Figure 4.32. Selecting the two `Treatment` columns by double-clicking these then select the Case of Labels as Upper. Clicking OK standardizes the factor labels.

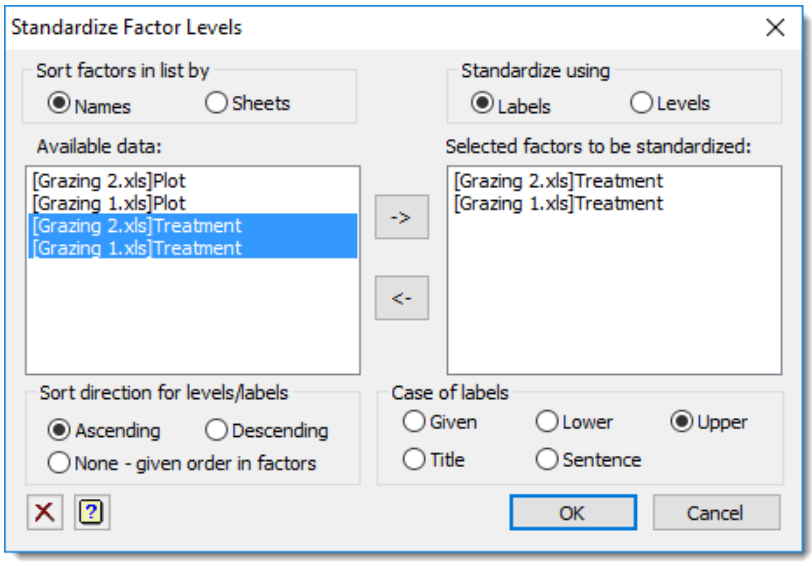


Figure 4.32

4.5 Summaries over groups

If you want to form summaries over groups from a data set and put these results into a vector spreadsheet, you can do this using the Summary Stats submenu. If you want the results to be in a table rather than a vector spreadsheet, you would use the Summary Tables submenu instead. Open the file `New Zealand Income Survey.gsh` then select Spread | Calculate | Summary Stats. Multi-select the factors `Gender`, `Qualification` and `Marital` and click -> to move them into the Summary Groups field. Now click the Counts button. Counts of rows will appear in the Summary Statistics field. In the Variates field multi-select `Age`, `Hours` and `Income` then click the Mean button (as in Figure 4.33) and then click OK. This produces the spreadsheet in Figure 4.34, 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_.

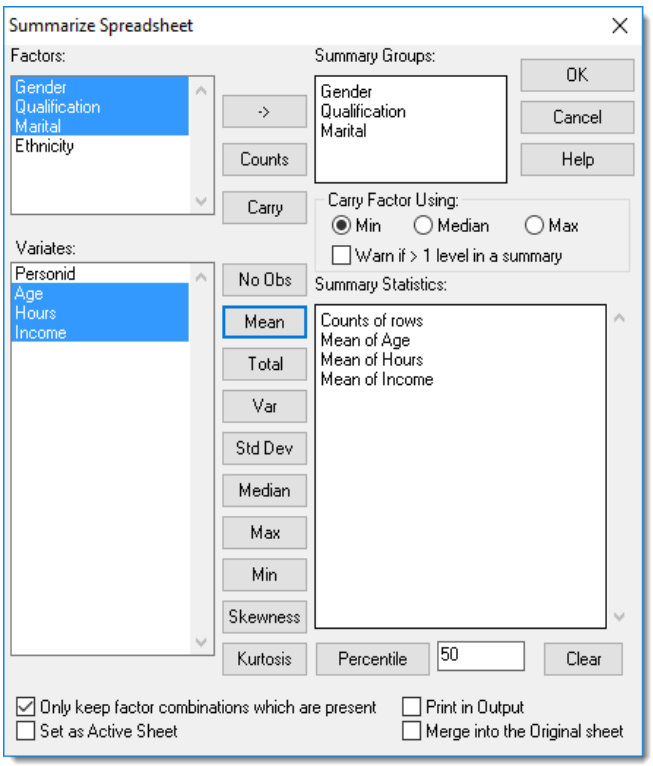


Figure 4.33

This is to avoid the summaries overwriting the original data in the Genstat server. If the Merge into the 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.

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
5	female	school	never	18	21.8333	16.6111	270.111

Figure 4.34

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 the Spread | Delete | Restricted rows menu or duplicate the sheet with just the included rows in the new sheet using the Spread | Manipulate | Duplicate menu. It normally is safer to uplicate 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 Spread | Restrict/Filter | Random Rows menu or a random subset of rows using the Spread | Manipulate | Split/Subset menu.

Let's create a subset of data by taking 100 random rows from a 200-row spreadsheet. Use the file `New Zealand Income Survey.gsh` that we worked with in the previous section. Click the spreadsheet to make it active then from the menu select Spread | Restrict/Filter | Random Rows (Figure 4.35). Clicking OK filters the spreadsheet to include 100 of the rows.

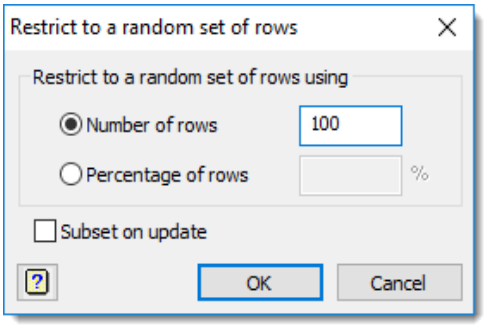


Figure 4.35

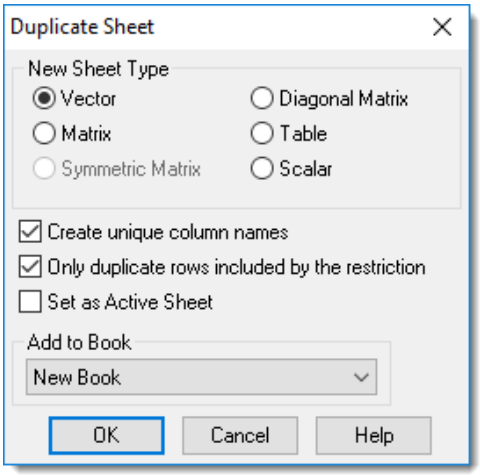
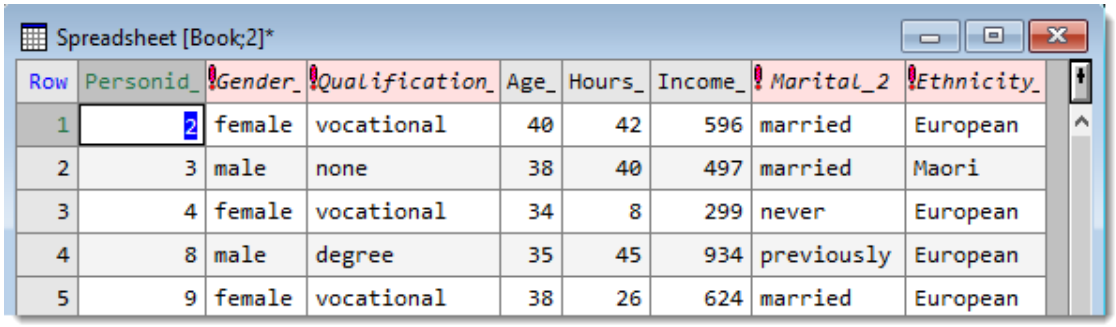


Figure 4.36

We can then use the Spread | Manipulate | Duplicate menu to obtain the dialog shown in Figure 4.36. It is important that the item Only duplicate rows included by the restriction is ticked (normally if the spreadsheet is restricted, this will be ticked by default). Clicking OK creates a spreadsheet similar to that shown in Figure 4.37.



Row	Personid_	Gender_	Qualification_	Age_	Hours_	Income_	Marital_2	Ethnicity_
1	2	female	vocational	40	42	596	married	European
2	3	male	none	38	40	497	married	Maori
3	4	female	vocational	34	8	299	never	European
4	8	male	degree	35	45	934	previously	European
5	9	female	vocational	38	26	624	married	European

Figure 4.37

Leaving the second spreadsheet aside, click the original spreadsheet again to give it the focus. Now select Spread | Manipulate | Split/Subset. This gives the dialog in Figure 4.38. We need to select Subset to a single sheet and select Random sampling as the Split sheet using option. Set the number of samples to `100` and Weighting to `<Equal>`. Clicking OK will create a sheet like that in Figure 4.37, but the names will end in _2 rather than _1 so that they are unique. If we had the Operation as Split into multiple sheets, we would have obtained two sheets with a set of 100 random rows in one and the remaining 100 rows in the second sheet.

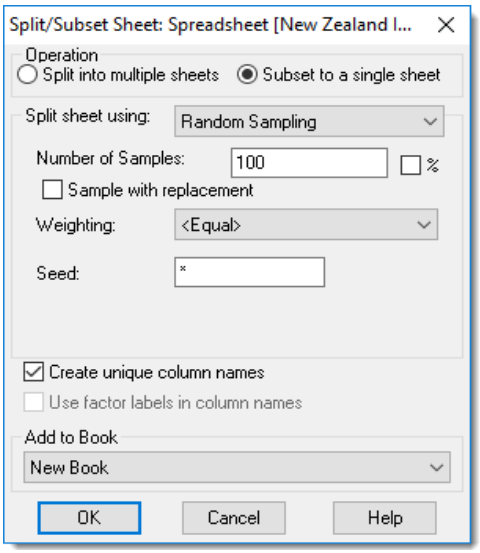


Figure 4.38

We can also use the Split/Subset menu to split the sheet into two data sets. To create one spreadsheet for males and one for females, click on the original spreadsheet again to give it the focus then select Spread | Manipulate | Split/Subset. Select options as shown in Figure 4.39, ensuring that you select Split into multiple sheets. Set Factor Groups to Keep as `<All Levels>` so that each level of `Gender` will have its own 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 to a single sheet.

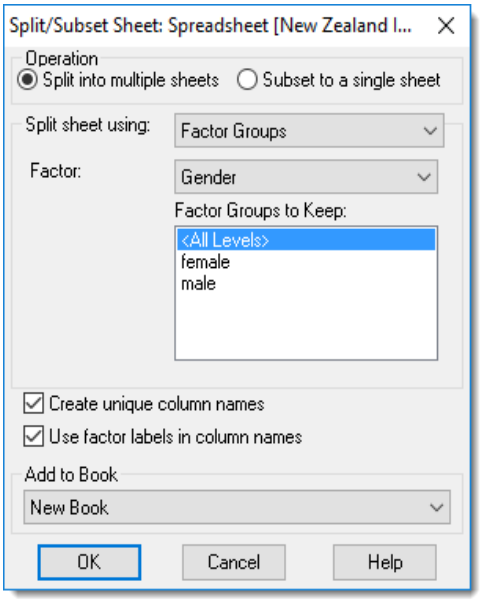


Figure 4.39

Clicking OK will create two sheets like that in Figure 4.40.

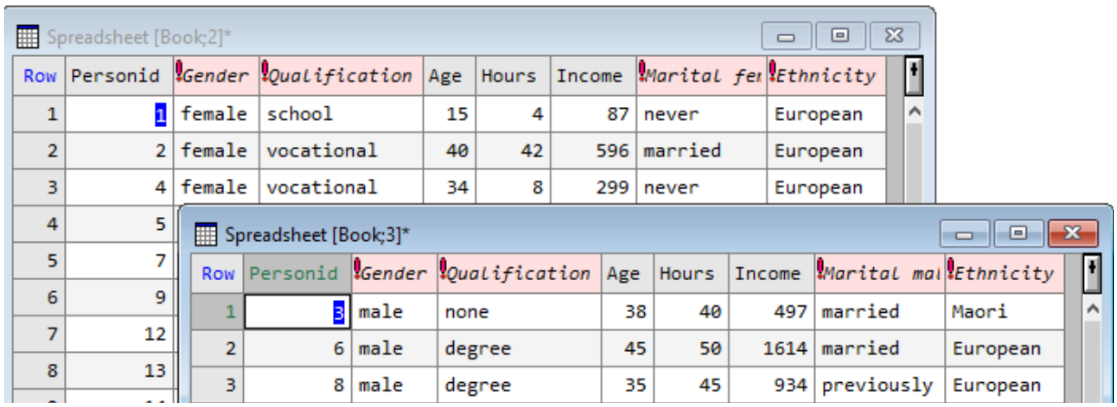


Figure 4.40

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. The following section will explain how tables can be manipulated in a number of ways within a spreadsheet.

5.1 Creating tables from Genstat menus

Several of the Genstat Statistics menus can create tables (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). 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.

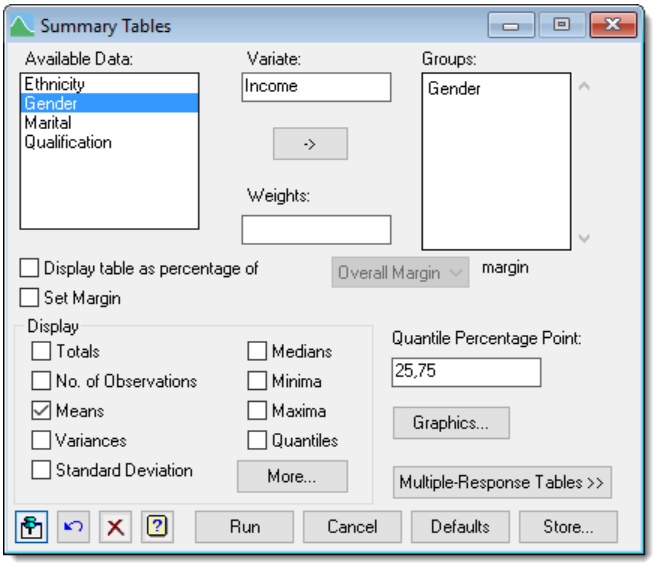


Figure 5.1

Open the spreadsheet file `New Zealand Income Survey.gsh` then select Stats | Summary Statistics | Summary Tables. This will open 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. Select No. of Observations, Means and Standard Deviation then enter the names as shown in Figure 5.2 to store these in 3 named tables. At the bottom of the dialog tick Display Tables in Spreadsheet using and click OK to close this dialog.

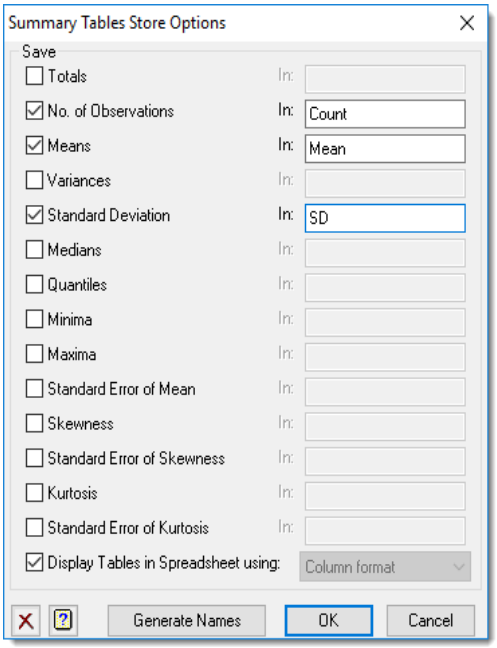
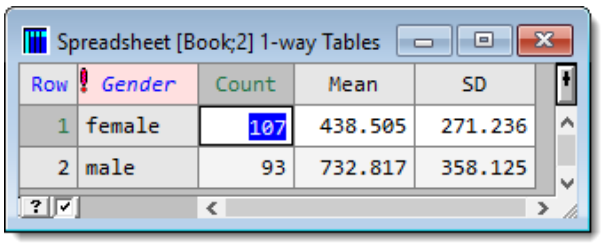


Figure 5.2

The tables are not saved until we click the Run button so do this now to produce the spreadsheet in Figure 5.3. Some dialogs have Save buttons rather than Store buttons and these are enabled after the main dialog has been run, and then create the structures when the Save dialog is closed.

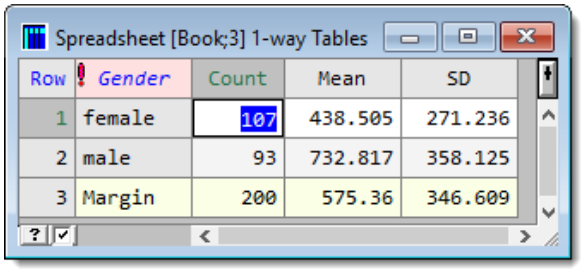
The Store button options need to be specified before the main menu is run and so do not create the structures when they are closed.


Spreadsheet [Book;2] 1-way Tables

Row	Gender	Count	Mean	SD
1	female	107	438.505	271.236
2	male	93	732.817	358.125

Figure 5.3

Note how we have the 3 one-way tables in the same spreadsheet. If you wanted to also produce overall summaries across both genders (margins), you would need to tick the option Set Margin in the Summary Tables dialog. Setting this and clicking the Run button would give you the spreadsheet shown in Figure 5.4. Note how the margin cells are shaded. You can change the default colour by selecting Tools | Options, then clicking the Fonts and Colours tab. Select `Spreadsheets` from the first dropdown list then select `Table margins background` from the Display items list. Use the colour selector to specify a different colour then click OK.

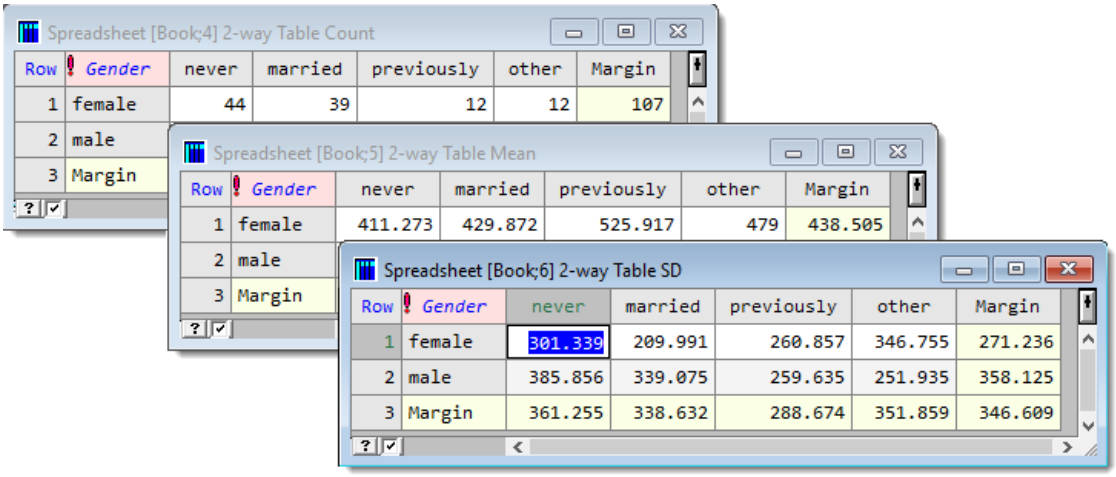

Spreadsheet [Book;3] 1-way Tables

Row	Gender	Count	Mean	SD
1	female	107	438.505	271.236
2	male	93	732.817	358.125
3	Margin	200	575.36	346.609

Figure 5.4

Now close the one-way tables.

If we want a cross tabulation of `Gender` by `Marital` status, we can add to the Groups field in the Summary Tables dialog by double-clicking `Marital` in the Available Data list and click Run. This will now create a two-way table. A spreadsheet can only hold a single two-or-more-way table so clicking Run will give us 3 spreadsheets for the counts, means and standard deviations. The resulting spreadsheets are shown in Figure 5.5.


Spreadsheet [Book;4] 2-way Table Count

Row	Gender	never	married	previously	other	Margin
1	female	44	39	12	12	107
2	male					
3	Margin					

Spreadsheet [Book;5] 2-way Table Mean

Row	Gender	never	married	previously	other	Margin
1	female	411.273	429.872	525.917	479	438.505
2	male					
3	Margin					

Spreadsheet [Book;6] 2-way Table SD

Row	Gender	never	married	previously	other	Margin
1	female	301.339	209.991	260.857	346.755	271.236
2	male	385.856	339.075	259.635	251.935	358.125
3	Margin	361.255	338.632	288.674	351.859	346.609

Figure 5.5

If you would prefer the table to display with the marital 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 marital 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.

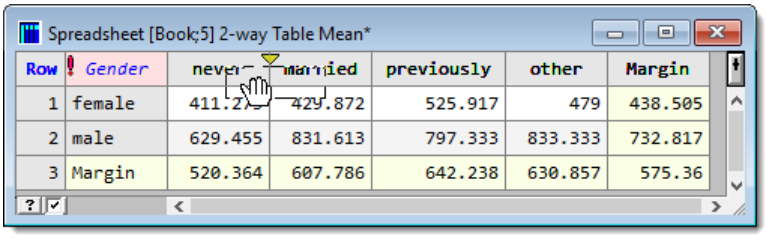


Figure 5.6

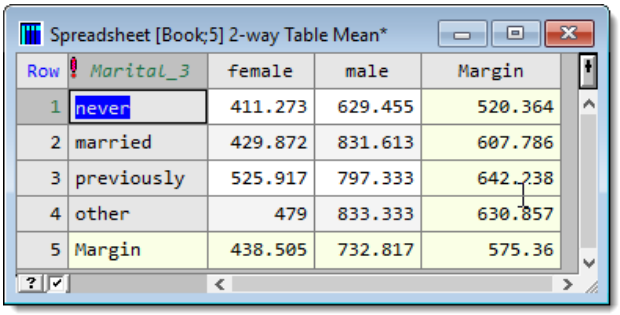


Figure 5.7

If you want to change the number of decimal places in a column select Spread | Column | Attributes/Format. To do this for all columns, select them by clicking on their headings before using this menu. This will open a dialog that lets you change the attributes of all selected columns as shown in Figure 5.8.

Now close the two-way tables.

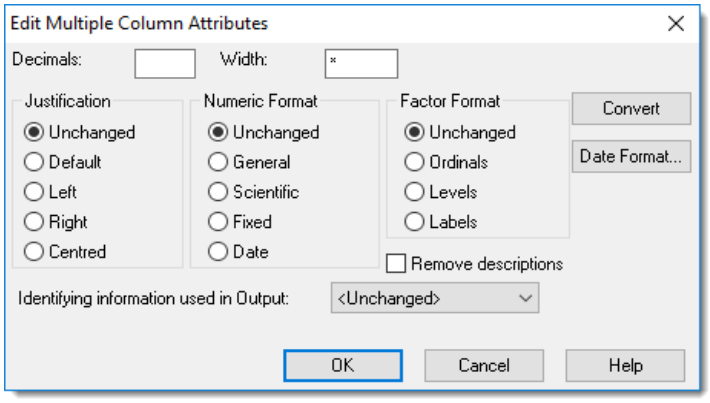


Figure 5.8

Tables with up to 9 classifying factors can be displayed in a spreadsheet. To create a three-way table with the extra classifying factor Qualification we can double-click this in the Available Data list to add it to the Groups field as shown in Figure 5.9.

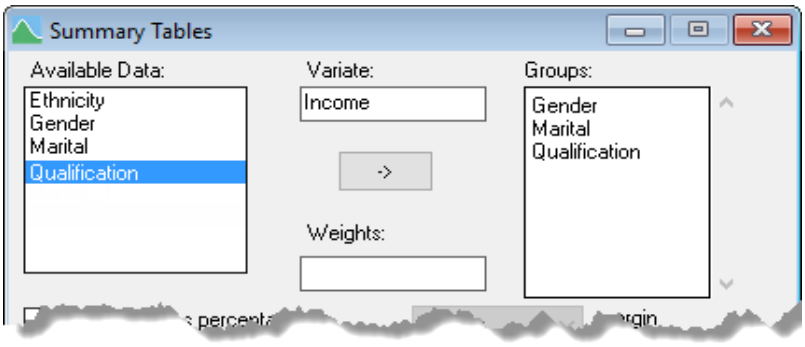


Figure 5.9

Clicking Run will create 3 spreadsheets again (the means table is shown in Figure 5.10). 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. Look at each table in turn to see that they all have the same structure.

Spreadsheet [Book;8] 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 5.10

If you want to hide individual rows, select them first then from the menu select Spread | Restrict/Filter | Selected Rows | Set as Excluded rows. Alternatively, if all the margins of a factor were to be hidden, you could put the focus on the cell showing Margins and select Spread | Restrict/Filter | Values Not equal to the current Cell to hide all the margins of this factor. Figure 5.11 shows the table with the margins of `Gender` hidden.

Spreadsheet [Book;8] 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

Figure 5.11

Columns in a spreadsheet can be hidden by selecting Spread | Column | Hide/Show as shown in Figure 5.12. 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 5.13 shows the table spreadsheet with the `Margin` column hidden.

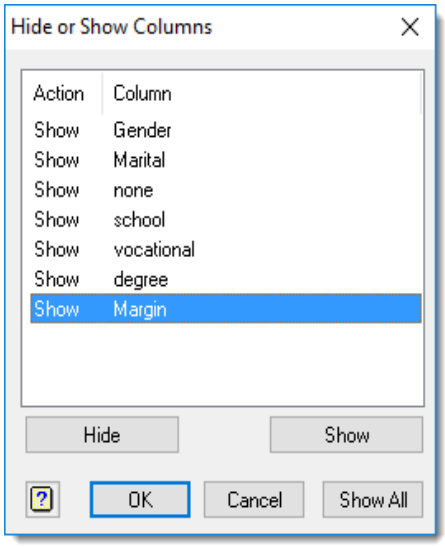


Figure 5.12

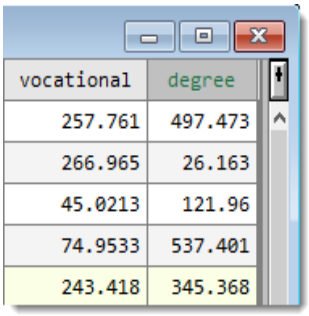


Figure 5.13

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 5.14 shows this being done for the final column to redisplay the hidden `Margin` column.

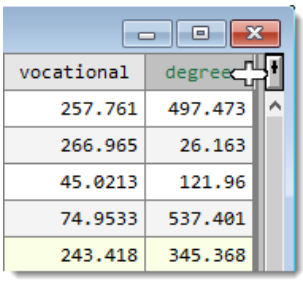


Figure 5.14

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 time using Spread | Calculate | Table Margins. This opens the dialog shown in Figure 5.15. 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.

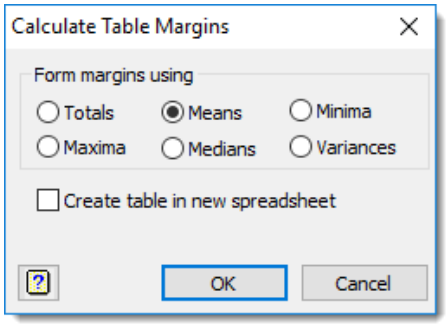


Figure 5.15

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.

On the Summary Tables dialog (Figure 5.16) click the Store button to re-open the Summary Tables Store Option dialog (Figure 5.17). Deselect Means and Standard Deviation, then select Tabbed format from the dropdown list at the bottom of the dialog and click OK.

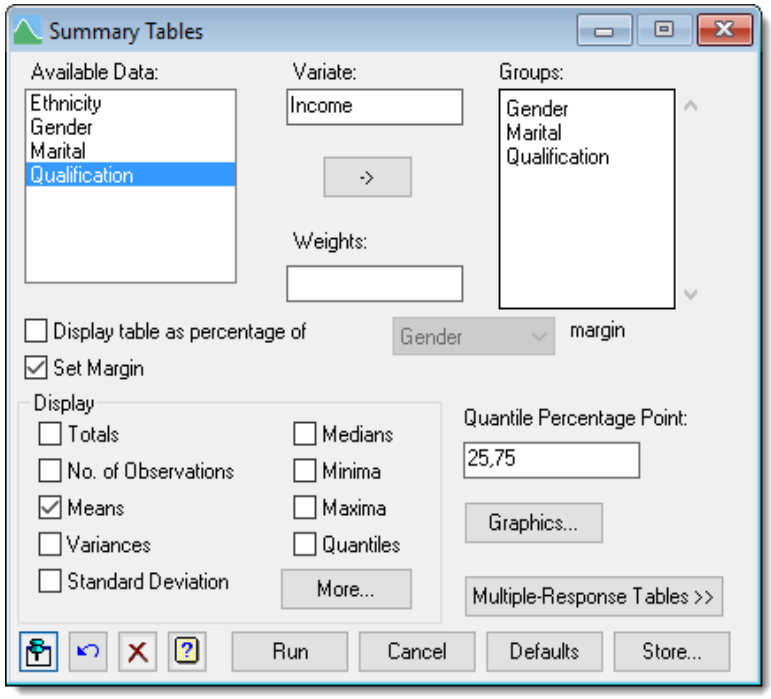


Figure 5.16

Clicking Run on the Summary Tables dialog produces the spreadsheet shown in Figure 5.18 with the genders across the tabs. The first group in the Groups field is used across the tabs.

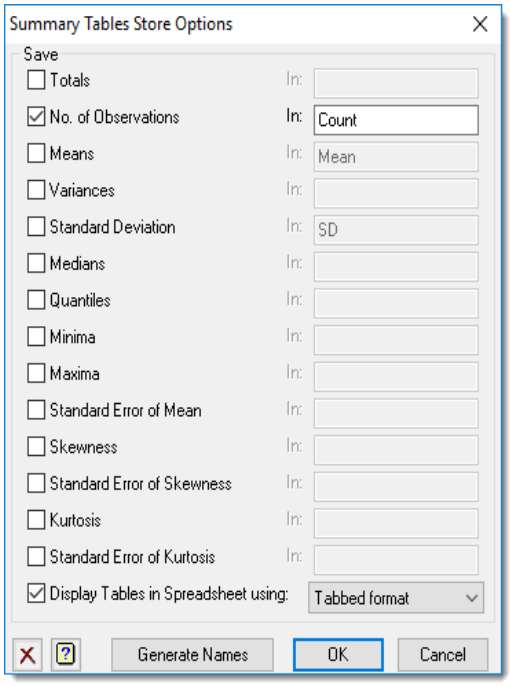
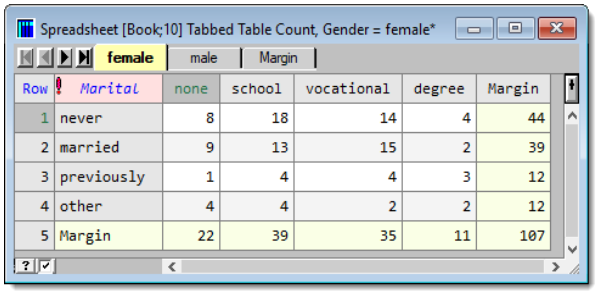


Figure 5.17



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 5.18

An alternative method for creating a tabbed table is to select Spread | New | Tabbed-table from Genstat. This opens the dialog shown in Figure 5.19 which displays all the 3+ way tables in the Genstat server (central data pool). Double-click the table `Mean` to move it into the Selected table field. Now put the cursor into Factor across tabs to list the available factors. Double-click the factor `Qualification` to put its groups across the tabs in the table. Note: you must be careful to specify a factor that is a classifying factor of the table, otherwise you will get an error. Set options in the dialog as shown then click OK to produce the spreadsheet displayed in Figure 5.20.

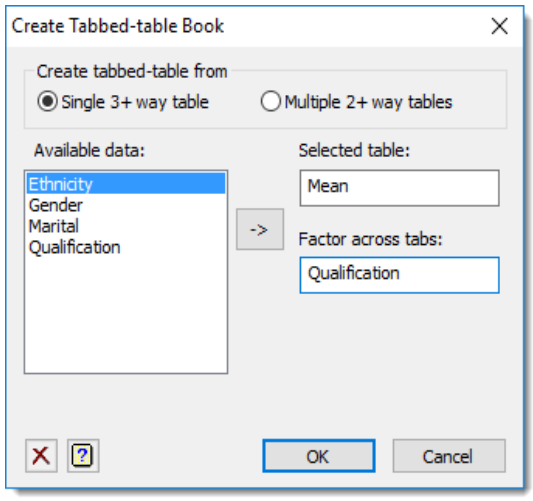


Figure 5.19

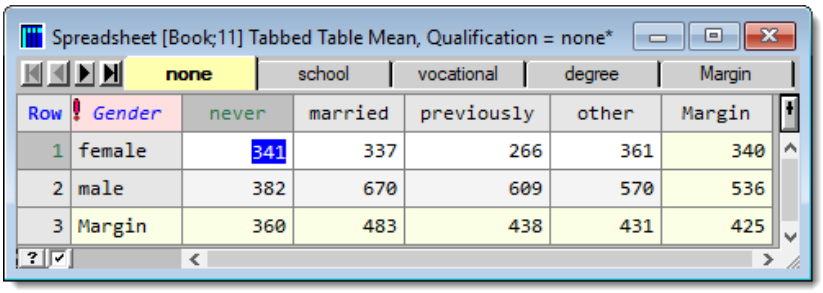


Figure 5.20

The other use of tabbed-tables is to put multiple 2+ way tables into a single table. Select Spread | New | Tabbed-table from Genstat then select the Multiple 2+ way tables option. To put the statistics in the `Count`, `Mean` and `SD` tables into a single tabbed-table, set options as shown in Figure 5.21.

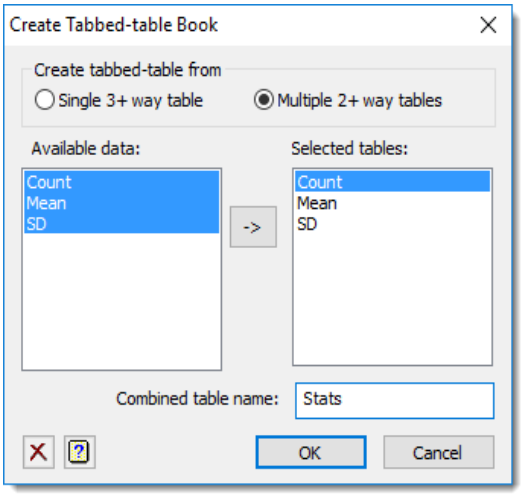


Figure 5.21

Clicking OK produces 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).

Spreadsheet [Book;11] Tabbed Table Count*

Count:Income | Mean:Income | SD:Income

Row	Gender	Marital	none	school	vocational	degree	Margin
1	female	never	8	18	14	4	44
2	female	married	9	13	15	2	39
3	female	previously	1	4	4	3	12
4	female	other	4	4	2	2	12
5	female	Margin	22	39	35	11	107
6	male	never	7	21	10	6	44
7	male	married	7	5	12	7	31
8	male	previously	1	1	3	4	9
9	male	other	2	0	7	0	9
10	male	Margin	17	27	32	17	93
11	Margin	never	15	39	24	10	88

Figure 5.22

We can now use this table with other spreadsheet menus to display the statistics side by side in rows or columns.
Select Spread | Manipulate | Reorder table. Set the factor order as shown in Figure 5.23 then click OK to produce the spreadsheet displayed in Figure 5.24.

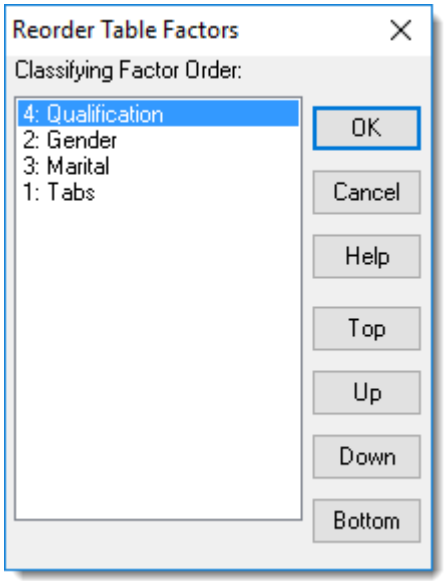


Figure 5.23

Spreadsheet [Book;7] Tabbed Table Stats, Qualification = none*

none | school | vocational | degree | Margin

Row	Gender	Marital 1	Count:Income	Mean:Income	SD:Income	Margin
1	female	never	8	341	205	*
2	female	married	9	337	137	*
3	female	previously	1	266	*	*
4	female	other	4	361	139	*
5	female	Margin	22	340	156	*
6	male	never	7	382	238	*
7	male	married	7	670	249	*
8	male	previously	1	609	*	*
9	male	other	2	570	131	*
10	male	Margin	17	536	254	*
11	Margin	never	15	360	214	*
12	Margin	married	16	483	253	*
13	Margin	previously	2	438	243	*
14	Margin	other	6	431	163	*
15	Margin	Margin	39	425	224	*

Figure 5.24

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 illustrate the bookmark facilities select File | Open then locate and open the spreadsheet file `Sulphur.gsh`.

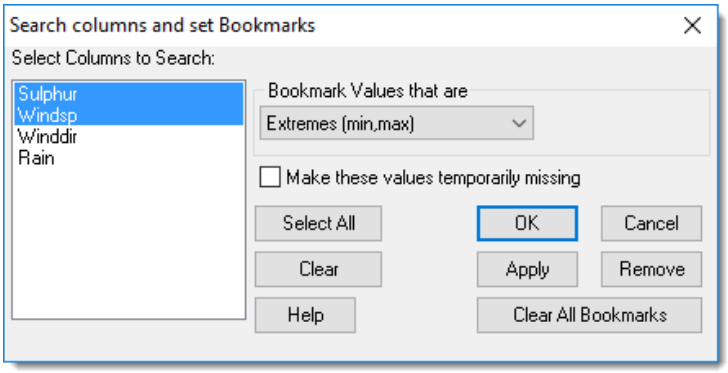
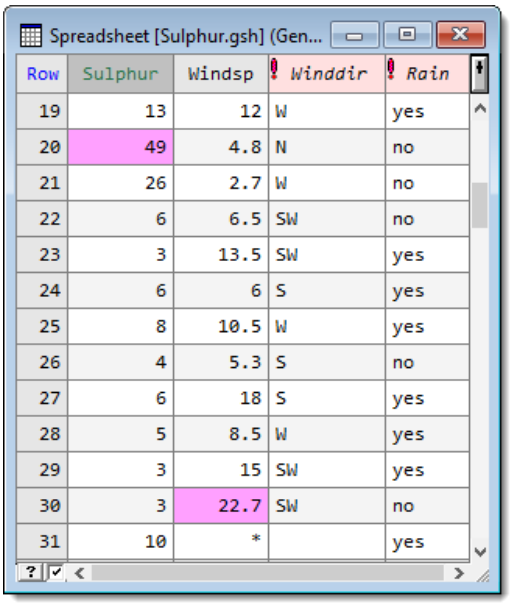


Figure 6.1

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. From the menu select Edit | Bookmark | By Value to open the dialog in Figure 6.1. Here we have selected both `Sulphur` and `Windsp` from the Select Columns to Search list, and Extremes (min, max) from the Bookmark Values that are 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 Edit | Bookmark | Next. 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. Click inside the bookmarked cell within the column `Sulphur` at row 20 and select Edit | Bookmark | Add Note to open the dialog in Figure 6.3.

This small, resizable text editor lets you enter a comment for the bookmarked cell. By default, this editor contains 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`. Click Cancel to close the editor.

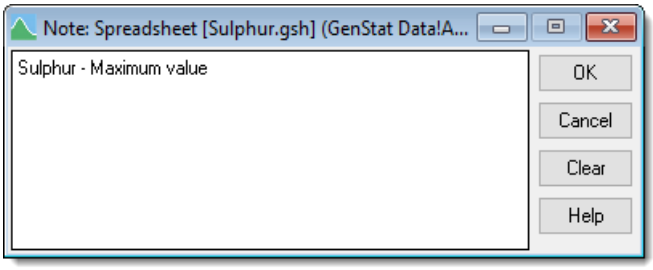


Figure 6.3

If a spreadsheet containing bookmarks is saved into a Genstat spreadsheet file, the bookmarks will be retained when the file is opened again. Clear the bookmarks by electing Edit | Bookmarks | Clear All.

Another useful feature for visually displaying values that fall into different categories or conditions is the Conditional Formatting dialog shown in Figure 6.4. Click anywhere in the spreadsheet to give it focus then select Spread | Column | Conditional Formatting. This dialog lets you set up to 3 conditions to differentiate your data by colour. Here 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). Conditions are matched sequentially, so if a cell matches condition 1 then conditions 2 and 3 will be ignored. Thus, the ordering of conditions can be important. 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. Set the options as shown in Figure 6.4 then click OK to redisplay the values within the column `Sulphur` in the chosen colours.

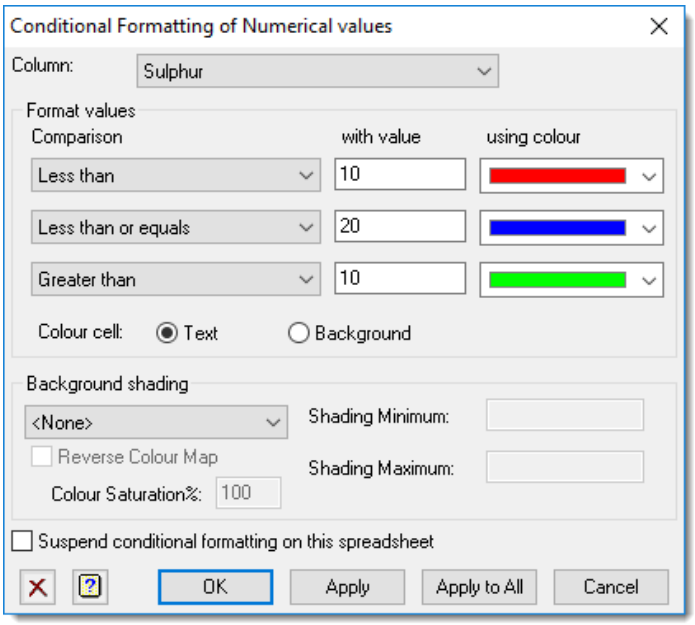


Figure 6.4

To turn off conditional formatting you can close the spreadsheet without saving your formatting changes or select Spread | Column | Conditional Formatting then select Suspend conditional formatting on this spreadsheet.

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 (you can find this under the menu Help | Genstat Guides). 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 Spread | Column | Temporary missing values to open the dialog shown in Figure 6.5. Here we have selected `Sulphur` and entered row 1 (where the value 0 is located).

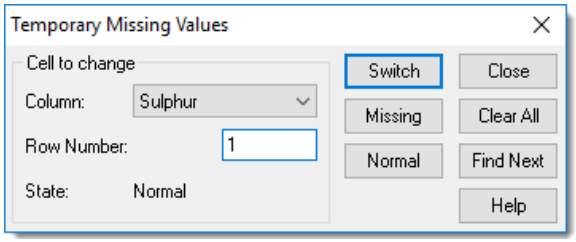


Figure 6.5

Click Missing to change the cell value to be temporarily missing then click Close to produce the spreadsheet in Figure 6.6.

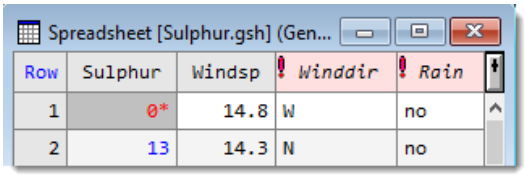


Figure 6.6

The temporary missing cell is disabled and has an asterisk appended to the value in the cell.

Alternatively, you can either toggle the status of the current cell with the Alt+F8 key or click the Temporarily Missing button on the toolbar  .

6.1 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.

Select Edit | Bookmark | By Value to bookmark the column time to show the maximum and minimum values. 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.

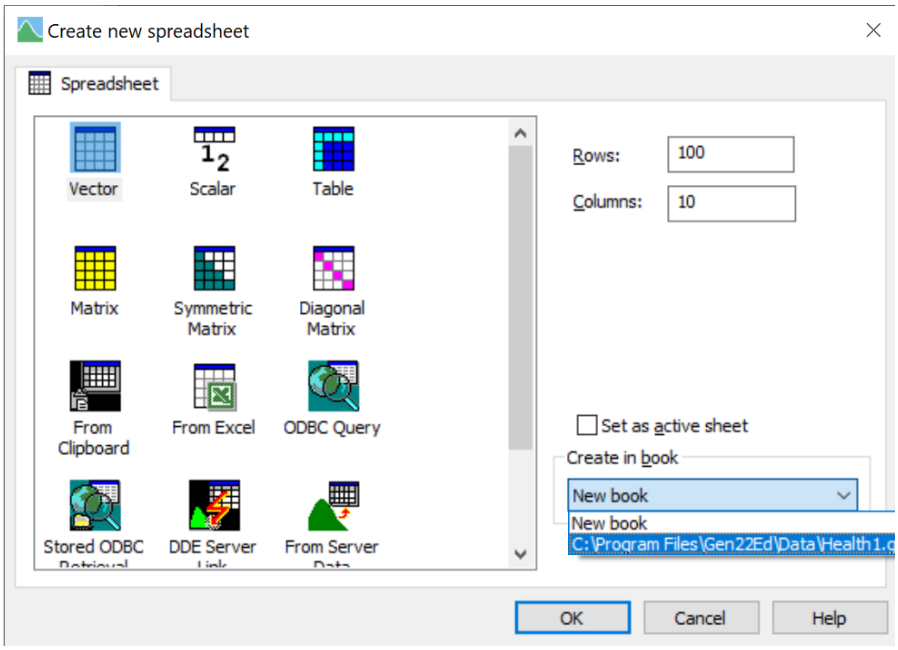


Figure 7.2

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 Spread | Book | Add Sheet which opens the dialog shown in Figure 7.1. The data in our example will be in 3 rows with 2 columns of variates, so we have selected the Vector spreadsheet icon, and entered 3 rows and 2 columns. 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.

Clicking OK adds a new sheet to the book called `Sheet2`. In `Sheet2`, rename the columns as shown in Figure 7.2 by double-clicking the column heading `C1` and typing `ID` into the Name field, then repeat with column `C2`, changing the name to `Year`. Now enter the data values into the cells as shown. The current sheet is identified by the highlighted tab, which is `Sheet2`. To view another sheet within a book, click the sheet's tab or use the arrow navigation buttons on the top left of the window.

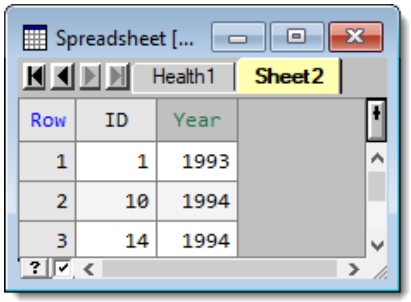


Figure 7.1

If a book contains many sheets, the quickest way to locate one is to select Spread | Book | Display Sheet. This produces the menu shown in Figure 7.3 where we have selected `Health1` as the sheet that we now wish to display.

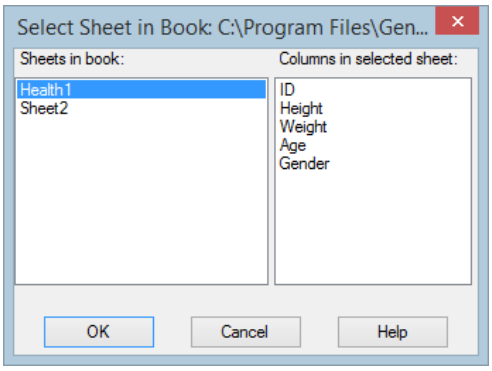


Figure 7.3

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. To change the name of `Sheet2` double-click its tab to open the editor shown in Figure 7.4. Type the name `Year` then click OK.

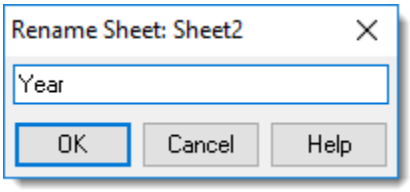


Figure 7.4

Spreadsheets can be moved or copied within a book or into other books. To illustrate this, we will add the data from the file `Health2.gsh` into a new sheet in the current book. Open the file `Health2.gsh` then select Spread | Book | Move Current Sheet. This opens the dialog shown in Figure 7.5, where we have selected Health1.gsh as the book to which the data from `Health2.gsh` are to be moved. Click OK. Genstat will warn you that the structure (column) name `ID` is duplicated. Click Yes to allow Genstat to make the structure name unique, which it will do by appending a number to the column name. The spreadsheet `Health2.gsh` will close and the new sheet will appear in your book as shown in Figure 7.6.

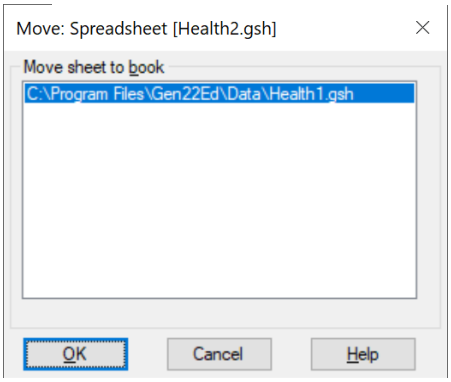


Figure 7.5

Sheets can be deleted from books by selecting Spread | Book | Delete Sheet, then selecting the sheet to remove from the displayed list. Alternatively, you can use the mouse to delete sheets. Click and drag the sheet outside the book then release the mouse. This creates a new spreadsheet which you can then close without saving (this deletes it). Similarly, sheets within a book can be reordered using the mouse. Click and hold the tab you want to move then drag the sheet to a different position. The cursor will change to a hand with a grid, 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.

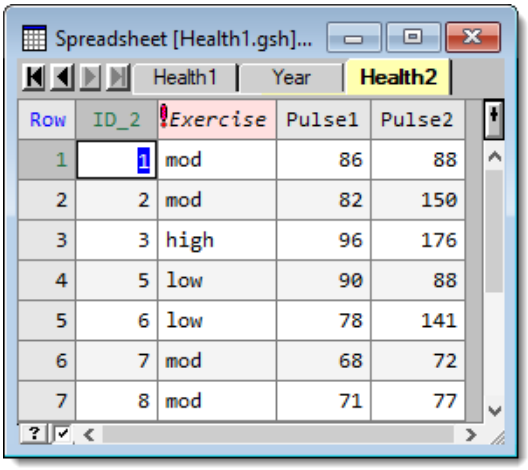


Figure 7.6

Figure 7.7 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 Spread | Book | Reorder Sheets.

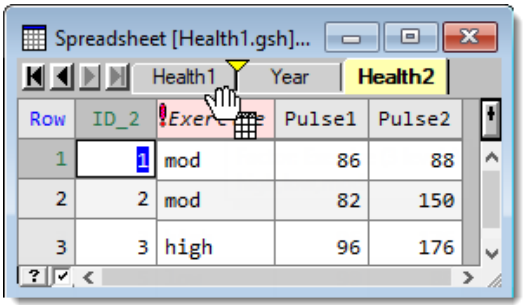


Figure 7.7

Sheets can be split into new books using the mouse or menus. To split the sheet called Health2 into a new book, we select Spread | Book | Split. This opens the dialog shown in Figure 7.8. Here we have selected the sheet [Health1.gsh]Health2, and clicked on -> 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. (If we selected Copy this would create a copy of the sheet in a new book and also keep the sheet in the original book.)

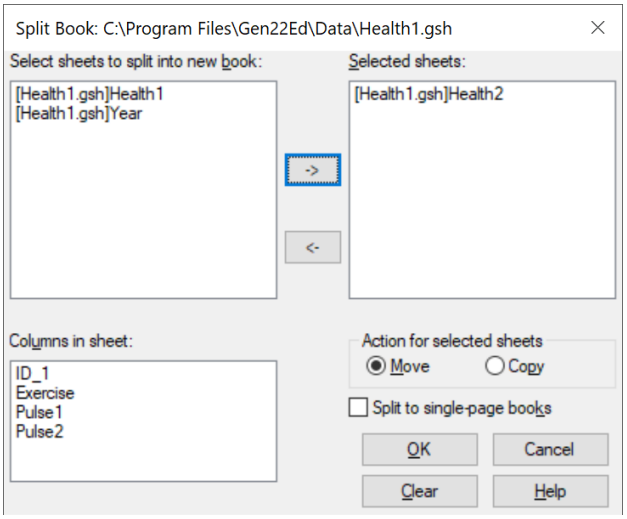


Figure 7.8

Clicking OK splits the book into two books (see Figure 7.9).

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 7.10).

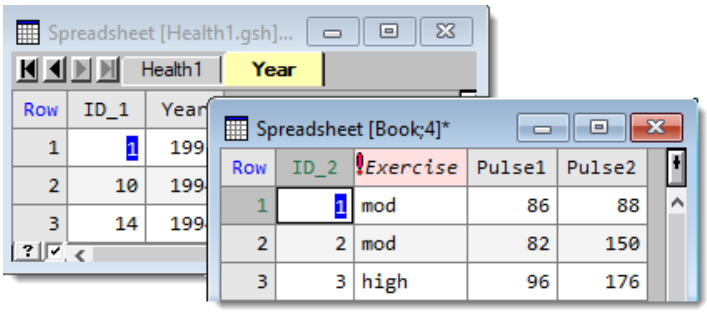


Figure 7.9

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) using File | Save As. The Genstat gwb and Excel formats are the only two formats which enable you to save multiple sheets within a book. Alternatively, individual sheets from a book can be saved as a Genstat Spreadsheet (*.gsh) file.

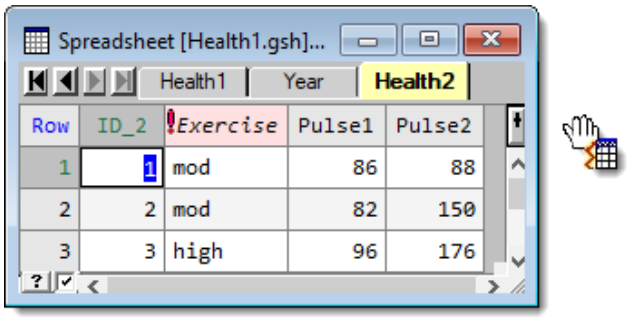


Figure 7.10

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 Spread | New | ODBC Data Query opens the dialog 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:

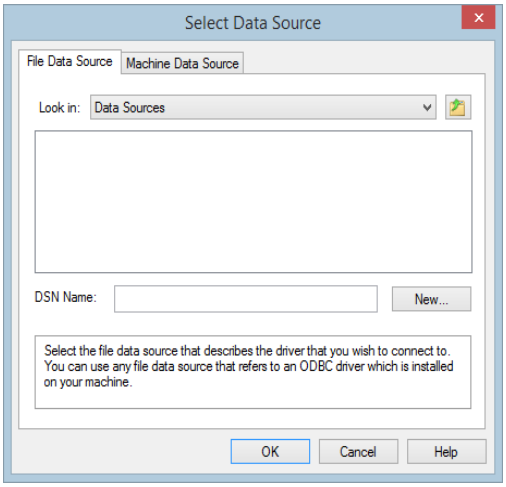


Figure 8.1

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`. Genstat, which is a 64-bit program, will see only the 64-bit version of ODBC. If you have a 32-bit version of Office, the Office ODBC drivers will be the 32-bit version. Therefore, the 64-bit version of Genstat will not be able to see the 32-bit Office ODBC drivers for Access or Excel. You should remove the 32-bit version of the Office ODBC drivers and install the 64-bit version. This can be found at: https://www.microsoft.com/en-us/download/details.aspx?id=13255.

On the dialog 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. You can find this file in the same folder as the other example data files, `C:\Program Files\Gen22Ed\Data`. However, as this is a read-only folder, before we begin this exercise, we'll need to save a copy a writable location. Navigate to `C:\Program Files\Gen22Ed\Data` and click on `cardata.mdb` to open it in Access. Using Save As… save a copy to a writable location, for example your Documents folder.

Selecting the File Data Source tab in the dialog shown in Figure 8.1 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, *.acdb) 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.

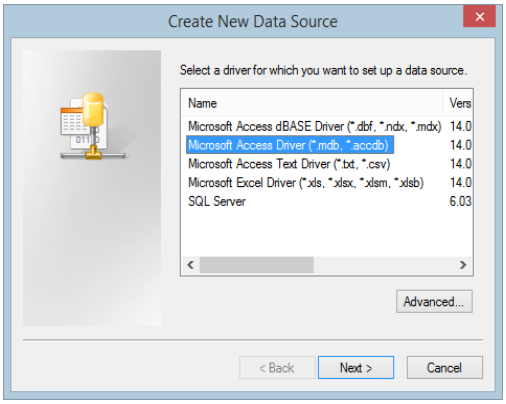


Figure 8.2

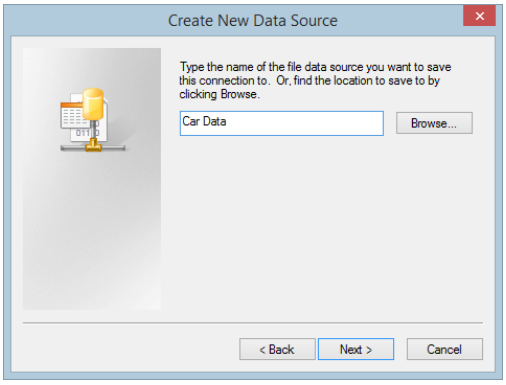


Figure 8.3

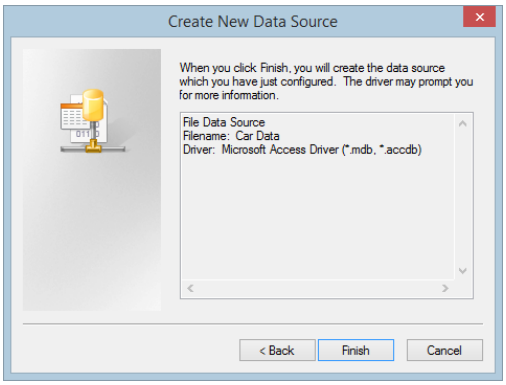


Figure 8.4

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`).

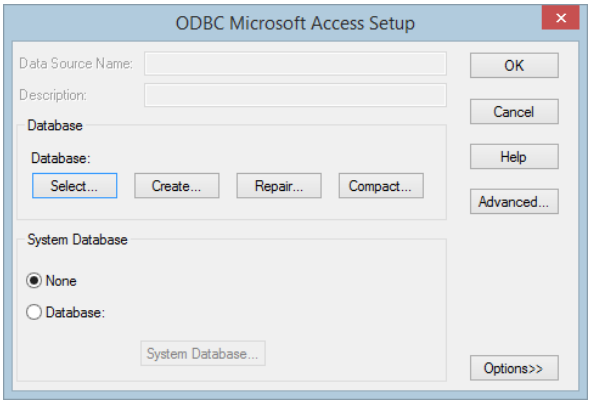


Figure 8.5

Clicking on Select opens a browse menu (see Figure 8.6) where we have selected the file `cardata.mdb`. (Note: we have previously saved a copy of `cardata.mdb` from `C:\Program Files\Gen22Ed\Data` to the Documents folder, a writable location). 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 Advanced 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.

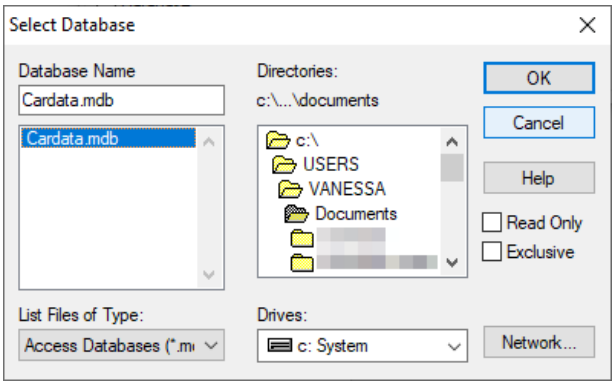


Figure 8.6

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

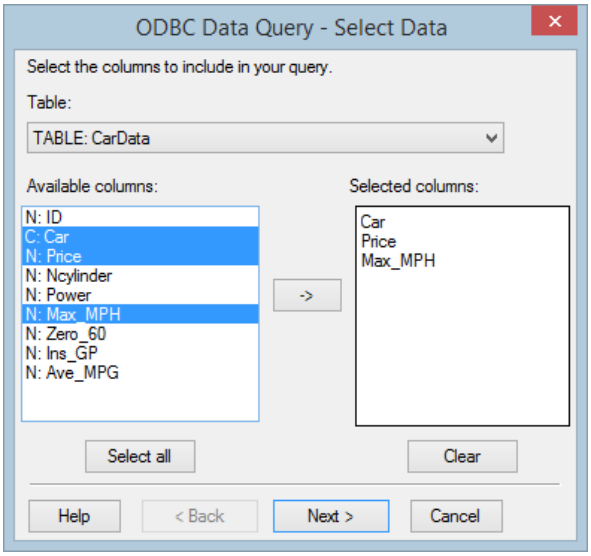


Figure 8.7

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 -> to copy the selected columns across to the Selected columns list.

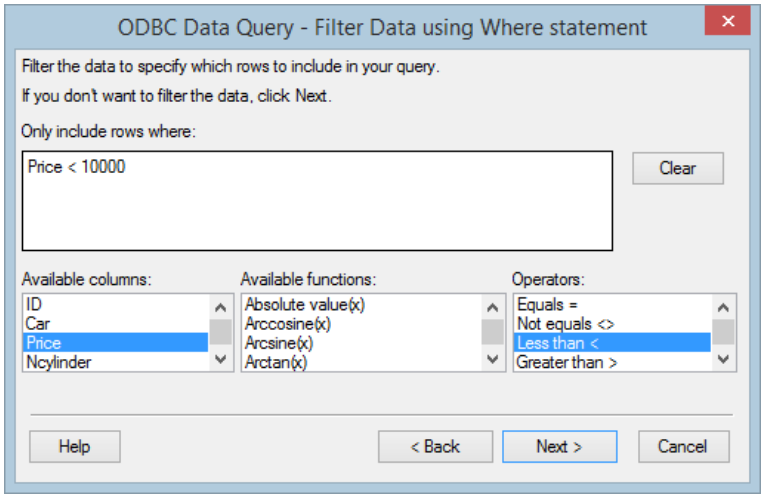


Figure 8.8

Clicking on Next opens a Filter dialog as shown in Figure 8.8. Here 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 create a filter to show only 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.

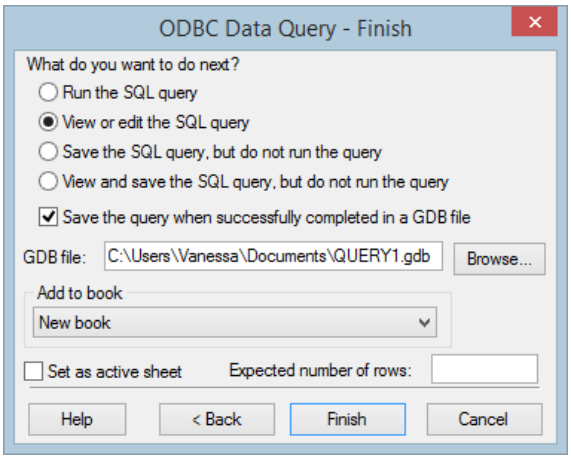


Figure 8.9

Figure 8.9 shows the final dialog 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 enables the field below it, and we can use the Browse button to select a location to save the data query. This will save the whole ODBC query process within a file called a Genstat .gdb file. A .gdb file can be opened File | Open and will automatically run the query on the ODBC Server specified within the file.

Clicking Finish opens the dialog 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.

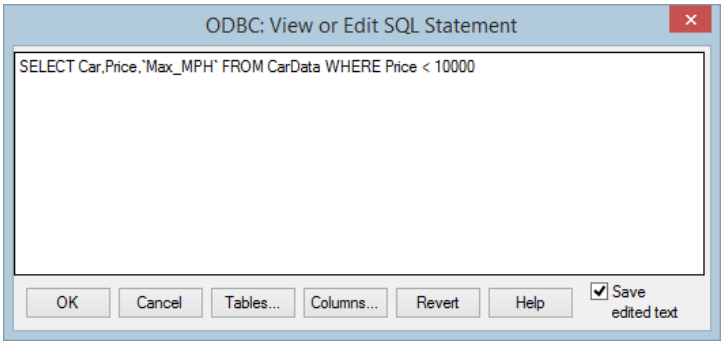
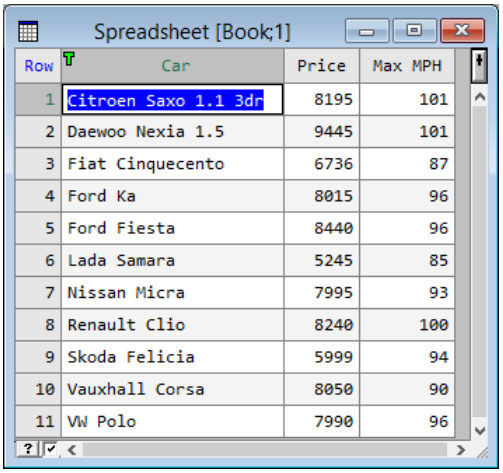


Figure 8.10

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.

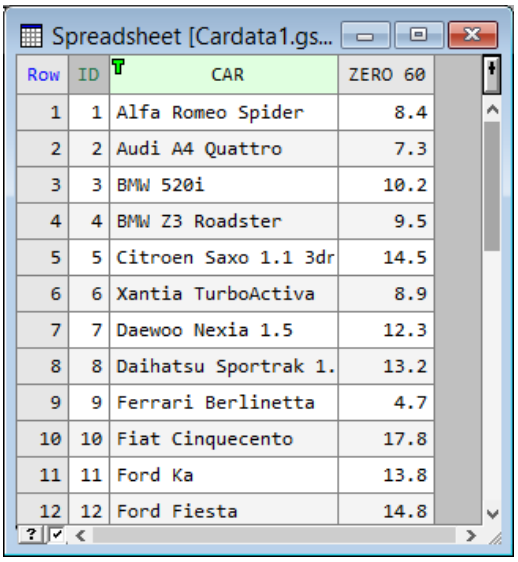


Row	Car	Price	Max MPH
1	Citroen Saxo 1.1 3dr	8195	101
2	Daewoo Nexia 1.5	9445	101
3	Fiat Cinquecento	6736	87
4	Ford Ka	8015	96
5	Ford Fiesta	8440	96
6	Lada Samara	5245	85
7	Nissan Micra	7995	93
8	Renault Clio	8240	100
9	Skoda Felicia	5999	94
10	Vauxhall Corsa	8050	90
11	VW Polo	7990	96

Figure 8.11

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. 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. Open this file in Genstat then select Spread | Export | Create Database Table. 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.	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.

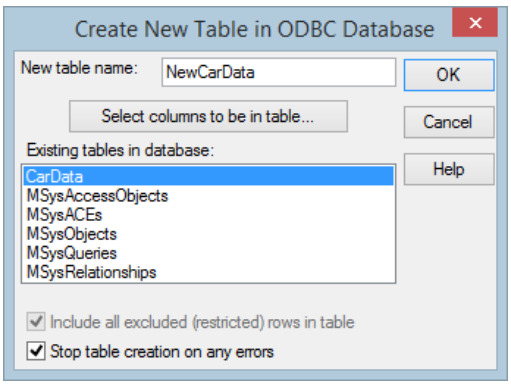


Figure 8.13

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.

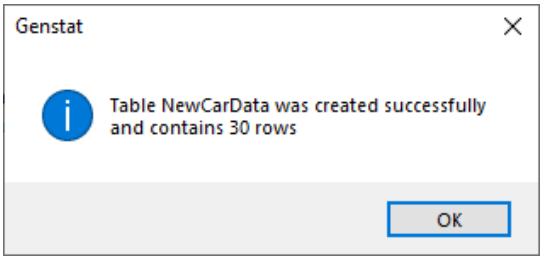


Figure 8.14

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

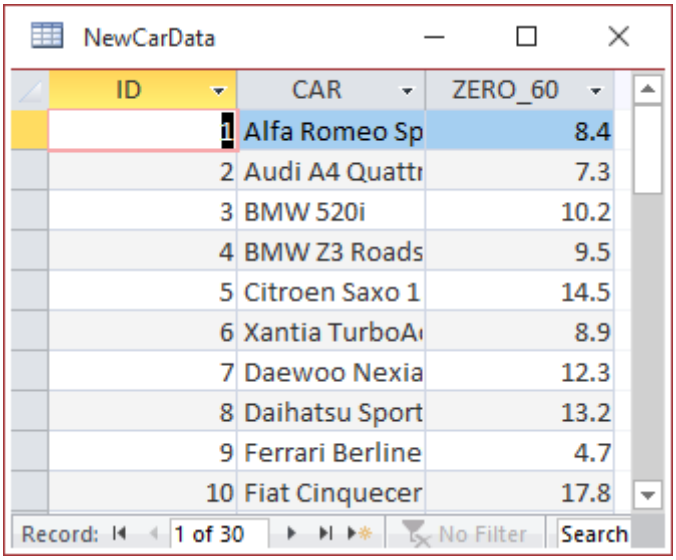


Figure 8.15

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), then select Spread | Export | Insert into Database Table.

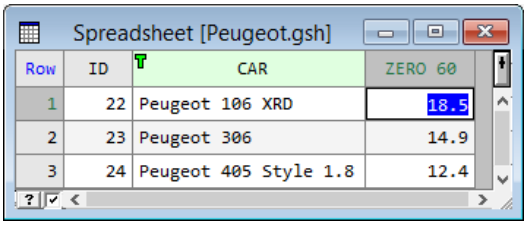


Figure 8.16

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.

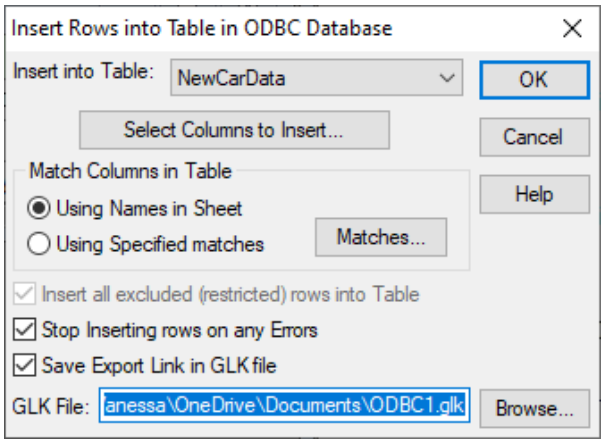


Figure 8.17

We have selected the Save Export Link in GLK file option and used the Browse button select a location to save the export link information.

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.

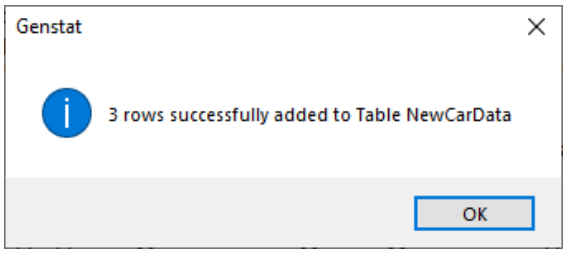


Figure 8.18

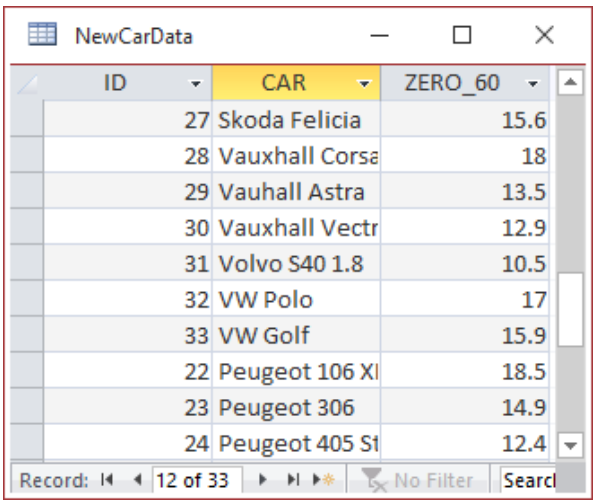


Figure 8.19

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.

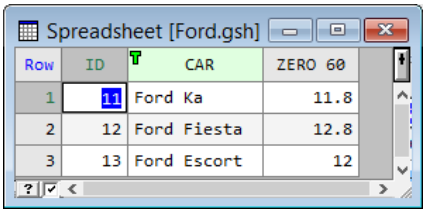


Figure 8.20

From the menu select Spread | Export | Merge with existing Database Table and double-click the Car Data DSN on the Select Data Source dialog. This opens the dialog 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 dialog 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 used the Browse button to select a location to save the Genstat ODBC Link file. A description of the other options on this menu can be found by clicking on the Help button.

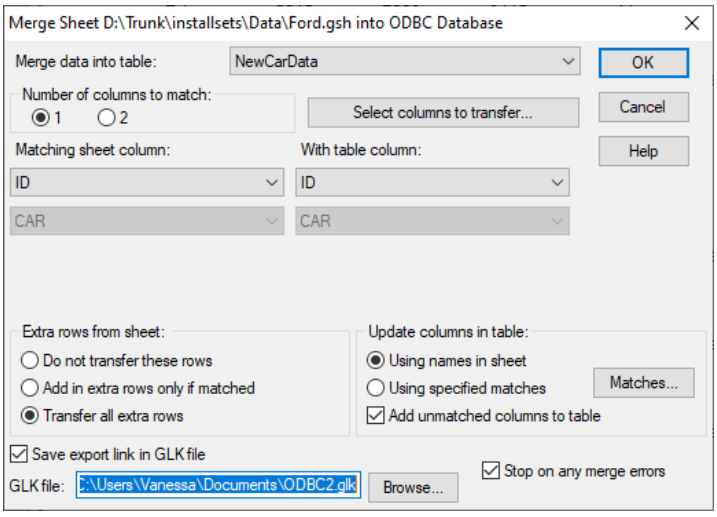


Figure 8.21

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

Figure 8.22

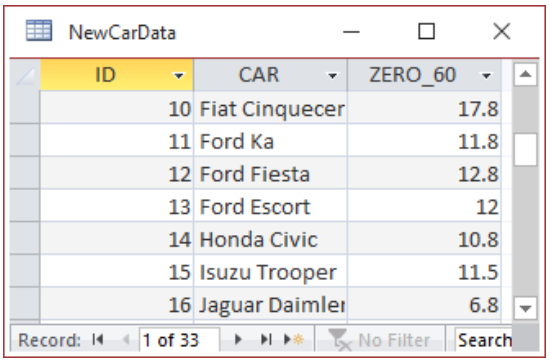


Figure 8.23

To run a Genstat ODBC Link file to automatically insert rows or merge data into a database, select Spread | Export | Run OBDC export link. This opens the dialog 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 `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.

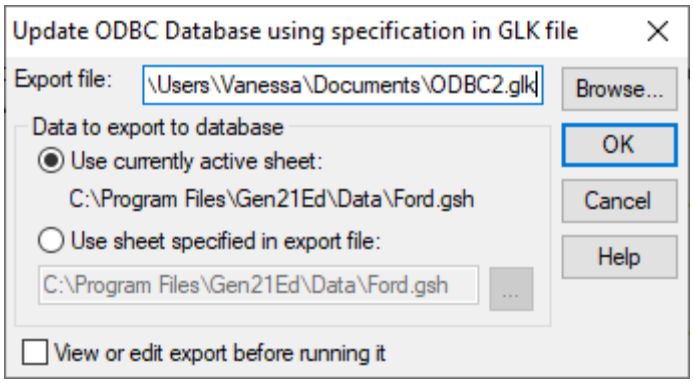


Figure 8.24

8.1 Exercise

Clear all the data from the Genstat data pool by selecting Data | Clear All Data. 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 £10,000 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. Re-open the .gdb file using the File | Open menu.

9 Other facilities

There are many other facilities for data manipulation using the spreadsheet menus 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 Spread | Set as active sheet. More details on active spreadsheets can be found in the online help.

Another method for rearranging data in Genstat is through the Paste Special menu. This is accessed using Edit | Paste Special. With this feature you can copy data onto the clipboard from another data source 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 Spread | Calculate | Summary Stats.

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 you can locate by selecting Help | Genstat Guides | Introduction to Genstat Command Language.

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. You can set the read only option and specify the minimum file size it applies to by selecting Tools | Spreadsheet Options and select the Books tab. 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 book can also be carried out using the command language. However, some of the features such as data verification, copying from the clipboard and bookmarks can only be performed in Windows™.

To use commands, open a new text window by selecting File | New, then click the General tab and select `Text Window`. Clicking OK opens an input window where you can type in your commands. To execute a typed command, select Run | Submit Line (or one of the other options).

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 formulas 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\\Gen22Ed\\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 8 you could supply the saved .gdb file as follows:

```
DBIMPORT GDBFILE='C:\\Program Files\\Gen22Ed\\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