

GenStat Introduction Workshop

Contents

1. What is GenStat?
2. How to run GenStat on the Lincoln Network
3. The windows
4. Loading data
5. Managing data
6. Editing a spreadsheet
7. Basic statistics
8. Drawing graphs
9. Saving your work
10. Exiting GenStat
11. Further information

1. What is GenStat?

GenStat is a statistical analysis software package developed at Rothamsted Experimental Station in the U.K. where many standard statistical techniques were first developed.

Brief comparison with other packages.

There are many other statistical software packages in use today, e.g., SAS, Systat, SPSS, Statistica, S-Plus, Minitab, etc. Several of these are available on the Lincoln Network, i.e. Systat, SPSS, and Minitab.

Some GenStat strengths:

- Offers many methods for transforming and rearranging data
- Very good for designed experiments
- Can fit a range of regression models, including non-linear models and GLMs
- Has useful diagnostic plots following analyses
- Has a powerful and flexible command language

How does GenStat work?

GenStat consists of two parts: a Server and a Client (Figure 1) (there is also a third part which draws the graphics, but that will be ignored just for now). When you run GenStat on a PC, you see the Client part which provides menus and helps you enter information. The client passes this information to the Server. The server runs in the background and analyses the data according to commands that you give it. You can either directly type commands in GenStat language and submit them to the server, or simply use the menus, or use a combination of commands and menus. Users new to GenStat often find the menus easiest to use, while advanced users find that using the command language gives more flexibility and power.

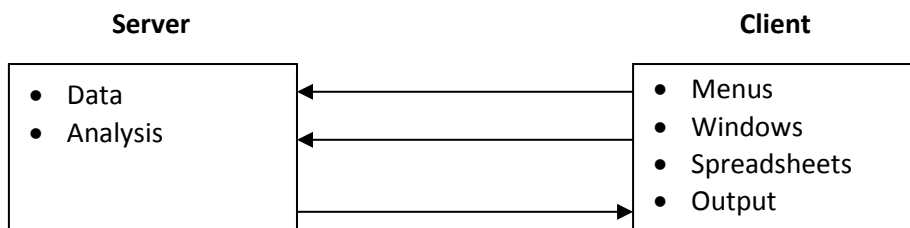



Figure 1 Simplified diagram of relationship between Client and Server.

2. Starting GenStat on the Lincoln Network

To run GenStat, start to type the word 'GenStat' into the Start menu's search box.

Once you start to type, all applications starting with 'G' will be displayed – click on  GenStat 16 to start the 16th edition.

Note: You may also find earlier versions of GenStat on the network. This workshop presents instructions and screen shots for the 16th edition.

(You can also access the GenStat software via:

Start > All Programs > LU Applications > Math & Stats Applications > GenStat 16)

When GenStat starts, the following Start Page might appear.

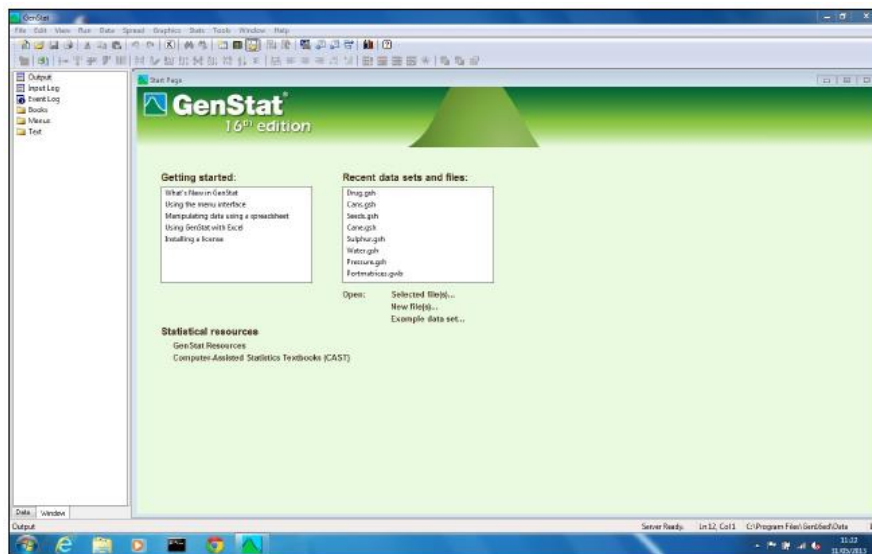


Figure 2a GenStat 16 screen upon start up.

The Start Page allows you to open a data file or view some helpful documentation. For now, close the page by clicking on the red cross at the top right hand corner of the window. (You can stop this window from appearing each time you start GenStat by *unchecking* the 'Show Start page' box on the General tab of the Tools, Options menu). With the Start Page closed, GenStat will look like the following:

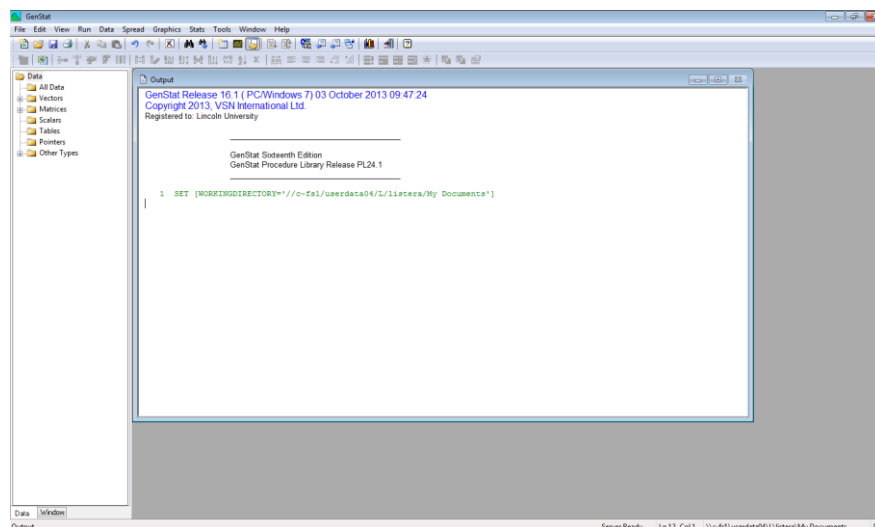


Figure 3b GenStat 16 screen with Start Page closed and left hand panel showing the Data View pane .

3. The windows



On startup, the GenStat client creates three different windows:

The main window is the **Output Window** - this will contain the text output from any analyses. You can edit, print, and save the contents of this window. You can also clear the contents of the Output window by using Edit, Delete from the toolbar. The window can only be closed by ending GenStat.

The **Input Log** contains the text of all the commands GenStat has processed. If you wish to see the commands generated by a particular menu action, you can look in the Input Log window (click the 'Window' tab at the left of the screen, then double-click 'Input Log'). You can clear the contents of the Input Log window by using Edit, Delete from the toolbar.

The **Event Log** window contains a log of events from the GenStat session. It will contain full details of any faults or warnings produced during the session.

In addition to these three windows, you can also open further text windows to type commands into and to edit commands.

An alternative way to move between windows is to use the 'Next Window' button  or the 'Previous Window' button  located on the toolbar.


You can also select a particular window via the Window menu on the toolbar.

Note that all the windows that are currently available are listed in the Window Navigator at the left of the screen (you may need to click on the 'Window' tab to see them).

4. Loading data

There are many ways to load data into GenStat. The following instructions describe two commonly used methods.

1. Importing an Excel spreadsheet

- Click on the 'Excel Import Wizard' button  on the toolbar
- Select the file Plant site results.xls
- A dialog box will open that will list the sheets and any named ranges present in this Excel spreadsheet. Select the one that you require (i.e., Sheet1) and Click [Next >].
- A preview will display the first few rows of the selected sheet (or named range). Click [Next >].
- You can choose which columns of data you wish to load from the spreadsheet. All the columns are already selected. Click [Next >].
- A dialog box will open with several tabs of options. Select the 'Rows' tab, and make sure the 'Remove empty rows from the spreadsheet' option is checked. Click [Next >].
- The 'Select Columns to Convert to Factors' dialog box appears next. This box allows you to select which columns will be factors. In this example, GenStat has determined that the `Site` column is *text* and the `Plant_no` column is a *variate*. To convert `Plant_no` to a *factor*, select it, then press the [Factor] button. If you make a mistake, simply select the column and press the [Leave] button. Click [OK] to import the data.

- The data will now be imported into a GenStat spreadsheet (Figure 4).

Row	Site	Plant_no	Size_cm
1	A	1	5.6
2	A	2	6.5
3	A	3	6.1
4	A	4	5.8
5	A	5	6.2
6	B	1	6.4
7	B	2	5.3
8	B	3	5.7
9	B	4	6
10	B	5	5.1

Figure 4 Successful import of the 'Plant site results.xls' Excel spreadsheet into GenStat.

Whenever you move from the spreadsheet window to another window, GenStat's central data pool will be automatically updated. Try clicking in the Output window and notice the data summary that appears there.

Also notice the Data View pane:

- the 'All Data' folder on the left of the GenStat screen has been updated (click the 'Data' tab if necessary).
- column names imported from Excel have had spaces replaced with underscores '_', and any brackets '(') and full stops '.' have been removed entirely.
- *variates* are identified by a blue 'v' and *factors* by a red '!' in the Data View pane.
- information about each data structure appears when you move the mouse over the name

2. Pasting data from the clipboard

Data from many applications, e.g., Excel spreadsheets, can be easily copied and pasted into GenStat. This method enables you to select a portion of a spreadsheet, rather than the whole sheet.

- First, open the file `cement.xls` in Excel.
- In Excel, select the required data (including the column headings) and copy to the clipboard.
- Open (or switch back to) GenStat.
- Select `Spread > New > From Clipboard` (the `Spread` menu is on the main menu bar.)
- The 'New Spreadsheet from Clipboard' dialog box will appear. Click [OK]. Notice the warning asking if you want a summary of errors in the Output Window. Click [No].
- The 'Select Columns to Convert to Factors' dialog box appears next. Click [OK].

- Notice that the data has **not** imported correctly: The names of the columns in Excel have been mistakenly read as data with the numeric characters at the start of the names being imported as the data values for row 1. Also notice that default labels, C1, C2, C3... have been used for column labels (Figure 5).

Row	C1	C2	C3	C4	C5	C6
1	3	4	2	*	*	
2	7	26	6	60	3.2	78.5
3	1	29	15	52	3.2	74.3
4	11	56	8	20	3.2	104.3
5	11	31	8	47	3.2	87.6

Figure 5 The data from the spreadsheet has been imported incorrectly.

Close the spreadsheet window, answering 'No' when asked if you want to save changes to the spreadsheet and and when asked whether you want to update the data in the server.

Now repeat the import **but this time**:

- In the 'New Spreadsheet from Clipboard' window, look for 'Column names in first row' and choose 'Yes' instead of 'Yes if All labels' (Figure 6). Click [OK].

Figure 6 Force column names to be imported from the first row of the clipboard data.

The data should now import correctly (Figure 7).

Row	%3CaOAl2O3	%3CaSiO2	%4CaAl2O3Fe2O3	%2CaSiO2	gypsum%	Heat
1	7	26	6	60	3.2	78.5
2	1	29	15	52	3.2	74.3
3	11	56	8	20	3.2	104.3
4	11	31	8	47	3.2	87.6
5	7	52	6	33	3.2	95.9
6	11	55	9	22	3.2	109.2

Figure 7 Correctly imported data from clipboard.

Notice that where the first character of a column name is a number, GenStat has added a '%' at the beginning of the column name.

- As before, click in the Output window to transfer the spreadsheet contents to GenStat's central data pool, and to see a brief data summary in the Output window.
- Notice the columns of data now listed in the 'All Data' folder on the left of the GenStat screen:

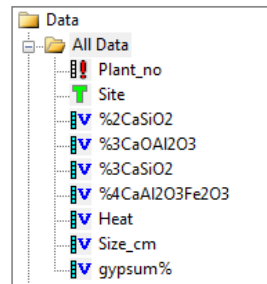


Figure 8 The 'All Data' folder showing the data available in GenStat.

5. Managing data

- To display data in the Output window use the 'Display Data in Output' option of the Data menu.

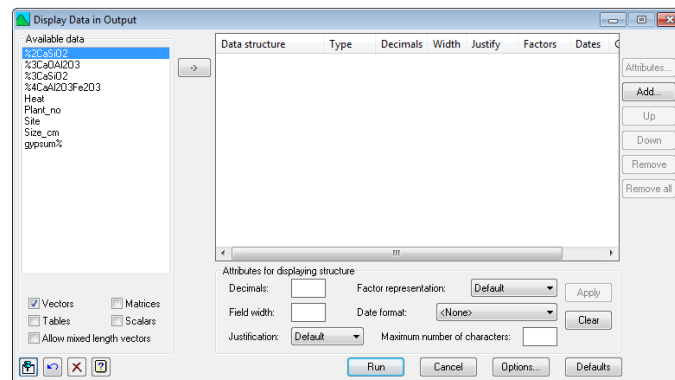


Figure 9 The 'Display Data in Output' window.


- Click on `Plant_no` in the list of available data at the left, then click the arrow to enter it into the right hand window. Notice the 'Available data' box now lists only data that is the same length as `Plant_no`. You can specify attributes for the display of the structure, or just use the default settings.
- Select `Site` and `Size_cm` in the same way.
- Click [Run] to display the selected columns of data in the Output window (Figure 9).

Plant_no	Site	Size_cm
1	A	5.600
2	A	6.500
3	A	6.100
4	A	5.800
5	A	6.200
1	B	6.400
2	B	5.300
3	B	5.700
4	B	6.000
5	B	5.100
1	C	4.300
2	C	5.700
3	C	4.800
4	C	3.800
5	C	4.300
1	D	3.200
2	D	3.900
3	D	4.100
4	D	4.300
5	D	3.700

Figure 9 `Plant_no`, `Site` and `Size_cm` displayed in the Output window.

6. Editing a spreadsheet

1. Changing a data type

- Select the spreadsheet *Plant site results*.
- To change the *Site* column from text to a factor, select *Spread > Column > Attributes/Format...* (or click on the 'Edit Column Attributes' button ).
Leave the default settings and click [Convert...].
- Make sure the *Site* column is selected and choose 'Column Type: Factor' (Figure 10). Click [OK].

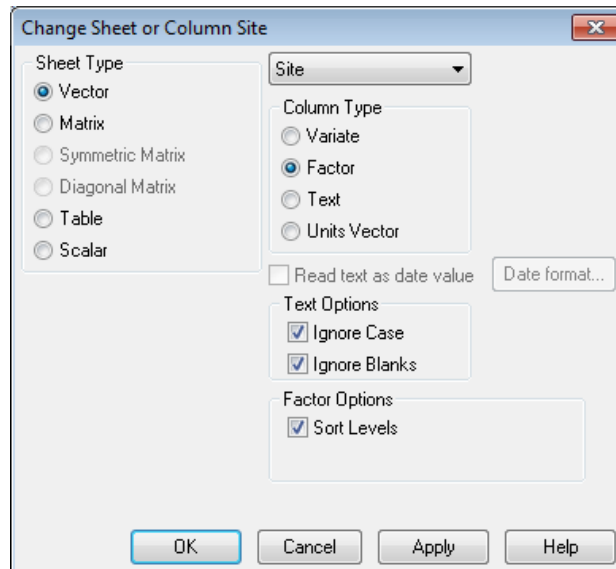


Figure 10 Dialog box for changing the *type* of a column. The text variable *Site* is being changed to a factor.

- Click [OK] again to exit the 'Column Attributes/Format' dialog box. The spreadsheet should now show *Site* with a red '!' alongside the column heading to indicate that it is now a factor.

Row	! Site	! Plant_no	Size_cm
1	A	1	5.6
2	A	2	6.5
3	A	3	6.1
4	A	4	5.8
5	A	5	6.2
6	B	1	6.4
7	B	2	5.3

Figure 10 The text variable *Site* has been changed to a factor.

Also note the short cut method to convert a column: right-click on the column and choose 'Convert to Factor' from the drop-down menu.

2. Renaming a column

Sometimes you will want to rename columns in a spreadsheet. For example, we will rename the column `Plant_no` to `Plant`. These instructions describe two ways to rename columns in a spreadsheet: the first method uses the mouse, and the second method uses the menus.

a. Using the mouse to rename columns.

- Select the spreadsheet `Plant site results`.
- Move the mouse pointer to the left hand edge of the column name you want to change (`Plant_no`), until the pointer changes to a red pencil shape (Figure 11). A small tool-tip that identifies the column may also appear.

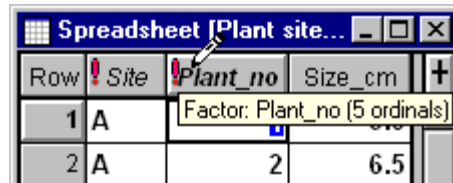


Figure 11 Renaming columns using the mouse. Move the pointer until it changes to a pencil shape.

- Click on the column name.
- A dialog box will appear (Figure 12). Type the new name `Plant` into the 'Name:' box and click [OK].

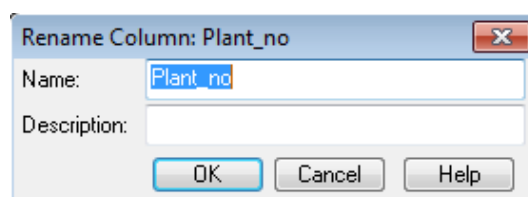


Figure 12 Rename column dialog box. Enter the new name and click [OK].

b. Using the menu to rename columns

- Select the spreadsheet `Plant site results`.
- Select `Spread > Column > Rename...`
- A dialog box will appear (Figure 13). Click on `Size_cm` and replace the name with `Size`, then click [OK].

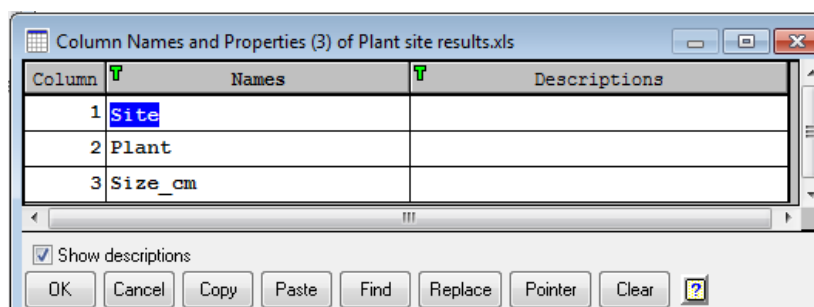


Figure 13 Rename Columns dialog box.

- Notice the list of data now in the 'All Data' folder - renaming the columns in the spreadsheet has created another factor and another variate in the GenStat data pool. If you want to delete the 'old' structures (`Plant_no`, and `Size_cm`), right click and select 'Delete' to remove them.

7. Basic statistics

To obtain some basic statistics

- Use Stats > Summary Statistics > Summary Tables...
- The Summary Tables dialog box will appear
- Select *Size* for the Variate: box and *Site* for the Groups: box (Figure 14):

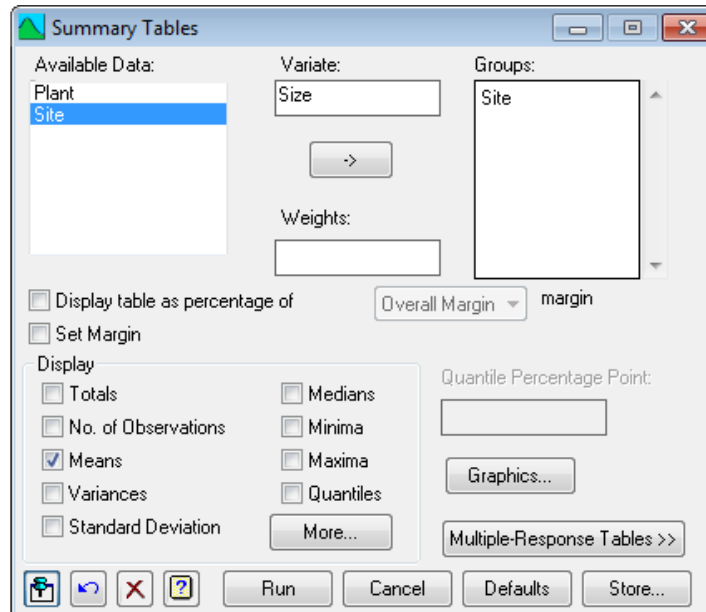


Figure 14 Completed Summary Tables dialog box to produce Mean Size for each Site.

- Click [Run]. The results will appear in the Output window but note that the dialog box remains open.
- Check the 'No. of Observations' box and click [Run] again.
- Check some of the other summary statistics boxes and when the output has displayed, close the dialog box.

8. Drawing graphs

1. Boxplots

The following instructions describe plotting the data from the variate *Size* as boxplots for each *Site*:

- Use Graphics > Create Graph...
- The 'Choose Graph' dialog box will appear. Select 'Single variate and grouping factor' as the Form of the data and 'Boxplot' as the Type of graph required (Figure 15). Click [OK].

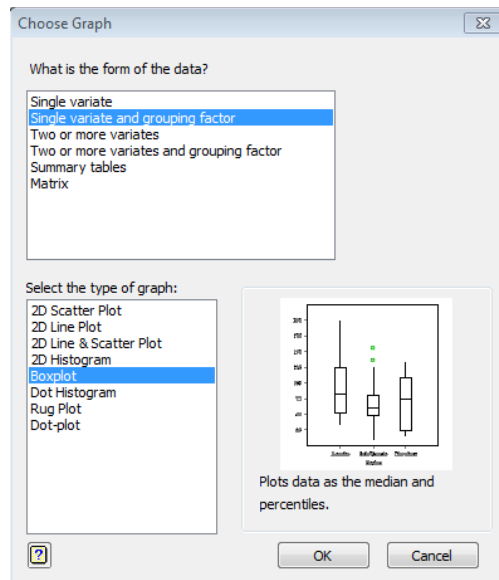


Figure 15 Specifying boxplots for a variate with a grouping factor.

- The 'Boxplot' dialog box will appear (Figure 16). Select 'Variate(s) with single grouping factor', then select *Size* as the 'Data variate(s):' and *Site* as the 'Grouping factor:'.

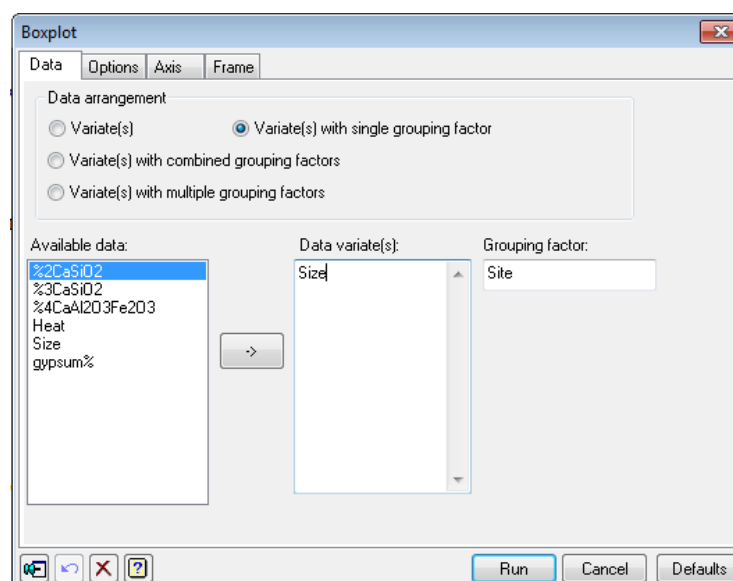


Figure 16 Boxplot dialog box to draw boxplots of the *Size* data for every *Site*.

- Add a Title and labels if you wish – use the ‘Options’ tab to add the Graph Title and the Box Title (the Grouping factor label). Use the ‘Axis’ tab to specify the Axis Title (eg Leaf size (cm)).
- Click [Run] to have the boxplots appear in the GenStat Graphics Viewer (Figure 17).

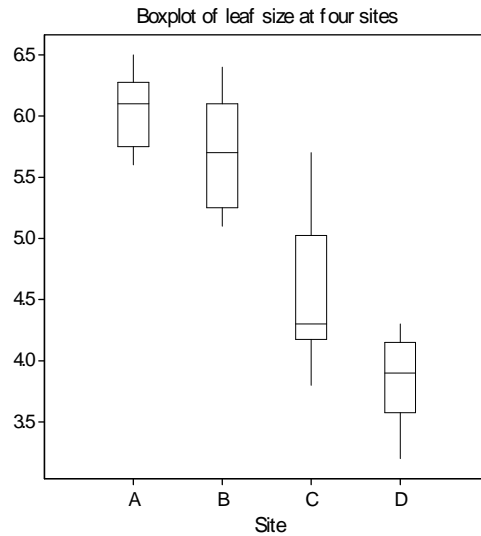


Figure 17 Boxplot of leaf sizes from plants sampled at four sites.

2. Scatter Plot Matrix

A scatter plot matrix is a useful plot for assessing how several variables are associated with each other (perhaps before performing a multiple regression).

- Use Graphics > Create Graph...
- The ‘Choose Graph’ dialog box will appear again. This time, select ‘Two or more variates’ as the Form of the data and ‘Scatter Plot Matrix’ as the Type of graph required (Figure 18). Click [OK].

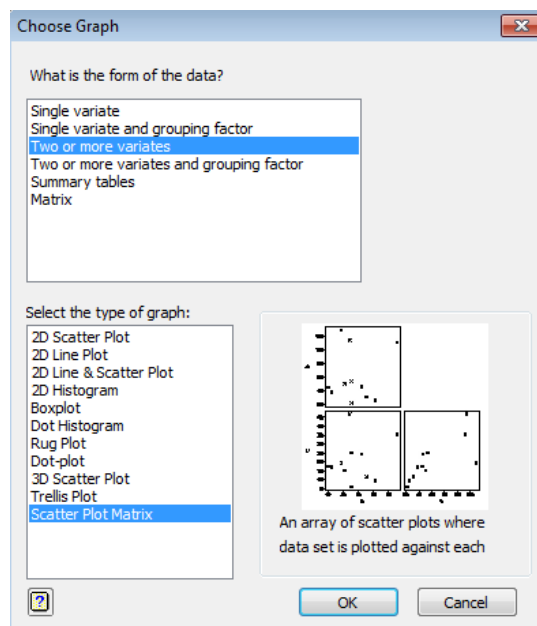


Figure 18 Specifying scatterplots to show the relationship between a number of variates.

- The Scatter Plot Matrix dialog box will appear. Select the cement data variables: %2CaSiO₂, %3CaOAl₂O₃, %3CaSiO₂, %4CaAl₂O₃Fe₂O₃, Heat (Figure 19).

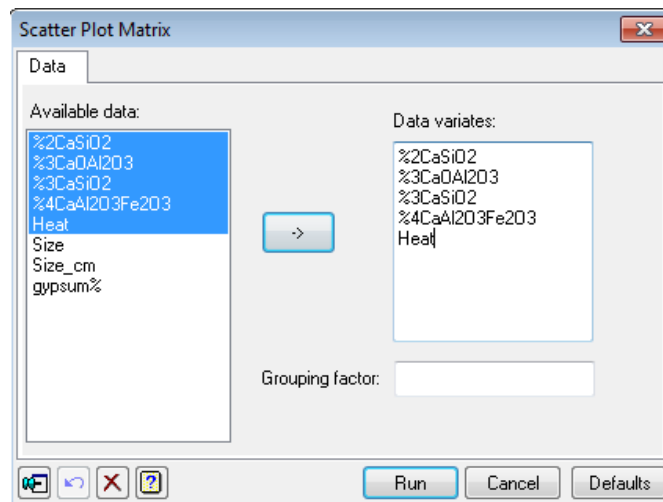


Figure 19 Completed dialog box to plot selected data in a matrix of scatter plots.

- Click [Run] to produce the scatterplot matrix in a GenStat Graphics window (Figure).

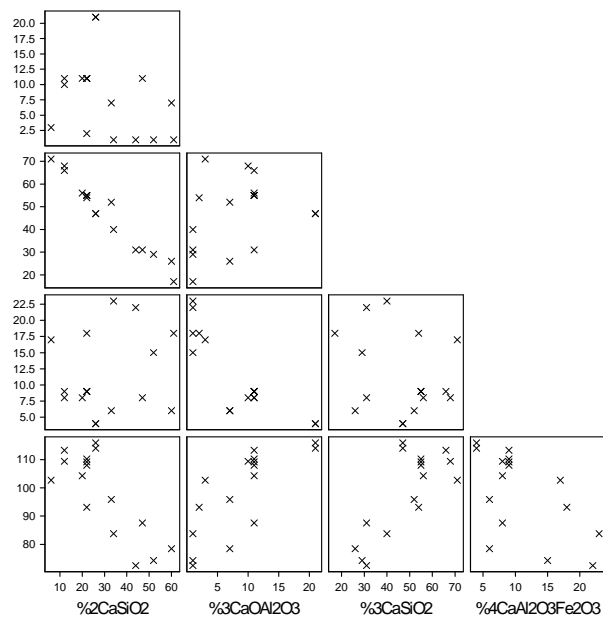




Figure 20 A matrix of scatter plots of the data from cement.xls.

9. Saving your work

1. Saving a GenStat spreadsheet

If you save your data as a GenStat spreadsheet, you can quickly load it back into GenStat in future.

- Make the Plant site results spreadsheet the *active* window (use  or  to move between windows) and click File > Save As...
- Make sure the file will be saved as a GenStat Spreadsheet (*.gsh).
- Navigate to the folder where you want to save the spreadsheet, and edit the File name if you wish (Figure). Click [Save].

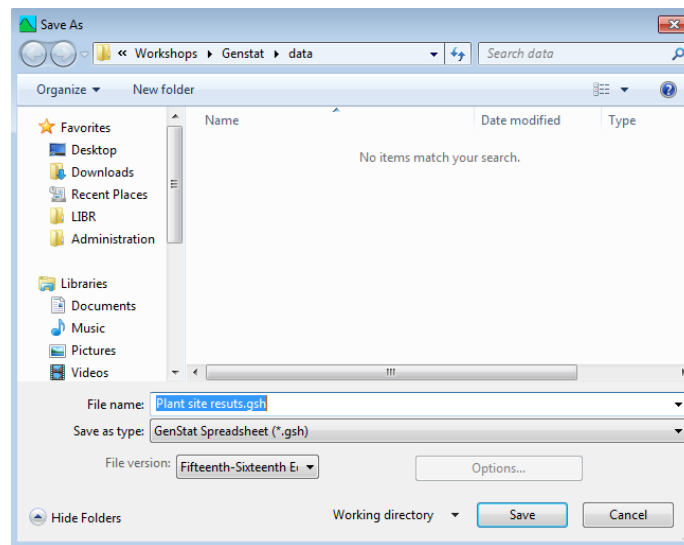




Figure 21 Dialog box for saving a spreadsheet.

- Now save the spreadsheet data that you imported from cement.xls by making *that* spreadsheet active and repeating the process described in the previous section. Remember to check what GenStat suggests as the file name and specify the filename you want before you click [Save].

2. Saving the output window contents

Once you have done an analysis, the results will appear in the output window. You can save the contents of all or part of this window for later editing or printing.

- Make the Output window active. (Use  or  to move between windows.)
- Use File > Save As... (**Error! Reference source not found.**)

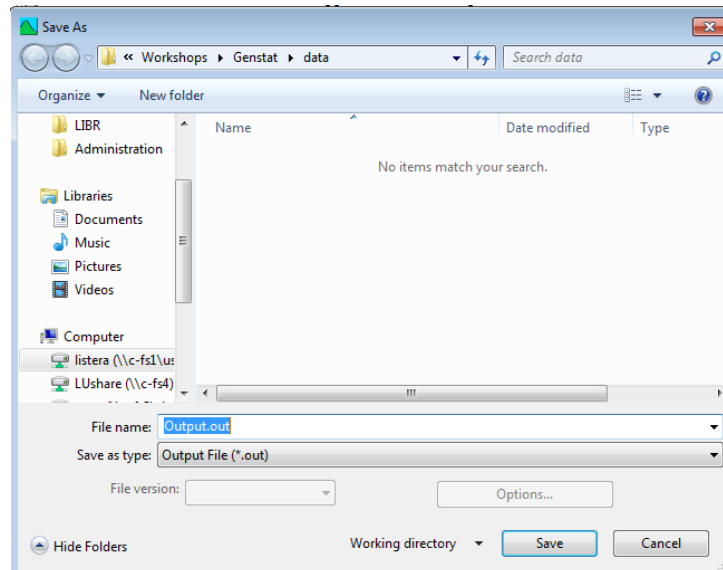


Figure 22 Dialog box for saving contents of the Output window. GenStat uses the default filename of Output.


- Select a folder and filename in the usual way. The default filename is **Output.out** – it is good practice to change the ‘Output’ part to something more meaningful to you. (However, I suggest you leave the file extension as ‘.out’ to remind you that the file is text output). Click [Save].
- The saved text file can be opened by Word or a text editor by using the File > Open menu in that application.

3. Saving a GenStat session

GenStat can save all of the output and data together into one file called a *session*. This is useful because it allows you to stop and save your work then come back later and restart where you left off.

- Use File > Save Session...
- Type in a filename and click [Save]. (Note: it often pays to name session files with a date or number so that you can easily find the latest session to use.)
- To open the session again, simply use File > Open... and select the session filename.

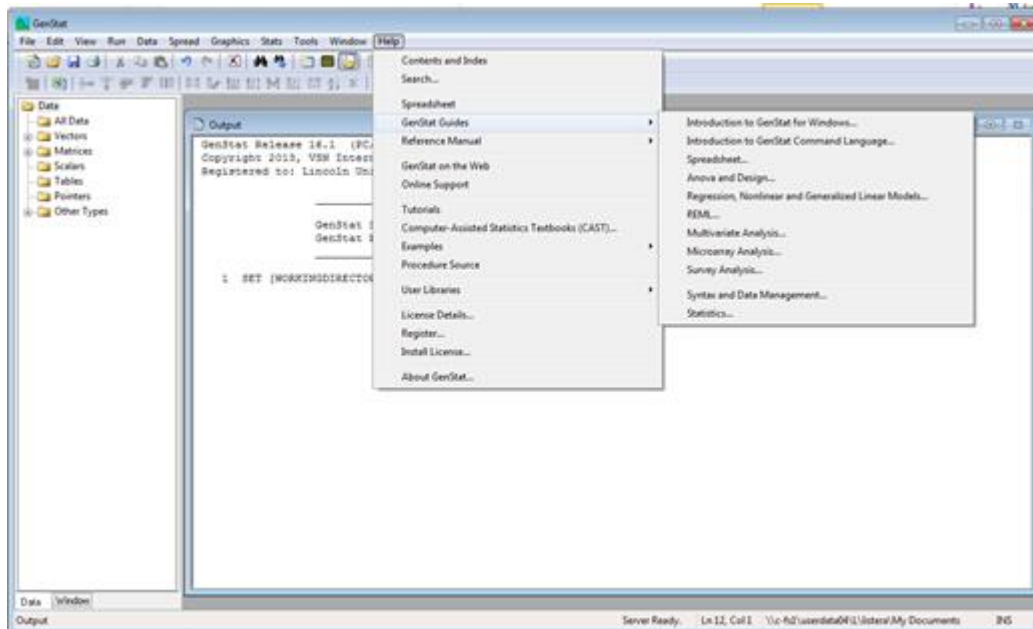
10. Exiting GenStat

- To exit GenStat, use File > Exit (or press the button, )

You may see a number of dialog boxes asking you whether or not you want to save various files or windows contents. Be careful to save anything you haven’t saved already.

11. Further information

- GenStat's Help menu provides access to a number of resources. The 'GenStat Guides', the 'Tutorials' and the 'Examples' will all be helpful for new users.



- New users are advised to look at the GenStat Guide, 'Introduction to GenStat for Windows...'. This is a .pdf booklet that can either be read from within GenStat or saved (or printed) as reference material.
- Help > Tutorials takes the user to VSN International's Resources website where you will find a range of Tutorials.
- You can also make an appointment to see me for assistance. Appointments can be made via the following page: <http://library.lincoln.ac.nz/Learn/Workshops-and-Appointments/>

12. Acknowledgements

Andrew McLachlan (previously of Lincoln University) designed a number of workshops for versions of GenStat available prior to 2007. The workshop presented today is based around one of the original workshops, updated for use with GenStat 16th edition (©2013).

*Alison Lister
Learning Advisor (Mathematics and Statistics)
Library, Teaching and Learning*