

2

Graphical User Interface

2.1 Introduction

S-PLUS has an excellent graphical user interface (GUI) that allows the user to access its functionalities by pointing and clicking with the mouse. This convenient way of working with S-PLUS was previously only available for the Windows platform but has now been added for the UNIX/Linux platform as well.

The advantage of the GUI for the novice of S-PLUS is that you don't have to know the syntax of S-PLUS to get started. All you need is a little familiarity with typical Windows software and a data set in some sort of standard format.

The descriptions in this chapter are, for the most part, based on S-PLUS for Windows. Users of UNIX/Linux will benefit from this chapter, despite it being aimed at the Windows user of S-PLUS, as we will show typical applications and examples and point out major differences between the two operating systems. The chapter will end with a section dedicated solely to the GUI available with UNIX/Linux.

The approach we will take in this chapter is to quickly introduce the S-PLUS system design under Windows, show the briefest of explanations of how it functions, and finish by describing in detail the various tasks that will be needed to complete a data analysis, from data input to printing and saving the results. It is by no means intended to be an extensive or exhaustive exploration of the GUI but merely a way of familiarizing you

with its structure, where to find things, and, most importantly, where and what to try for more detailed options.

2.2 System Overview

When you open S-PLUS under Windows by double-clicking on the S-PLUS icon, you are greeted by a screen layout as shown in Figure 2.1. This may vary slightly according to the version of S-PLUS you are using. The main elements that are visible include the Object Explorer, the Commands window, the menus, and the toolbar. Optionally, a graphics window can be opened. A short description of each of these components is given in the next several subsections.¹

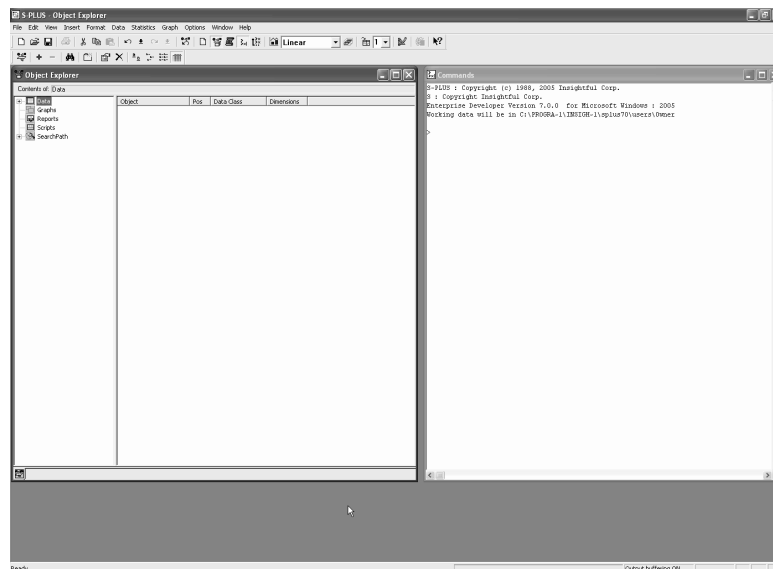


Figure 2.1. The S-PLUS screen and its components: the Object Explorer, the Commands window, and the toolbars.

The general layout of the S-PLUS system is similar to that of many popular Windows systems in that it has pull-down menus at the top and toolbars just below the menus. To use such a system it is useful to be a little familiar with basic point-and-click operations and how to use a mouse.

¹According to the version of S-PLUS being used, some components might not be present or look slightly different.

For those who are not that comfortable with window- and icon-based software, the following subsections provide a crash introduction to the essentials. The pull-down menus across the top are used to group categories of commands or options to be set. Under the **File** menu, we find actions relating to files (**Open**, **Close**, **Import Data**, **Save**) as well as to exiting the system. The toolbars below the menus contain buttons that are convenient shortcuts to commands found through the layers in the menus. Some of the toolbar buttons (e.g., **2D Plots**) open a palette containing a myriad of options to complete your task.

An online help facility is available through the main menu. The typical approach to handling the help system (i.e., search by index or topic) has been used. The big advantage of the help system in S-PLUS is that it includes online manuals. We encourage liberal use of the help system.

2.2.1 Using a Mouse

Using a mouse efficiently is important to get the most out of the system. Clicking once on the left mouse button is usually used to highlight an item in a list (e.g., a file out of a list of files) or to select a menu heading or button from a toolbar. You will not always be able to guess at the function of a toolbar button merely by looking at the icon, but if you are at a loss, simply position the mouse over the button in question and a short text description will appear below it. Double-clicking on the left mouse button is used to select and execute. Examples include double-clicking on a file name to select it and start the function, or on a part of a graph to select and edit it. Clicking once with the right mouse button opens a context menu that changes depending on the item selected.

2.2.2 Object Explorer

The Object Explorer is used to get an overview of what is available on the system, including data, functions, and graphs. It operates in much the same way as the Windows Explorer in that there is a tree-like structure in the left pane and the details are displayed in the right pane. The Object Explorer can be opened either from the menu or from a toolbar button. Not only can it show what objects exist, but with data, for example, the Object Explorer is used to view (browse) it and even edit it. Double-click on the data you want to view or edit, and a spreadsheet will open containing the selected data. Once the spreadsheet is open, the data can be edited.

2.2.3 Commands Window

The Commands window is actually the heart of the S-PLUS system. Every command that is performed via menus and buttons can be issued as a

command directly from the prompt in the Commands window. In addition, there are many functions that can only be run in the Commands window. As examples, the functions of programming, personalized functions, data subsets, and logical operations are only available, or more extensively available, through the use of the Commands window.

2.2.4 Toolbars

The main toolbar contains many familiar commands, including **Open**, **Save**, **Print**, **Copy**, **Paste**, and **Undo**, and is shown in Figure 2.2. In S-PLUS, however, one may also open the Commands window and Object Explorer, open a new graph sheet, open the History log, create 2D and 3D graphs, and so forth.



Figure 2.2. The S-PLUS main toolbar.

An additional feature of the toolbars in S-PLUS is that they are context-sensitive (smart). Open a graph sheet, for example, and extra buttons will appear in the toolbar area that are specific to graph sheets.

2.2.5 Graph Sheets

Graphs are drawn in windows referred to as graph sheets. Components of graphs can be edited and redefined by double-clicking on the component of interest. Labels can be changed, colors modified, axes redefined, and more, all through menus and dialog boxes available simply by clicking in the graph sheet. The **Insert** menu is useful for adding features and components to graphs that already exist in a graph sheet, and the **Format** menu is useful for changing the design of many components of an existing graph.

2.2.6 Script Window

The Script window also consists of two panes, one on top and one on the bottom. The top pane can be used as a development space in which to create or fine-tune a section of code. When the commands in the top pane are run, the output from them appears in the bottom pane. In function, the Script window is similar to the Commands window, the difference being that the former runs commands in segments, whereas the latter is completely interactive and executes commands one at a time as they are input. More details of the layout of the Script window and how it operates are provided in Section 2.16.

2.3 Getting Started with the Interface

If you want to use S-PLUS having had no introduction, use the menus and toolbar buttons. For the most part, you should find them to be self-explanatory. However, we have described a few key functions in a bit more detail to get you on your way.

2.3.1 Importing Data

You probably have your own data that you want to analyze, so the first thing you have to know is how to import it into S-PLUS. There is an **Import Data** facility located in the **File** menu. In the **From File** dialog, you will be asked for the name you want for your data (which may be more than eight characters) and have the usual Windows boxes for specifying the name and location of your data file using the **Browse** button. Pay careful attention to the type of data (**File Format** pull-down menu) that you have and properly define it in the corresponding box. The data file types available are shown in Table 2.1.

Table 2.1. File types available with the Import From File facility.

File	Format		
ASCII	ASCII formatted	dBase	Epi Info
Excel	FoxPro	Gauss	Gauss UNIX
Lotus 1-2-3	Matlab	Minitab	MS Access
MS SQL Server	Paradox	QuattroPro	SAS
SAS Transport	SigmaPlot	S-PLUS	SPSS
SPSS Portable	Stata	Systat	

When the spreadsheet containing newly imported data is closed, the new data automatically appear as a new entry in the Object Explorer.

2.3.2 Graphs

Graphs are created by connecting the data shown in the Object Explorer to a graphical function represented by a button in a palette. In the Object Explorer, the data are displayed with its full name, and after a click on the data object's name, the elements of the selected data set are displayed in the right-hand pane of the window. By clicking on these elements, the set of variables to be plotted can be selected. To select a second and third variable, hold down the CTRL key and click on the element.

Before actually displaying the data graphically, the graph palettes need to be opened if they are not open yet. Open the 2D or 3D graph palette by

using the appropriate icon in the toolbar. The 2D graph palette appears in Figure 2.3 and shows all the types of 2D graphs available.



Figure 2.3. The 2D graph palette.

Once the palette is open, you might want to move the mouse over the different icons slowly. If the mouse stops for a second, the name of the method represented by the button is displayed. Select the plot type by clicking on it. S-PLUS executes the graphical method right away by using the selected data.

Note For creating a graph, the order of selection of variables is important. The variable selected first becomes the x -variable, the second variable selected becomes the y -variable, and if a third variable is selected, it becomes the z -variable. ◀

Once data have been selected, another graphical method can be applied simply by clicking on a different graph button in the palette.

Note If you click on a graphics button and a graph is set up but not drawn, this probably means that the data selected and the graph type chosen are not compatible. Selecting two variables and clicking on a 3D graphics method is an example of such a situation. ◀

Once a basic graph has been created, it can be edited. Double-click on the element to modify, for example, the axis. A window opens and offers the possibility of modifying all the components such as range, tickmarks, color, width, label, and much more.

Using the Annotation palette that appears once a graph sheet is open, further elements, such as text, or graphics symbols, such as rectangles and arrows, can be inserted into the graph.

A graph sheet may be printed using **File - Print Graph Sheet...** or saved using **File - Save**. All that is left is to specify the destination file name and directory.

2.3.3 Data and Statistics

Data can be summarized by using graphical display techniques but also by statistics such as the mean, median, minimum, and maximum values.

There are **Data** and **Statistics** menus in the top menu bar that contain a large set of routines for manipulating and transforming a data set (**Data**) and routines to process the data (**Statistics**).

First, select the statistical method to use and then specify the data set (or vice versa) and the options you want in the menu fields. The method is applied to the data by clicking on the **OK** button, and the results are displayed in a Report window.

In this way, data can be summarized by looking at minimum and maximum values, at quantiles, or at the correlation between variables. Data can be processed by carrying out a *t*-test, a regression model, or any other method, and tick boxes can be ticked or not to choose options such as paired or unpaired or the degree of a polynomial to fit.

2.3.4 Customizing the Toolbars

The toolbars and palettes can be customized to some extent. It is possible to select which palettes show up on the screen. Select **View** from the top menu and click on **Toolbars**. A window opens that lets you choose which toolbars should be shown. Clicking on any of the selection fields immediately pops up or removes the corresponding palette on the desktop. These settings are stored and reused when S-PLUS is started again. Be careful which ones are chosen, remember that the Graph toolbar will open when a graph is created, and so forth.

The shapes of the palettes might not be satisfactory, depending on your screen resolution, personal preferences, and more. Clicking on one of the four edges allows you to change the size of the palettes, which is automatically set such that all icons fit.

Furthermore, the palette buttons can be displayed in palette form or in toolbar form. Click on the palette and drag it to the top, where the primary toolbar is located. Drop it and the palette becomes a horizontal toolbar itself. Dragging the bar away from the top recreates a palette.

Adding New Palettes and Buttons

The palettes and menu buttons are open for extension. You can add menu buttons to an existing palette and create new palettes. You can generate new toolbars via the **View** menu by selecting the **Toolbars** entry and clicking on **New**. This produces a toolbar dialog box that can be used to create a new toolbar.

You can add a new button by selecting the desired toolbar and opening the context menu with the right mouse button. The entry **New Button** opens up a window where you can specify the name of the button, the text to show when the mouse is over the button, and the S-PLUS function to carry out. (You will learn more about functions as you proceed.)

You can modify an existing button by selecting the button and opening its context menu (right mouse click). A property dialog displays the current settings.

2.3.5 Chapters

S-PLUS gives you the ability to save files from different projects into different areas called chapters. Chapters provide an efficient mechanism for project management. All elements open during a session, such as Object Explorer and graph sheet, can be saved to a chapter such that on reopening the chapter, everything is as you left it. Since S-PLUS keeps all data items ever created, we highly recommend the use of chapters as an easy way of organizing your work and switching between project files.

The actual operation of chapters is quite straightforward. Under **File-Chapters** you have the options of **Attach/Create**, **Detach**, and **New Working**.

Create a new chapter by choosing the **Attach/Create** option and specifying the directory where it should be located. There is a space for defining a label for the new chapter that will appear in the search path. Be sure to choose position 1. S-PLUS will create several subdirectories that it needs, including a `.Data` directory where data are stored. The new chapter is completely empty.

We want to create some data in our new chapter so that when we switch between chapters we can verify that they actually perform the way we want them to and keep our projects separate.

Creating Data

- Click on **Data**, then on **Random Numbers...**
- For **Data Set:**, specify `test`
- For **Sample Size:**, specify 50

- Choose **OK**
 - The first entry should appear in the Object Explorer.
- Click on **Data**, then on **Transform**
- For **Data Set:**, specify **test** (chosen by default)
- For **Target Column:**, specify **Square**
- For **Variable:**, specify **Sample**
- For **Function:**, specify **X²**
- Click on **Add**
 - **Sample²** appears in **Expression:**
- Choose **OK**
 - Can see that the **test** data sheet has been opened and contains the variables **Sample** (the original Gaussian distributed data) and **Square** (the square of the variable **Sample**).

The new data set appears in the right pane of the Object Explorer. We can switch back to the default chapter by choosing **File-Chapters-New Working Chapter...** and choosing the default user directory. “New” refers to the fact that you are defining which of the existing chapters to use, not that you are creating a new one. Does the newly created data set (**test**) appear in the Object Explorer of the default chapter? Switch back to the new chapter. Is **test** the only data set that appears in the Object Explorer here?

Note The **New Working Chapter** option automatically detaches the old chapter as it attaches the new one. However, you do not have to detach the old chapter when accessing a new one. You could simply attach the new chapter and assign it to position 1 (**File-Chapters-Attach/Create**), moving the old chapter further down the search path. This action has two implications. Data sets from a nonactive chapter cannot be seen through the Object Explorer but can be accessed. If a data set in the active chapter has the same name as one in a nonactive chapter, the one in the active chapter will be accessed. ◁

In summary, the chapter feature is a convenient way to help keep yourself organized. We highly recommend that you begin using it right from the start. For practice with chapters, try creating one for each “chapter” of the book and keeping chapter-specific data sets in the respective S-PLUS chapter.

2.4 Detailed Use of the GUI Interface

The previous section provided a quick introduction to some of the basic concepts involved with using the graphical user interface (GUI). Brief descriptions of how to perform several tasks were given, but without much in the way of specifics. The details of how to understand and work with the GUI are given throughout the following sections.

We cannot emphasize enough that to really profit from this book, you need to be running S-PLUS and following the examples while reading the book. Only by trying something on your own will you really learn it to the extent that you could do it on your own. We have left a lot of tasks to be done by the reader (you!) so as to build on the concepts immediately.

2.5 Object Explorer

The Object Explorer plays the same role in S-PLUS as the Explorer does in Windows in that the “contents” are displayed in a tree-like diagram on the left-hand side of the screen and the details on the right-hand side. Similar or related objects can be stored in folders, giving structure to your work. When an object is put into a folder, a link (shortcut) is created such that the actual location of the object is not changed, just its virtual location. Folders can be dragged and dropped between Object Explorers, between Explorer pages within an Object Explorer, or onto the toolbar as a button.

Objects can be classified into three types: computational engine, interface, and document. Computational engine objects include data frames, matrices, lists, functions, and others, and are objects containing data or the functions that are used to process data. Interface objects relate to communication with the system and include search paths, menu items, toolbars, dialogs, and the like. Document objects refer to output from the system and include graph sheets, Reports, Scripts, and the like.

All of the object types mentioned can be saved in folders in the Object Explorer. Folders can be organized according to object type, data set, project, model class, or in any way that seems to be helpful.

Open the Object Explorer, if it is not already open, by clicking the Object Explorer button on the standard toolbar. It should look like the one in Figure 2.4.

If you have not yet worked with S-PLUS, the Object Explorer will be empty, and if you have, you will see different object types in the left pane and their details in the right pane.

We want you to follow an example and to get practice exporting a data set, so we will take a built-in S-PLUS data set, export it, and then import it before we begin analyzing it. First, we have to use the Object Explorer to find the data set.



Figure 2.4. The Object Explorer toolbar.

S-PLUS uses a system of “positions” that are searched sequentially until the desired object is found. A message that an object has not been found is only issued once all the positions have been searched. Position number 1 is the current working directory. By default, the Object Explorer is used to show the objects in the current working directory. Other positions are used to store general functions, statistical functions, other types of data sets, and the like. The data we want to use are a few positions down in the list because the built-in data sets are not meant to be worked on in a working directory but merely read from a safer location.

The search path (written `SearchPath` in S-PLUS) lists all the positions currently available to S-PLUS. By examining the search path, we can find the built-in data sets, locate the one we want, and write it to the hard disk. The search path is conveniently located in the left pane of the Object Explorer. By default, the first directory shown in the `SearchPath` is the one that can be written to, the others are read only.

S-Plus Data Sets

- Expand the search path list by clicking on the **+** next to the `SearchPath` icon in the left pane of the Object Explorer
 - The S-PLUS data sets are located in the ***data*** directory (see Figure 2.5).
- Expand the list of data sets by clicking on the **+** next to the ***data*** icon
- Scroll down to find the Puromycin data set

Double-clicking on Puromycin runs the application, which, in this case, opens a spreadsheet-like editor that can be used to view and edit the data. By single-clicking on Puromycin in the left-hand pane of the Object Explorer, a list of its variables is shown in the right-hand pane. The Puromycin data set contains the three variables ***conc***=concentration, ***vel***=velocity, and ***state***, as shown in Figure 2.6.

2.6 Help

At some point, you will want to use (or have to use) the help system in S-PLUS, and we have reached a point where it is possible to explain this

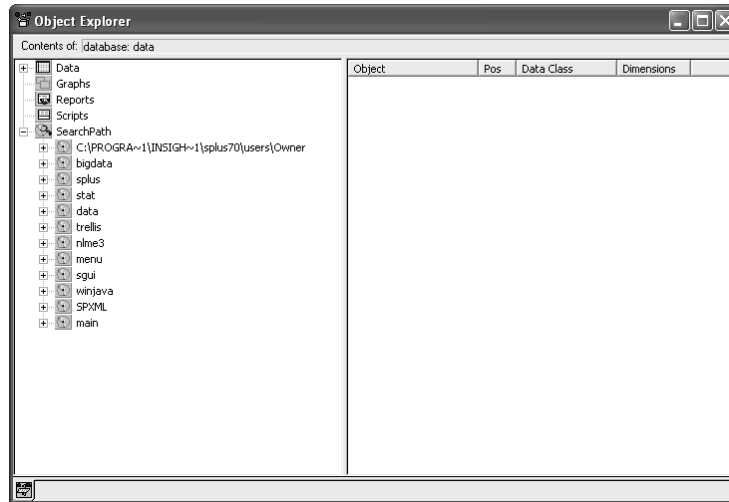


Figure 2.5. The SearchPath.

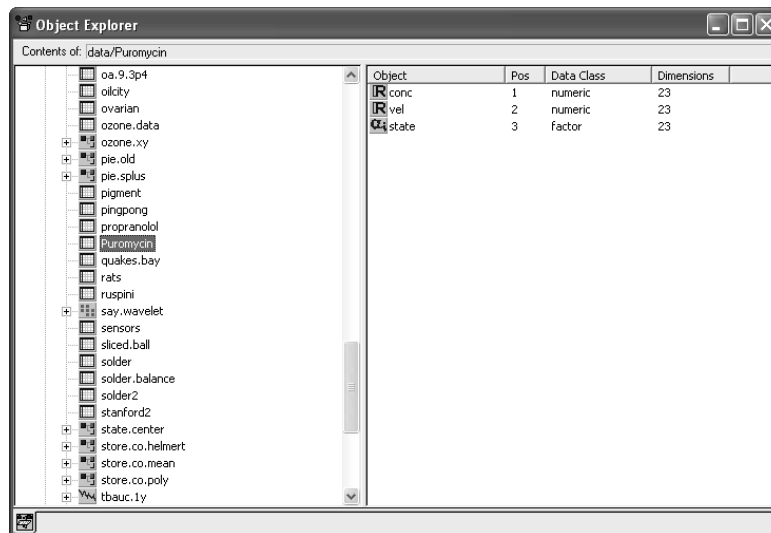


Figure 2.6. Details of the Puromycin data set.

simple procedure. Suppose we want to export the Puromycin data but don't know how to start. Simply invoke the help system to find out how.

Using the Help System – Exporting Data

- Click on the **Help** main menu
- Choose **Available Help-S-Plus Help**
- Choose the **Contents** tab
 - Expand the section on **Importing and Exporting Data**
 - Choose **Exporting Data Sets**
- Or, choose the **Index** tab
 - Type in “export”
 - Click on **Exporting Data Sets**
 - Choose **Display** at bottom
- Or, choose the **Search** tab
 - Type in “export”
 - Click on **List Topics**
 - Click **Export to File** and choose **Display**

Using the Help System – Puromycin Data Details

- Click on the **Help** main menu
- Choose **Available Help-Language Reference**
- Choose the **Index** tab
- Type “pur”
 - Puromycin is highlighted
- Choose **Display** at bottom
 - Window explaining the Puromycin data set appears

The help system can be used in the same way to find information on (almost) any other topic. It is helpful to note that details on the built-in S-PLUS data sets as well as on the Command window language are found in the **Language Reference** section of the help menu.

2.7 Data Export

As promised, we’re going to export the Puromycin data, import it again, and then work with it. The entire process is menu-driven.

Exporting Data

- Click on Puromycin in the left-hand pane of the Object Explorer
- Click on the **File** main menu
- Choose **Export Data-To File ...**
- Use the **Browse** button to specify a location for the file, if desired
- Specify the file type using the **Files of type:** pull-down menu
 - We chose **SAS Version 7/8/9 - Windows (sas7bdat; sd7)** and suggest that you do as well (you don't need to have SAS installed on your computer).
 - Puromycin is already highlighted in **From Data Frame:** (see Figure 2.7).

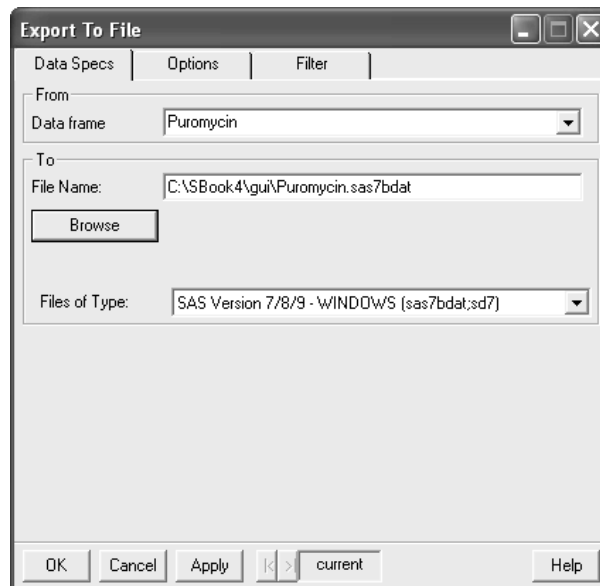


Figure 2.7. Data export.

- The file extension and name are automatically updated to reflect the file type choice.
- Choose **OK**

The file should now be created with the name (Puromycin) and type (.sas7bdat) that we specified, residing in the directory of our choice. Use the Windows Explorer to verify that the desired data set has been created. Save the same data in different formats and open them with the appropriate program; for example, use Excel for *.xls files.

2.8 Working Directory

When working on several different projects, it is often convenient to set up a different “directory” for each separate study on which you are working. Directories can be set up in two different ways: as shown in Section 2.3.5, or by defining multiple S-PLUS icons on the desktop. The advantage of the former approach is that you never have to leave S-PLUS to switch between project-specific directories, whereas the advantage of the latter approach is that the danger of inadvertently working in the wrong directory is minimized.

The method of defining and switching between chapters is easy, flexible, and orderly (see p. 16), but, as people work in different manners, the method of defining a working directory and creating an S-PLUS icon will be shown here.

Defining a Working Directory

Create a copy of the S-PLUS shortcut icon as follows:

- Right-click on the **S-Plus** icon on the Windows desktop
- Choose **Copy**
- Right-click on open space on the Windows desktop
- Choose **Paste**
- Right-click on the new icon and choose **Rename** to give it a new name
- Right-click on the new icon and choose **Properties**
- Choose the **Shortcut** tab
 - **Target:** is highlighted and shows the command for running S-PLUS.
- After the existing text that is already in the box for **Target:**, add **S_PROJ=pathname**
 - Use the path where you saved the Puromycin data.
 - The *pathname* must be enclosed with double quotes if there is a space in it.
 - The addition of **/MULTIPLEINSTANCES** will allow this session to run concurrently with other sessions, each using different working directories (chapters)
- **Start in:** should be blank (see Figure 2.8)
- Choose **OK**

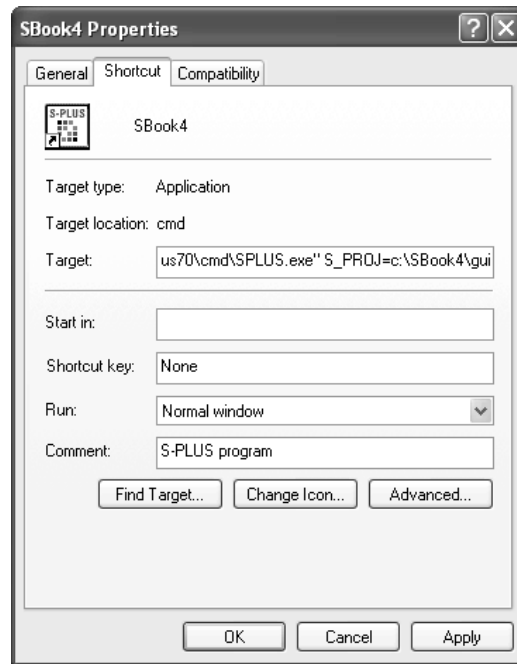


Figure 2.8. Defining a working directory.

You can now start S-PLUS by double-clicking on the new icon, which will have the effect of working in the directory specified in the path name. Open the Object Explorer and verify that the new S-PLUS icon is using the database (empty) defined by the new path name (we did this earlier).

2.9 Data Import

After this preparation, we are finally ready to look at how to read a data set into S-PLUS. The procedure is essentially the same as with exporting data and works in the following manner.

Importing Data

- Choose **File** from the main menu
- Choose **Import Data**
- Choose **From File ...**
- Use the **Browse** button to locate the Puromycin data set (in SAS format) that we created earlier and highlight it by clicking on it (and choose **OK** if necessary)

- Check that **File Format** appears as **SAS Version 7/8/9 (sas7bdat)**
- Change the **To Data Set** name to **puro** to distinguish it from the source
- Click on **Update Preview** to see a preview of the data that is about to be imported
- Options can be chosen to change desired properties as in Figure 2.9
- Choose **OK**

As soon as **OK** has been chosen, a Data window should open containing the data set displayed in typical spreadsheet fashion. It contains 23 rows and 3 columns (variables), just like the original data set.

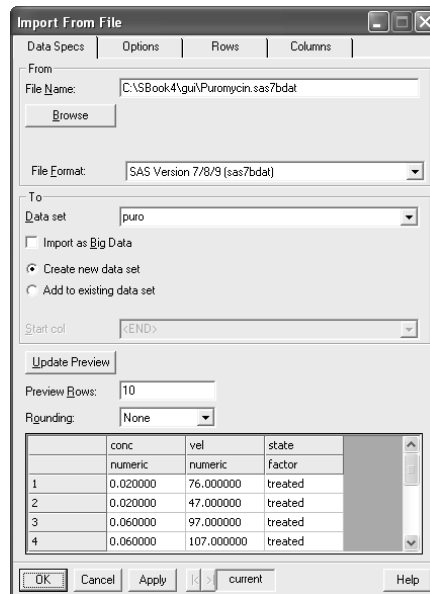


Figure 2.9. Data import window.

As a reminder, you should always check the data to make sure that there are no unpleasant surprises.

Notice that the **puro** data set is currently sorted by **state** and then by **conc**. It would make sense to have the data sorted by **vel** as well. The following steps can be used to sort a data set.

Sorting Data

- Click on the Data sheet to be sorted

- Choose **Data** from the main menu
- Choose **Restructure**
- Choose **Sort ...**
- Using the pull-down menu for **Sort By Columns:**
 - Click on **state**
 - CTRL-click on **conc**
 - CTRL-click on **vel** (see Figure 2.10)
- Choose **OK**
- Close **Data window**

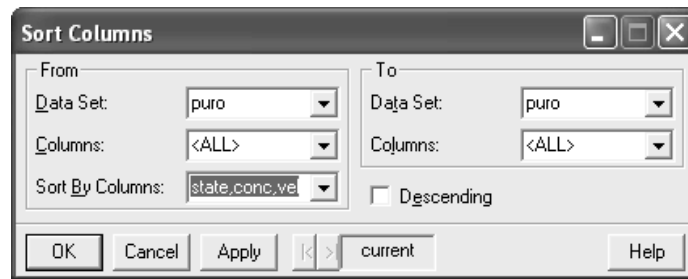


Figure 2.10. Sorting a data frame.

See Figure 2.11 for the resulting sorted data set. Note that the first 2 rows both have **state**=treated and **conc**=0.02 but that now the first row has a smaller **vel** (=47.00) than the second row (**vel**=76.00). In fact, the original order of the rows appears in a column just to the left of the data, reflecting the fact that these two rows have switched positions.

		1	2	3
		conc	vel	state
1	2	0.02	47.00	treated
2	1	0.02	76.00	treated
3	3	0.06	97.00	treated
4	4	0.06	107.00	treated

Figure 2.11. Sorted Puromycin data.

S-PLUS refers to this data set as a “data frame.” Data frame is the term used by S-PLUS to denote a rectangular collection of data, possibly including variables with different data types (e.g., character and integer). Many data sets typically encountered are actually data frames.

2.10 Data Summaries

Now that we have a data frame in the Object Explorer, we want to “look” at the data in some sense. One of the easiest ways of getting a sense for your data is to generate simple summary statistics (mean, standard deviation, etc.). The following command box shows how this is done.

Generating Summary Statistics

- Click on **puro** in the left-hand pane of Object Explorer to select it
- Choose **Statistics** from the main menu
- Choose **Data Summaries**
 - Options are **Summary Statistics**, **Crosstabulations**, and **Correlations**.
- Choose **Summary Statistics ...**
 - The Summary Statistics dialog opens to a data tab.
 - **puro** should be chosen as the default data frame (see Figure 2.12).
- Choose **All Variables** (default)
- Choose **No Grouping Variables** (default)
 - Can save results by supplying a name using **Save As**.
- Choose **Statistics** tab at the top of dialog window.
 - Can choose specific statistics to be calculated (see Figure 2.13).
- Choose **OK**

A **Report** window is opened (Report1) containing the count of each level of the variable **state** and the summary statistics for the continuous variables **conc** and **vel**, as shown in Figure 2.14.

We see, for example, that there are 12 treated and 11 untreated patients and that the mean **conc** is 0.3122. The contents of the **Report** window can be saved into a file by choosing **File-Save As ...** and specifying a path and file name. The default file type is .rtf (rich text format), which can be opened by any standard editor or from within S-PLUS.

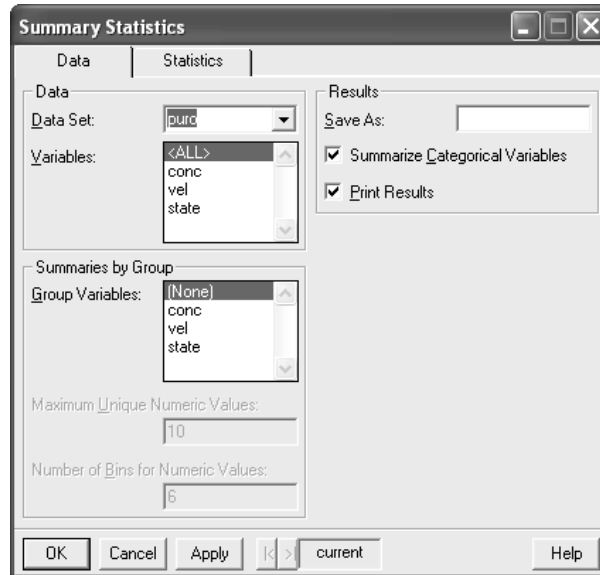


Figure 2.12. Summary Statistics data dialog.

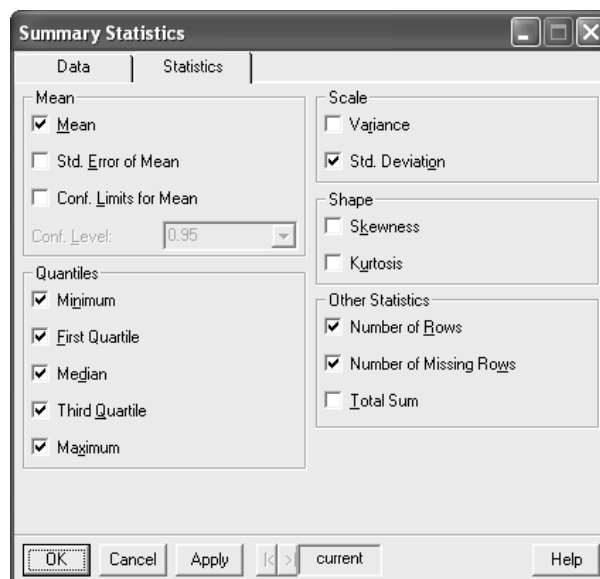


Figure 2.13. Summary Statistics dialog.

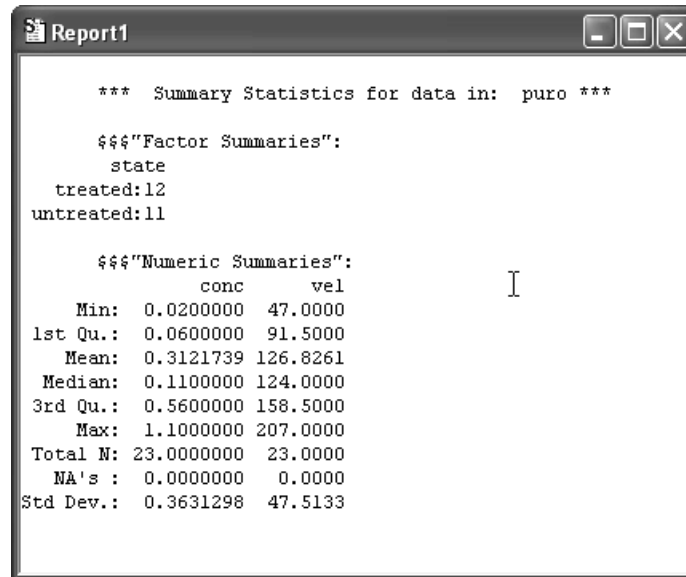


Figure 2.14. Summary Statistics Report window.

2.11 Graphs

The problem with data summaries is that you don't actually "see" the data itself. For a more visual summary of our data, we need to construct a graph. There are actually several different ways of constructing a graph in S-PLUS using the GUI (menu buttons), but perhaps the easiest and most flexible method is to use the palettes. The palettes exist only with S-PLUS for Windows, but the UNIX/Linux user can do the same things using the main menu options.

Creating a Graph Using a Palette

- Open the Object Explorer if it is not already open
- Expand the list in the left-hand pane of the Object Explorer such that the three variables of **puro** are visible in the right-hand pane
- Click on **conc** to highlight it
- CTRL-click on **vel** so that both variables are highlighted
 - The order in which the variables are chosen (clicked) is important. The first will be used for the *x*-(horizontal) axis and the second for the *y*-(vertical) axis.
- Click on the 2D palette button on the main toolbar

- Click on the button for scatterplot (upper left)
 - A new graph sheet will open, containing a scatterplot of the two variables **conc** and **vel**.
 - If you don't like the type of graph that you have chosen, simply click on a different one (e.g., line scatter, loess) and a new graph sheet will be opened with the same data but using a different plot type. Be adventurous and try several.

Although the default graph looks quite good, we'll wait until after we explain some editing techniques before actually showing it. UNIX/Linux users don't have the convenience of editing the graph after the fact. Instead, the details of the appearance of the graph have to be specified upfront.

This leads us to the topic of how to edit a graph once it has been created. In particular, we are going to change the color and symbol for the points on the scatterplot so that they will be more visible.

Changing Color and Symbol of Points

- Click on any point in the graph
 - A green knob will appear in the lower left of the graph.
- Double-click on the green knob
 - A dialog window named **Line/Scatter Plot [1]** will appear, as in Figure 2.15.
 - The first tab for the window shows the **Data to Plot** panel. The data that are used in the plot could be redefined here.
- Click on the **Symbol** tab
- Using the pull-down menu, change the **Style:** to • **Circle, Solid**
- Using the pull-down menu, change the **Color:** to **Black** (see Figure 2.15)
- Choose **OK**
 - The symbol has changed both style and color.

Making the graph look “nicer” (depending on individual tastes) is usually of interest, particularly when it comes to labeling different parts of the graph. We will investigate how to change the axis labels, add a title, and add general text to the graph.

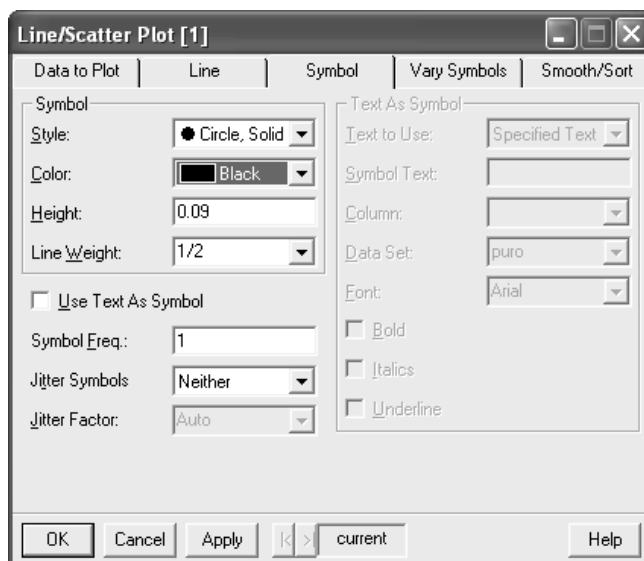


Figure 2.15. 2D Plot dialog (symbol).

Changing Axis Labels

- Click on the default x -axis label (**conc**)
 - It is now surrounded by green knobs.
- Choose one of the following methods:
 - (A) One more click allows you to change the label by simply typing over the default (Default is automatically taken as the column name and is denoted '@Auto').
 - (B1) Right-click on the default label (dialog appears).
 - (B2) Choose **Edit In-place...** (see Figure 2.16).
- Change the text to “Concentration”
- Click outside of the box to finish without hitting RETURN

The label has now been changed and the text is better now; however, the font is rather small. We can easily change this.

Changing the Font

- Right-click on axis label text
- Choose **Font...**

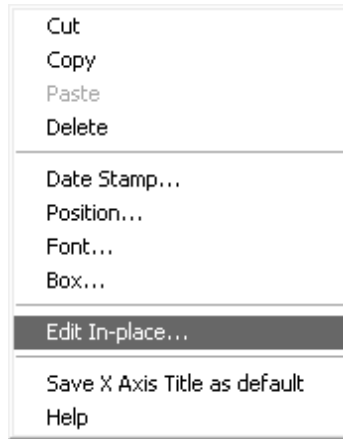


Figure 2.16. Axis dialog box.

- Define font size by varying **Size:**
 - Try 20 (see Figure 2.17).
- Choose **OK**

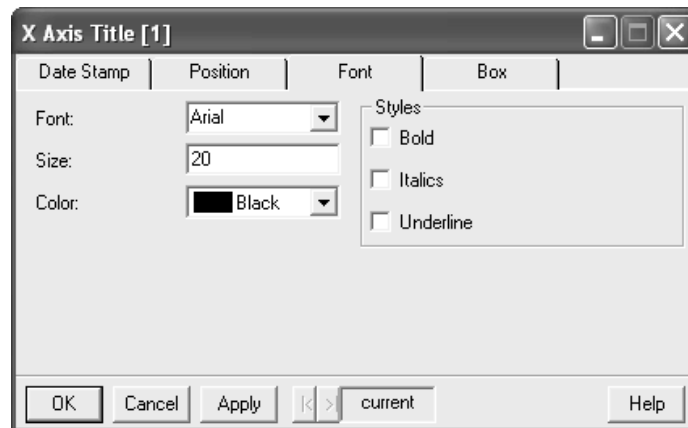


Figure 2.17. Axis font dialog box.

We have now changed the label and font size for the x -axis and leave it up to you to repeat the procedure for the y -axis.

Adding a title to a graph is most easily done using **Insert** from the main menu. The exact usage is explained below.

Adding a Title

- Click on **Insert** in the main menu
- Click on **Titles**
- Choose **Main...**
- Type in the desired title (Use 'Scatterplot')
- Click outside the title box to end
- Change the font as desired (see Changing the Font)
- Change the box as desired

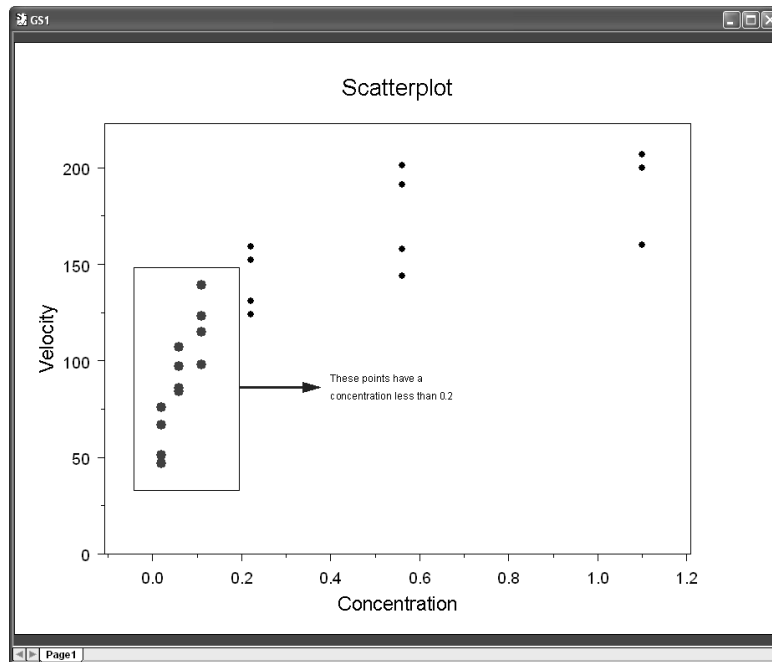


Figure 2.18. Annotated scatterplot.

Graphs in S-PLUS are designed to be interactive in more than one sense; that is, we created the graph to learn something about the data and want to get information out of it. One of the nicest features about S-PLUS is that the graphs are not static; they are not merely nice pictures to look at and study. Try positioning the cursor over a point and holding it there for a second. What happens? A small box appears that contains three pieces of information: the index number and the x and y coordinates. The index

refers to the number of the entry in the data sheet. Choose some points and look at their values and indices.

There are two very useful buttons on the graph sheet toolbar that open the Graph Tools palette and the Annotation palette. By holding the cursor over the individual buttons in the palettes, you can see what their functions are. Explore the possibilities open to you with these and other buttons. We created the graph in Figure 2.18 with just a few clicks of the mouse.

Now that the graph looks exactly the way we want, we're ready to print it and save it for possible future use.

Printing a Graph

- Click anywhere on the **graph sheet** to make it the active window
- Choose **File** from the main menu
- Choose **Print Graph Sheet...**
 - Make sure the settings for the printer are correct.
- Choose **OK**

Saving a Graph (Sheet)

- Click anywhere on the **graph sheet** to make it the active window
- Choose **File** from the main menu
- Choose **Save As...**
- Specify the location and file name for the graph (see Figure 2.19)
- Choose **Save**

We have just saved the graph into a file, so we should be able to read the graph back into S-PLUS. A file extension was not explicitly added to the file name, but the association was established with Windows so that the file is an S-PLUS graph sheet. We leave it up to you to open the graph from the file that you just created.

Another useful task that is often done is to create a graph in S-PLUS for use in another application (e.g., MS Word). There are two ways of exporting a graph from S-PLUS: the direct method and the indirect method. The quality is better with the indirect method but takes a couple of extra steps.

Exporting a Graph – Direct Method

- Open a graph sheet and make sure it is the active window

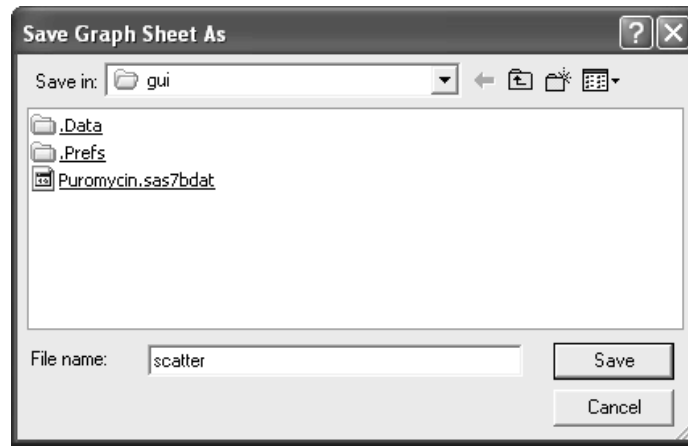


Figure 2.19. Saving a graph sheet.

- Press CTRL-C (or Edit-Copy or Send Graph to Other App button from main toolbar) to Copy
- Move to MS Word (or other Windows application)
- Press CTRL-V (or Edit-Paste) to Paste
 - The Graph is now in MS Word.

Exporting a Graph – Indirect Method

- Open a graph sheet and make sure it is the active window
- Choose **File** from the main menu
- Choose **Export Graph...**
 - From this window, you can choose from many different file types. MS Word, for example, can easily read a Windows Metafile (*.WMF).
- Specify the file type, file name, and location
- Choose **Export**
 - A Windows Metafile can be inserted into MS Word using **Insert-Picture-From File** and then specifying the file name and location.

According to what software you have available on your system, you should try saving a graph using different file types and importing it into the appropriate package.

2.12 Trellis Graphs

Trellis graphs are a convenient way of exploring relationships in data. They work by graphing one or more variables while conditioning on one or more others. Panels are constructed for each level of the conditioning set, and a plot of the variables of interest is made in each panel. An example of their use might be to plot height and weight, but to have one panel for males and one for females. In this manner, it is then easy to see if the relationship between height and weight is the same between the sexes. We will now show how to make a Trellis graph, although we have devoted an entire chapter (Chapter 6) to it later.

Trellis Graphs – Drag-and-Drop (Windows only)

- Create a scatterplot from the Puromycin data as we have done before
- Left-click on **state** in the right-hand pane of the Object Explorer and keep the mouse key depressed
- Drag the mouse to the upper portion of the graph sheet until a dashed rectangular box appears, as in Figure 2.20

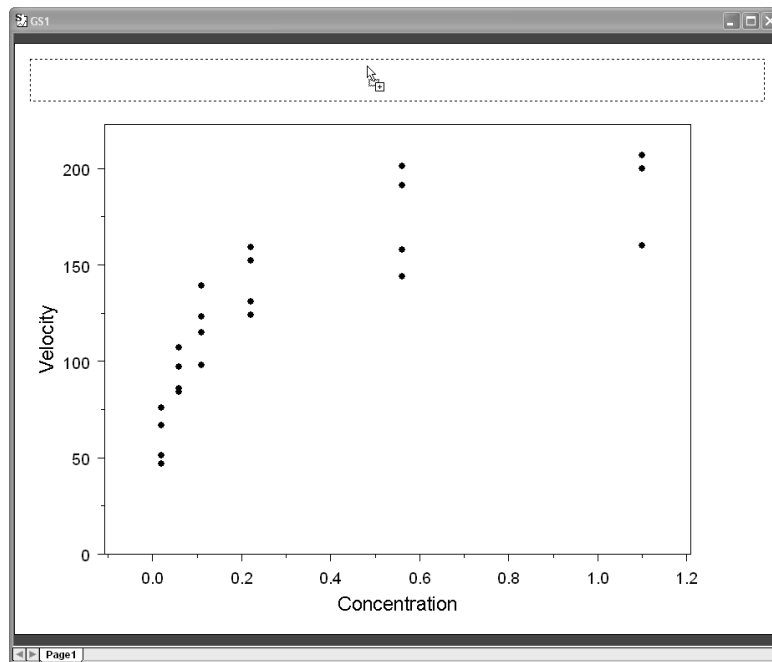


Figure 2.20. Trellis graphs – drag-and-drop.

- Release the mouse
- The graph is automatically transformed into a Trellis graph, as in Figure 2.21

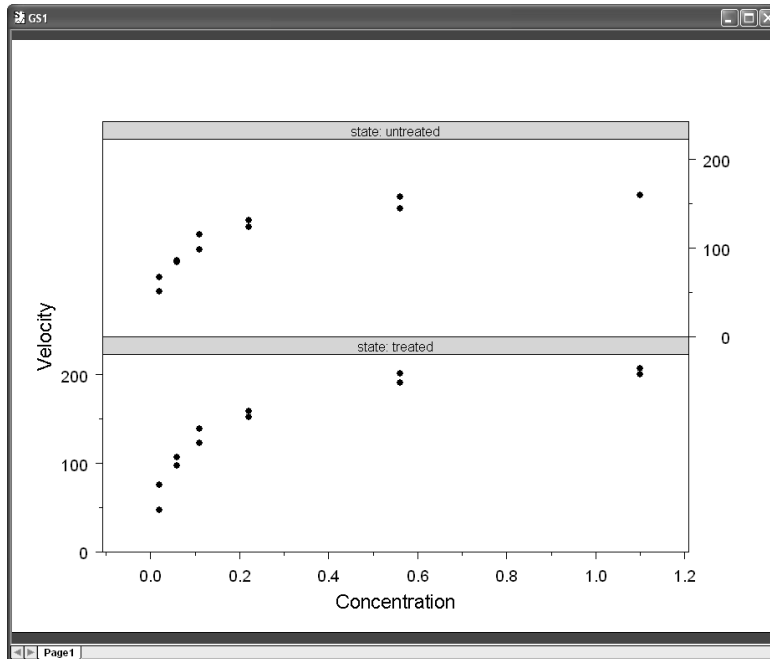


Figure 2.21. A Trellis graph conditioning on state.

There is a small trick to producing Trellis graphs using the menu; we will show that approach separately.

Trellis Graphs – Using the Menu

- In the Object Explorer, click on the variable to be used on the x -axis (**conc**). In the Object Explorer, CTRL-click on the variable to be used on the y -axis (**vel**). In the Object Explorer, CTRL-click on the conditioning variable (**state**).
- Click on **Graph-2D-Plot** in the main menu
- An **Insert Graph** dialog appears
- Change the axes type to **Panel** (the default is **Linear**)
- Proceed as before to create a graph as in Figure 2.21

The only difference in the final graphs between the two approaches is that the same symbol is used in all panels with the drag-and-drop method, whereas each panel gets a different color/symbol using the menu approach.

The resulting message that we can draw from the Trellis graph is that the relationship between concentration and velocity is the same for *state*=treated as it is for *state*=untreated.

2.13 Linear Regression

We have already looked at the summary statistics for the variables in our data set and at a scatterplot of concentration versus velocity. We are now ready to try to fit a linear regression model to our data.

Performing Linear Regression

- In Object Explorer, click on data frame **puro**
- Choose **Statistics**
- Choose **Regression**
- Choose **Linear . . .**
 - Puro is the default data frame.
 - Tabs for specifying Model, Results, Plot, and Predict (see Figure 2.22)
- Note that a default formula for the regression has been chosen, Delete the formula as we will create a new one.
- Choose **Create Formula**
- In the Variables section, click on **vel**
- In **Add** option, choose **Response**
 - Velocity will be the response (*y*) variable.
- Click on **conc**
- In **Add** option, choose **Main Effect: (+)**
 - Concentration will be the predictor (*x*) variable.
 - Can transform variables, add quadratics, and so forth.
 - In the Remove section, can tick box for removing *y*-intercept (see Figure 2.23)
- Choose **OK**
- Try the following options:
- In the Save Model Object section, specify a name for storing the regression object (use the name **my.reg**).

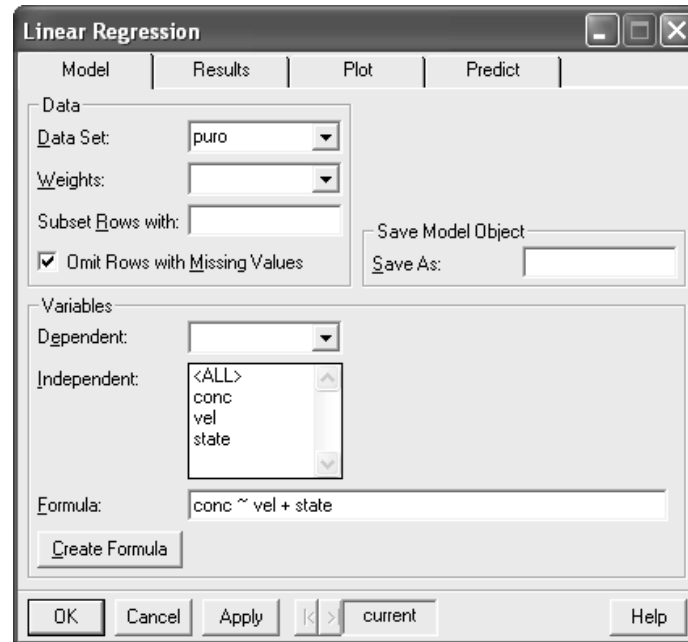


Figure 2.22. Linear Regression dialog.

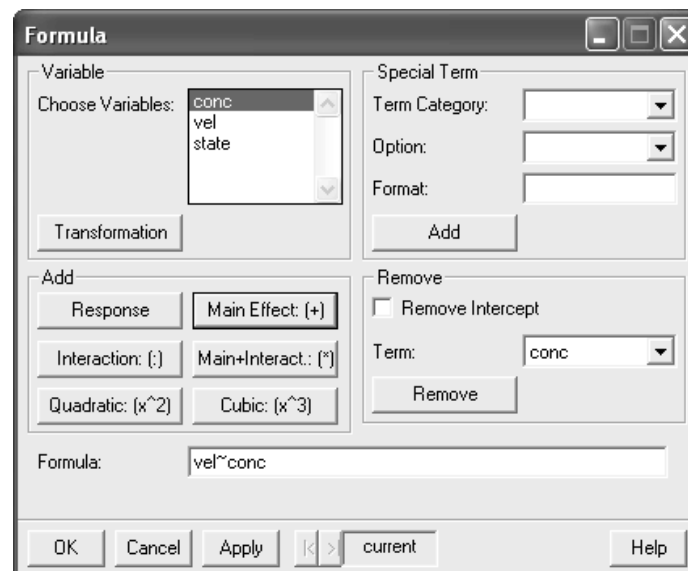


Figure 2.23. Regression Formula dialog.

- Click on the **Results** tab to move to that sheet
- Tick ANOVA table (as an option)
 - Can also save fitted values and residuals
- Click on **Plot** tab to move to that sheet
 - Can choose which plots to look at. None are ticked by default, but this doesn't matter, as we'll see later.
- Click on **Predict** tab to move to that sheet
 - Can save predicted values and confidence intervals
- Choose **OK**

The Linear Regression dialog opens a new Report window, shown in Figure 2.24. Looking in the Report window at the ANOVA table, we see that **conc** has an F -value of 38.8 and a corresponding p -value of 3.5×10^{-6} . At first glance, it appears as though the model provides an excellent fit to the data. Recall from the scatterplot, however, that the relationship between the two variables does not look particularly linear. We should explore the fit of the linear regression model further by looking at certain graphical representations (i.e., the residual plots).

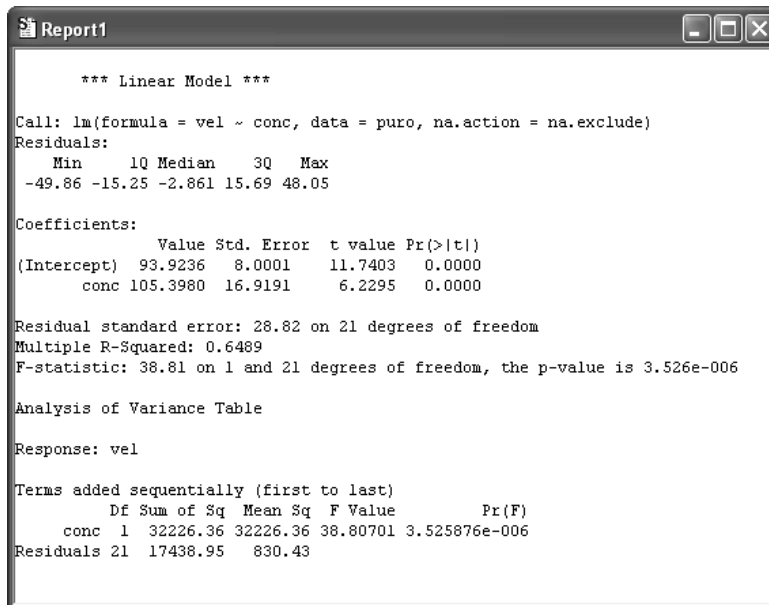


Figure 2.24. Regression Report window.

You will recall that we did not specify that any of the residual plots should be created, but we do not have to start from the beginning. Notice

that the object `my.reg` now appears in the right-hand side of the Object Explorer. Right-click on `my.reg` and choose **Plot....** The original menu for the residual plots appears, and we can now specify which plots we want to explore. Choose Residuals vs Fit, Response vs Fit, Residuals Normal QQ, and Cook's Distance as a start by ticking the appropriate boxes and then choosing **OK**. A graph sheet is created with four pages, one for each plot (not shown).

Note It should be mentioned at this point that graphs produced by statistical dialogs can be either editable or noneditable. The noneditable versions have the advantages of being quicker to produce and requiring less space for storage, but the disadvantages are that they cannot be edited to change labels, add annotations, and so forth. Since most computers have ample computing speed and space, we recommend the use of editable graphs. The desired type of graph can be defined using **Options - Graphs Options...** and ticking (or unticking) the desired default. The option **Create Editable Graphics** appears twice in the window and both should be ticked or unticked (as desired). The next section will not be possible to perform unless the residual plots were created as editable graphs (i.e., change the option and redo the graphs above). ◀

If we look at page 1 of the graph sheet, however, notice that the plot of residuals (y -axis) versus fitted values (x -axis) does not look like a cloud of points randomly distributed about the line $y = 0$. Furthermore, three points have been indicated as “suspect”: points 1, 10, and 23. A smoothed line has been added to make it all that much more apparent that there is a pattern to the points. Page 2 of the graph sheet shows the response variable (**vel**) versus the fitted values (x -axis) with an overlaid regression line (dotted) and a smoothed line (solid). Notice that the points, from left to right, are mostly below the line, mostly above the line, and then mostly below the line again. This trend indicates a poor fit of the model. Page 3 shows the residuals (y -axis) versus the quantiles of the standard normal (x -axis), which also indicates the points 1, 10, and 23. The fourth page contains Cook's distance (y -axis) versus an index (count of the order of the points) on the x -axis. Here, we see that points 1, 13, and 23 have the largest values of Cook's distance.

The graphs suggest that the simple linear regression provides a poor fit to the data, but you will perform a better analysis in the exercises.

This is a good place to mention that we can put multiple plots on a single graph page by copying and pasting. This feature is particularly relevant and helpful when trying to deal with a multipage graph sheet.

Placing Multiple Graphs onto a Single Graph Sheet (Windows only)

- This will not work if you have not followed the instructions in the Note above concerning editable graphs!
- Click **New** on the toolbar
- Click on **graph sheet**
- Choose **OK**
- Click on **Window** from the main menu and then **Tile Vertical** (just to simplify the copying process)
- Click on page 1 of residual plots to highlight it
- Choose **Edit** from the main menu and then **Copy Graph Sheet Page**
- Click on the header of the target (empty) graph sheet to highlight it
- Choose **Edit** from the main menu and then **Paste**
 - The graph is now copied in the target graph sheet.
- Repeat the previous three steps to copy pages 2–4 onto the target graph sheet.
 - Notice how space is made at each step to accommodate each new graph.
- Save the new graph sheet (calling it residuals). The result is shown in Figure 2.25.

2.14 PowerPoint (Windows Only)

If you intend to make an MS PowerPoint presentation that includes graphs you have created in S-PLUS, then your task has been made easy for you by an S-PLUS wizard designed to set this up for you.

Putting Graphs into PowerPoint

- Click on the **PowerPoint Presentation** button on the toolbar (upper right)
 - Creates PowerPoint presentation of selected, saved graph sheets
 - Opens the S-PLUS **PowerPoint Presentation wizard**
- Choose **Next**
- Choose **Add Graph...**

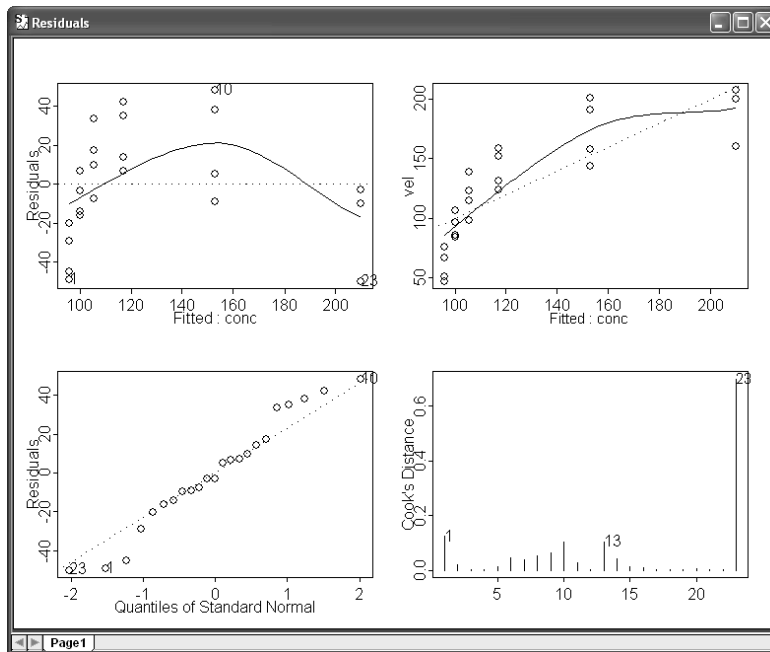


Figure 2.25. Combining graphs onto one graph sheet.

- Choose scatter.sgr
- Choose **Add Graph...** again
- Choose residuals.sgr
 - Could add additional graphs using the same process
- Click on **Next**
- Choose **Finish**
 - PowerPoint is opened and creates the slides.
- When complete, exit the wizard
- It asks to save the lists of graphs – you can, but choose No
- PowerPoint is open and is the active window
- Choose **File** from the main menu (in PowerPoint)
- Choose **Save As...**
- Specify the file name and location for the PowerPoint file

If several graph sheets have been saved, they can be put into the same PowerPoint presentation or one per presentation if you prefer. If a graph

sheet containing more than one page has been saved, PowerPoint will create one slide per page.

2.15 Excel (Windows Only)

The Windows-based version of S-PLUS also installs add-ins for MS Excel. You should notice a floating toolbar in Excel with three buttons as well as a new entry for S-PLUS in the main menu. These tools give you S-PLUS graphics capabilities from within Excel. If you have a data set in Excel and want to produce S-PLUS graphs, you don't have to transfer your data to S-PLUS but can stay in Excel and use the add-ins to produce your S-PLUS graphs. The three buttons in the S-PLUS toolbar (within Excel) are for creating a graph, modifying the layout of a graph, and modifying a plot.

There are essentially three steps to create S-PLUS graphs from within Excel. In Step 1, you are asked for the range of the data to be plotted. Keep in mind that you should specify the upper-left and lower-right corners of your block of data, as in A1:B10. If you highlight the data you want to plot before starting the process, the data range will be prespecified. In Step 2, you are asked to specify details for conditioning the data if you don't want to plot certain data items. The last step is where you specify the graph and plot type, which uses the same dialog boxes as in S-PLUS. The best way to test this is to open the Puromycin Excel file we created earlier and give it a try.

There are now three options for working with Excel: importing an Excel data file (which converts the data to an S-PLUS data frame), as explained earlier; using Excel directly to produce S-PLUS graphics while in Excel, as explained above; or linking to an Excel data file while in S-PLUS. The last option is discussed below.

Linking to Excel Files

- Click on **File** and **Open...**
- Choose the Puromycin Excel file
 - An Excel sheet is opened within S-PLUS
 - Can see letters as column headers
 - Can enter Excel formulas, etc.
- The following new buttons have been added to the toolbar (see Figure 2.26)
 - Left button starts the Excel to S-PLUS Link Wizard
 - Middle button updates a link
 - Right button removes a link



Figure 2.26. The Excel toolbar.

Running the Link Wizard creates a link so that the Excel data can be accessed in future S-PLUS sessions. Be sure to tick the box **Allow S-Plus to store link information in Excel**. Selecting the Excel data name and then clicking on the Update Link button will allow S-PLUS to update its version of the data to reflect any modifications that were made in Excel. Other S-PLUS functions can now be run directly on the Excel data.

2.16 Script Window

The Script window offers an easy way to build and run a program or to keep track of what has been done so far. The idea behind this type of window is that it is a place where commands can be entered, run, and saved, with the output appearing in a separate section of the window. It can also be used with GUI-style commands, but, as these can be quite cryptic, its greatest use is to be run in conjunction with line commands covered in the rest of the book.

Opening a Script Window

- Choose **New** from the main toolbar
- Choose **Script File**
- Choose **OK**

There are two panes in the window: the upper pane is used for entering and running commands, and the lower pane is for showing output. In the example to follow, you should open the Scatterplot graph sheet and choose **Window** (main menu)-**Tile Vertical** so that both the graph sheet (containing the scatterplot created earlier) and Script window can be accessed at the same time.

Viewing GUI Commands in the Script Window

- Click and hold the mouse button on the scatterplot
- Drag the mouse to the upper pane of the Script window
- Release the mouse button
 - All graph options used for the scatterplot appear (see Figure 2.27)

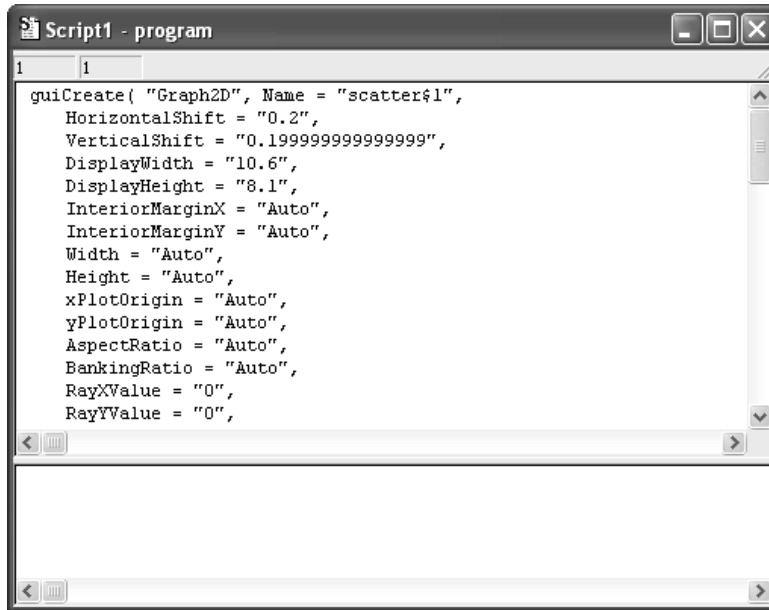


Figure 2.27. The Script Window.

Saving GUI Commands from the Script Window

- Click in the upper pane of the Script window to make it active
- Choose **File** from the main menu
- Choose **Save As...**
 - Default file type is *.ssc
- Specify the file name and location
- Choose **Save**

The GUI options and commands that have been shown here are sometimes difficult to interpret and are beyond the scope of this book. It can be helpful, however, simply to view what the current settings are at any given point in time.

The real power and usefulness of the Script window in the context of this book is to edit and submit segments of code. Up until now all the “commands” we have entered have been in the form of mouse movements and interacting with the GUI. The commands to be entered in the Script window are more akin to programming language commands on which the S-PLUS processor then acts. Suppose you have a data set **x** consisting of one vector and you wanted to calculate the mean of **x**. You would choose

Statistics, Data Summaries, and then **Summary Statistics....** However, you could also ask for the mean to be calculated directly by specifying `mean(x)` in either the Commands window (executed with the ENTER key) or in the Script window (executed with the **Run** button).

You may have noticed when you opened the Script window that three buttons were added to the toolbar, as in Figure 2.28. The left button has a right arrow for running scripts, the middle button is to stop the processing of the commands in the script window, and the right button is for finding specified text.



Figure 2.28. Script Window toolbar.

Running Commands in a Script Window

- Clear the existing text in the Script window
 - Choose **Edit** from main menu
 - Choose **Select All**
 - **Delete**
- In the upper pane, enter `mean(1:10)`
 - Calculates the mean of the numbers from 1 to 10
- Click on the **Run** button
 - Command and output appear in lower pane

In this manner, you could work on groups of commands to fine-tune a certain process and submit them as a group. It is a good way to work when building a function or longer program, tasks that will appear towards the end of the book.

2.17 UNIX/Linux GUI

The GUI available with S-PLUS under UNIX/Linux works in more or less the same manner as under Windows. The biggest difference is that it doesn't offer all the functionality as under Windows. Rather than point out all the similarities and differences between the two GUIs, we will simply

provide an overview of how to use the GUI under UNIX/Linux, and the similarities and differences will become apparent.

Help System

The help system on UNIX/Linux is slightly different than on Windows in that there is really only one menu choice. Click on **Help** from the main menu and the three choices are **Contents**, **Index**, and **Search**. It doesn't matter which of the three choices is made as they are simply the three tabs of the window that is opened. The **Index** tab is used to look at functions and built-in S-PLUS data sets, the **Contents** tab gives an overview, and the **Search** tab can be used to search for information on any topic.

Chapters

Chapters must be specified manually. This means that if you are working on two different projects and want to create two chapters to keep them both separate, you have to create the chapters manually outside of S-PLUS. Fortunately, the procedure is quite easy. Simply move to the desired directory and issue the UNIX/Linux command **Splus CHAPTER**. This will have the effect of creating a **.Data** subdirectory in which S-PLUS will store data from that chapter. To access data from that chapter, simply invoke S-PLUS from within that directory with the command **Splus -g &**. When S-PLUS is started, it opens with only the Commands window open and should look like Figure 2.29.

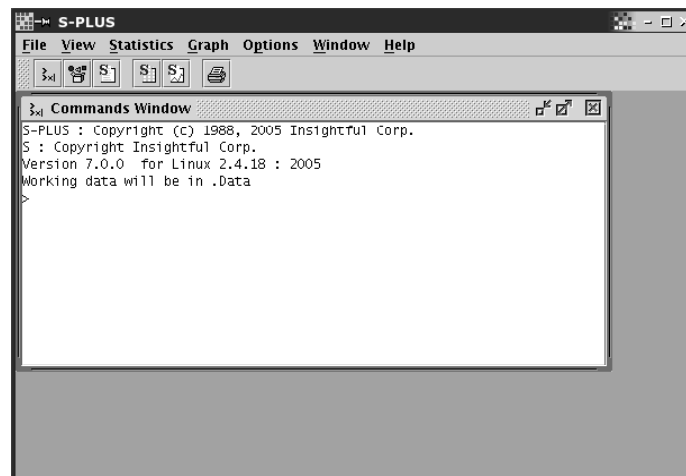


Figure 2.29. The UNIX/Linux Startup window.

Objects Summary Window

With the UNIX/Linux GUI, there is no Object Explorer; instead, there is an **Objects Summary Window**. To see the types of information it provides, we need some data. Go to the Commands window and type `x <- 1:10` to create the variable `x`. There is a button on the main toolbar for the **Objects Summary Window**. Pressing the button will open the window, showing summary information on all the variables available to the user. Figure 2.30 shows the **Objects Summary Window** and the description of our new variable `x`.

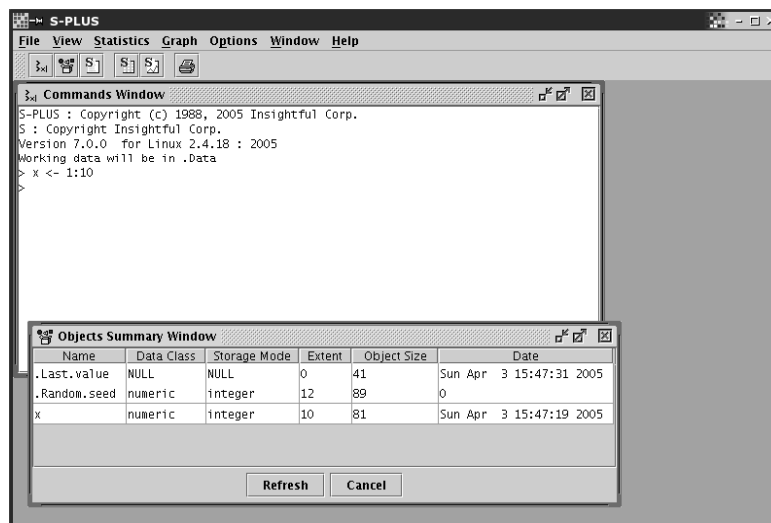


Figure 2.30. The Objects Summary Window.

The **Objects Summary Window** is not as versatile as the Object Explorer but is still a good way of keeping track of what variables exist.

Importing and Exporting Data

Importing and exporting data sets is mostly the same as with Windows in terms of how the GUI operates and the file types that are compatible, and the topic won't be covered here in any more detail.

New Data Viewer

Just as the **Objects Summary Window** is a little different from what we saw with Windows, viewing data is also done a little differently.

- Choose **View** from the main menu
- Choose **New Data Viewer**
- Type **Puromycin** in the **Data Set** field (see Figure 2.31)

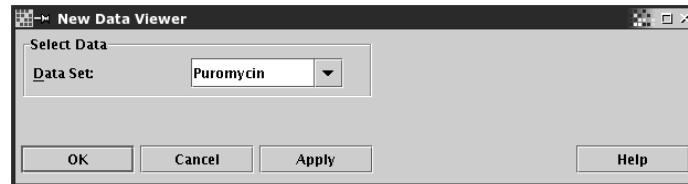


Figure 2.31. Data specification dialog.

- Choose **OK**
 - The Data viewer is opened (similar to Data sheets in Windows) (see Figure 2.32)

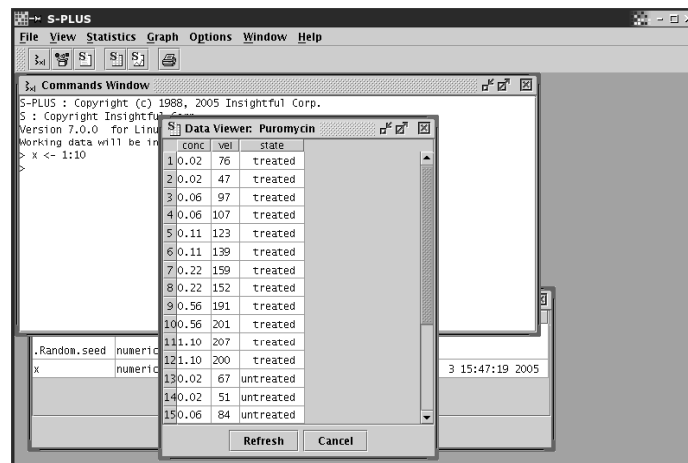


Figure 2.32. Data viewer.

Graphs

The biggest change to keep in mind when creating graphs with the UNIX/Linux GUI is that editing graphs works a little differently. Once you have clicked **OK** after specifying a graph, you won't be able to edit it. The way around this is to click on **Apply** instead. That way, the dialog box with all the tabs for specifying everything about the graph is still open and

able to accept modifications. The only disadvantage is that every time **Apply** is clicked, a new graph window is opened. Other than that, the GUI operates in much the same way as with Windows, and a demonstration appears below.

- Choose **Graph** from the main menu
- Choose **Scatter Plot**
- Define **Puromycin** as the **Data Set**
- Define the **x Axis Value** and the **y Axis Value** from the pull-down menu ($x=\text{conc}$, $y=\text{vel}$)
- Conditioning on a variable will create a Trellis scatterplot as opposed to a conventional one. Choose **state** as the conditioning variable (see Figure 2.33)

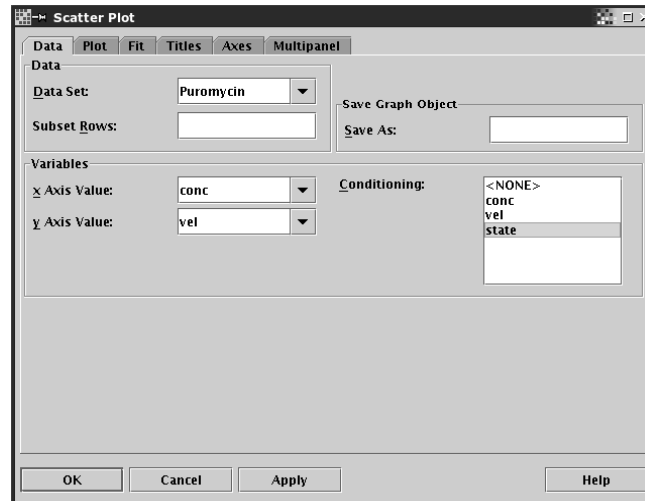


Figure 2.33. Trellis scatterplot data specification.

- Click on the **Plot** tab and notice that the available options are similar to those in Windows
- Change **Color** to black with the pull-down menu
- Change the **Symbol Style** to Circle, Solid with the pull-down menu (see Figure 2.34)
- Tabs also exist for **Fit**, **Titles**, **Axes**, and **Multipanel**
- Choose **OK**

The final Trellis graph appears in Figure 2.35.

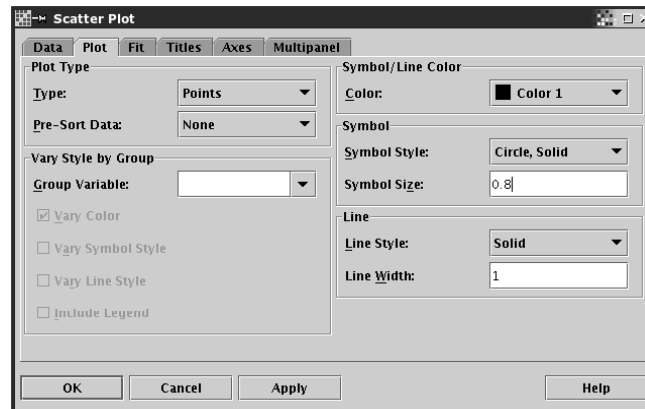


Figure 2.34. Trellis scatterplot symbol options.

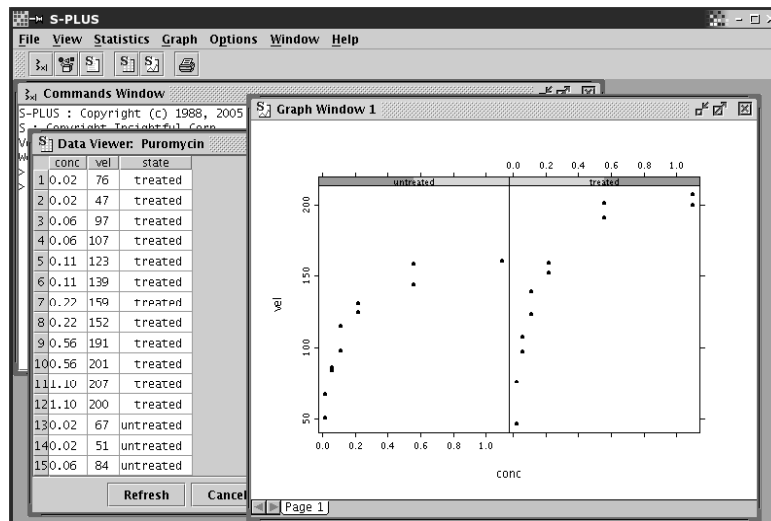


Figure 2.35. Trellis scatterplot.

Summary Statistics

Summary statistics are calculated in the same way with the UNIX/Linux GUI as with the Windows GUI, but an example of the choices and output is shown for reference.

- Choose **Statistics** from the main menu

- Choose **Data Summaries**
- Choose **Summary Statistics...**
- Type **Puromycin** in the **Data Set** field (see Figure 2.36). Clicking in the **Variables** field will show all the variables that exist in this data set.

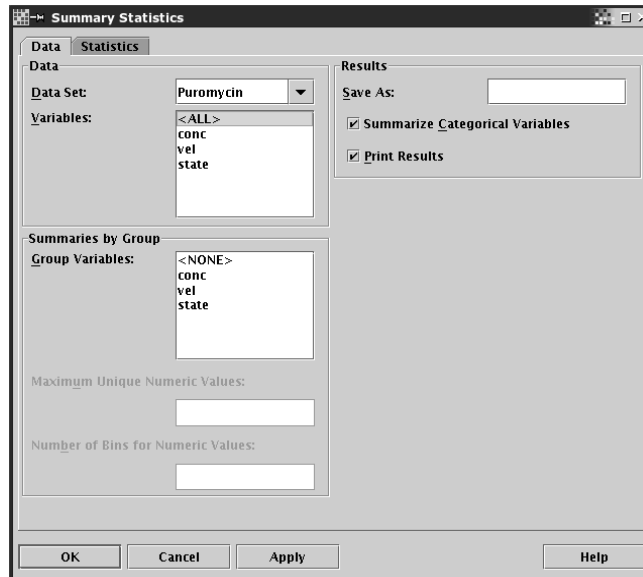


Figure 2.36. Summary Statistics dialog.

- Choose **OK**
 - A Report window is opened with the results as in Figure 2.37

Linear Regression

An example of how to perform linear regression with the UNIX/Linux GUI is shown below, although it works the same way as with the Windows GUI.

- Choose **Statistics** from the main menu
- Choose **Regression**
- Choose **Linear...**
- Choose **vel** as the **Dependent Variable** from the pull-down menu
- Choose **conc** as the **Independent Variable** from the pull-down menu (see Figure 2.38)

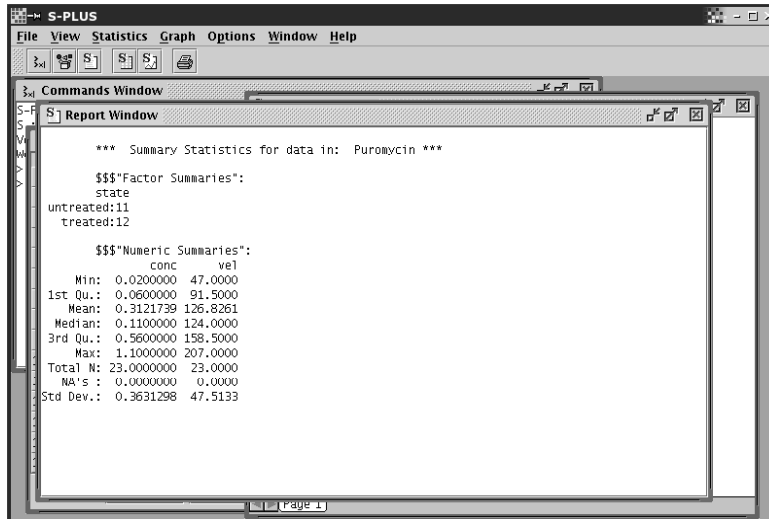


Figure 2.37. Summary Statistics Report window.

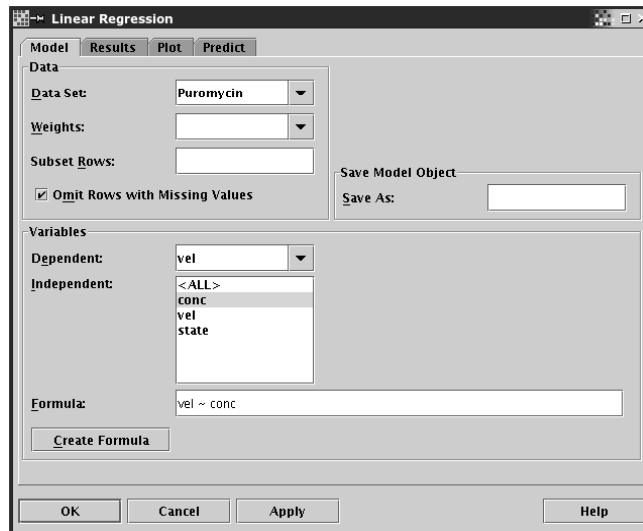


Figure 2.38. Linear Regression specification dialog.

- On the **Results** tab, choose **ANOVA Table**
- Choose residual and diagnostic plots, as appropriate, from the **Plot** tab
- Choose **OK**
 - Results appear in the Report window (see Figure 2.39).

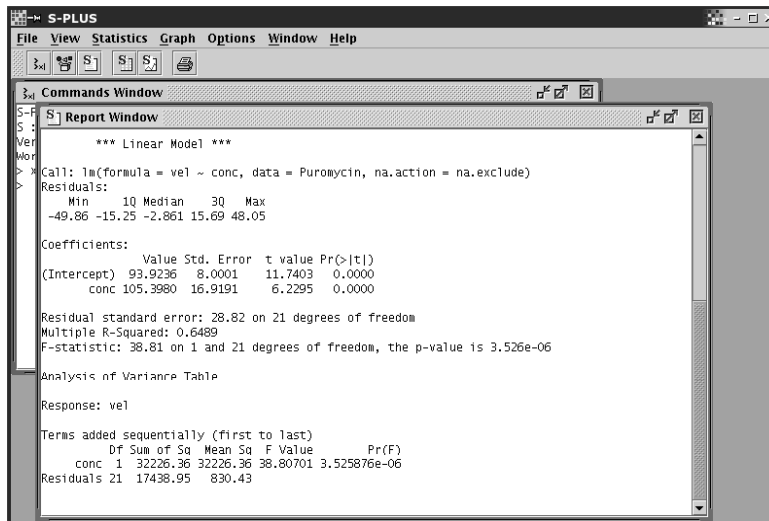


Figure 2.39. Linear Regression Report window.

- Residual and diagnostic plots appear in the graph window (see Figure 2.40).

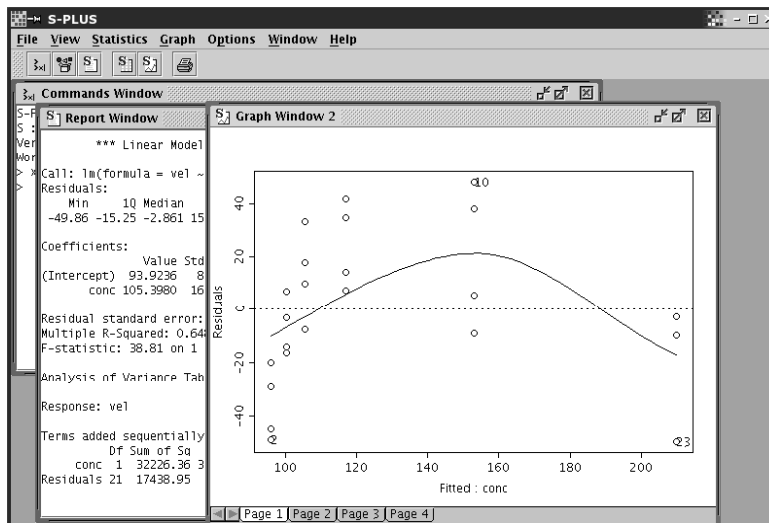


Figure 2.40. Linear regression residual plots.

We hope this quick introduction to the UNIX/Linux GUI, combined with the information from the rest of the chapter, will get you on your way, no matter what the platform.

2.18 Summary

The GUI you just looked at provides a convenient, easy-to-use approach to unlocking the possibilities of S-PLUS for the novice and expert alike. The whole power of S-PLUS is in a way hidden behind the simple clicks required to use the GUI.

The remainder of the book will primarily focus on the S-PLUS language. You will complete several sessions of analyzing data and encounter a variety of statistical routines along the way. You will see the commands involved in running some of the procedures that you have already learned in this chapter.

We strongly recommend that you do not stop here after seeing the “easy” point-and-click approach to S-PLUS. By continuing with the remaining chapters, you will not only understand what S-PLUS is doing and how it works, but you will learn some things that you will really want to be able to do that you simply cannot do with just the GUI.

2.19 Exercises

Exercise 2.1

Perform a more complete analysis of the Puromycin data, including such exploration as transformations of the predictor variable (square root and log are typical), the effect of **state**, and the interaction of **state** and the transformed variable. Be sure to use only the GUI to do your analyses and visually check the fit of the model.

Exercise 2.2

Using the search path, locate the S-PLUS built-in **Geyser** data set. Use the help system to find out something about the data. Create a scatterplot. (Hint: This data set is not a data frame, and you will have to be creative to coerce it into one before you can create the plot.) What do you notice about the points? Are they randomly distributed, linear, or grouped? Use the Annotation palette and Graph Tools palette to group the points. Label the groups of points. Calculate the means of each group (use the **Grouping variable** option in the **Summary Statistics** menu). Write the means of each group onto the scatterplot. Add a linear regression line to the graph. Save the graph sheet and then make it into a PowerPoint presentation.

2.20 Solutions

Solution to Exercise 2.1

We know from our scatterplot that the relationship between the variables is not particularly linear and perhaps look quadratic. We will try two transformations of the data and create scatterplots to make a first check of which model might fit best. We will try the two transformations square root and the natural logarithm (\ln , although the function in S-PLUS is `log`), but first we need to add these variables to our data frame.

Inserting a Calculated Column of Data

- Double-click on `puro` in the Object Explorer to open the data frame.
- Right-click on the top of any column
- Choose **Insert Column...**
- Define a name for **Name(s)**
 - We chose `sqrtconc`
- Define the transformation in **Fill Expression:**
 - Use **`sqrt(conc)`** (see Figure 2.41)
- Change **Start Column** to `vel`
 - Default is to put new column at the end
 - New column will go before `vel`
- Choose **OK**

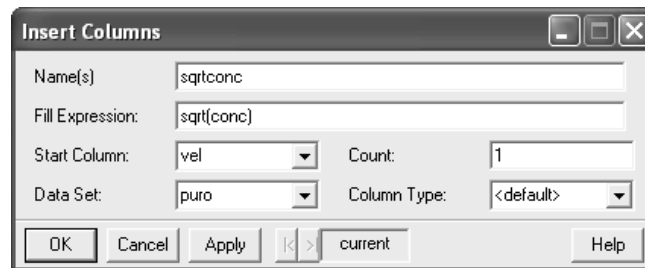


Figure 2.41. Inserting a column of data.

Repeat the procedure to create the variable `logconc` using the `log` function.

We want to replot the data using our transformed variables to look for a linear relationship. This time, however, we will add a smoothed line (curve) to the data that will help us visualize any trends that might exist.

Smoothed Line Plot

- Use the palette as before, OR...
- Click on top of the column for *sqrtconc*
 - Defines *x*-axis variable
- CTRL-click on top of column for *vel*
 - Defines *y*-axis variable
- Choose **Graph** from the main menu
- Choose **2D Plot...**
- Change **Plot Type:** to **Smoothing-Loess Plot (x, y1, y2, ...)**
 - Example of layout appears
- Choose **OK**
- Repeat for *logconc* and put both on one graph sheet (see Figure 2.42)

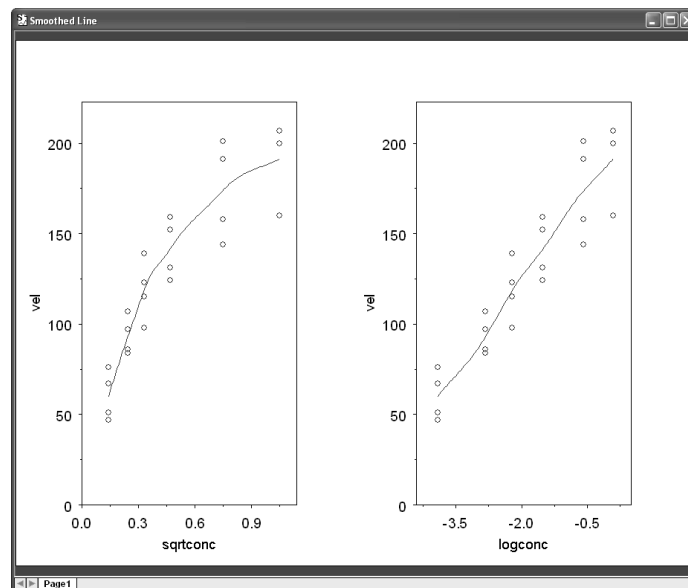


Figure 2.42. Loess smoother curves.

Notice that the general shape of the data and the smoothed line are much more linear than with the original data but are still curved. Repeat the process above using `logconc`, and you will find that the log transformation of the data is the one that produces what appears to be a linear relationship. Anyone who has worked with concentration data could probably have guessed this right from the start.

Now that we are satisfied with a transformation on `conc`, we can go back to the task of actually performing a better regression analysis.

Regression with Transformed Variables

- Choose **Statistics** from the main menu
- Choose **Regression**
- Choose **Linear...**
- Choose **Create Formula**
- Add `vel` as the response
- Click on `conc`
- Choose **Transformation**
 - We already have the transformed variable in our data frame, but this is shown for information.
- In the **Function** section, Choose **Log: $\log(x)$**
 - **New Variable(s)** shows `log(conc)` (see Figure 2.43)

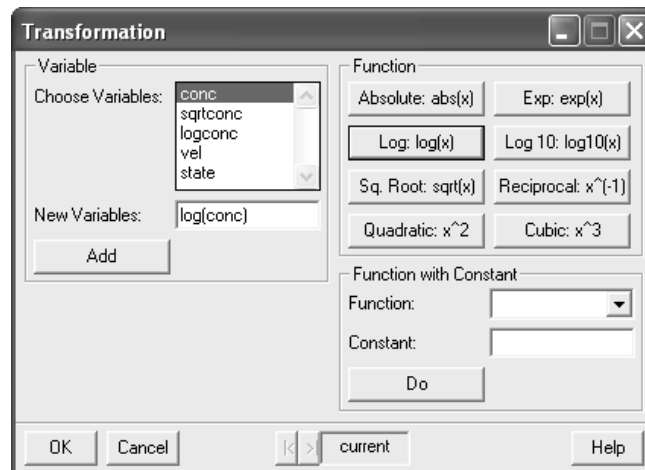


Figure 2.43. Variable transformation for regression.

- Choose **Add**
- Choose **OK**
- Click on **log(conc)** in the Variable section of the Formula dialog
- Add **Main Effect: (+)**
- Choose **OK** for **Formula**
- On the **Results** tab, choose **ANOVA table**
- On the **Plot** tab, click residual plots of interest
- Choose **OK** for **Linear Regression**

The F -statistic is still highly significant ($F=146.9$ on 1, 21 df, $p=6.0 \times 10^{-11}$) (see Figure 2.44), but the graphs look much better this time (see Figure 2.45).

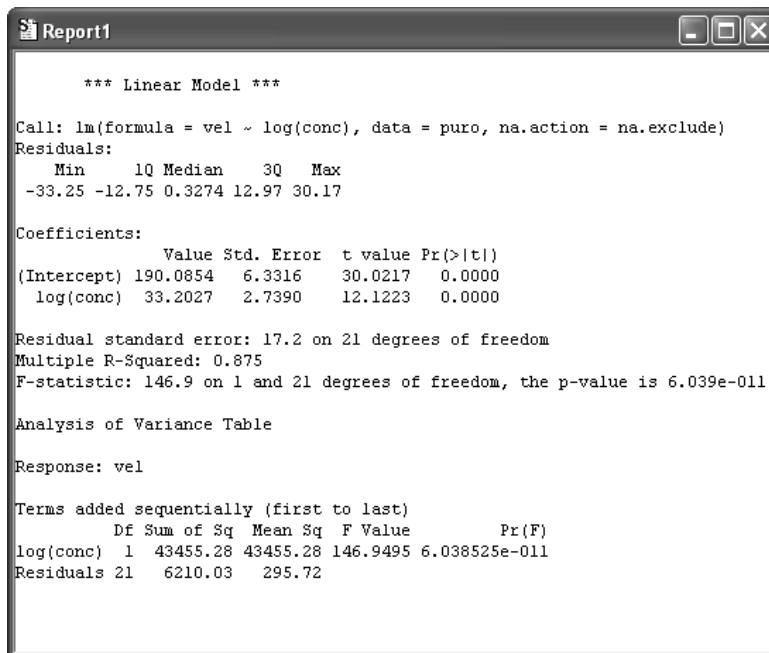
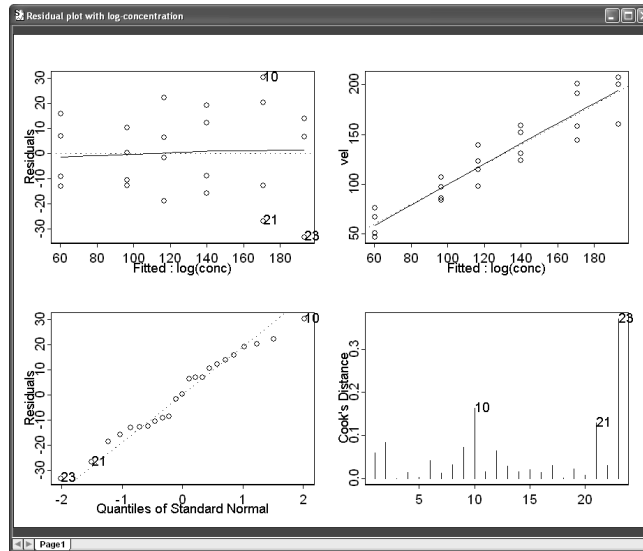


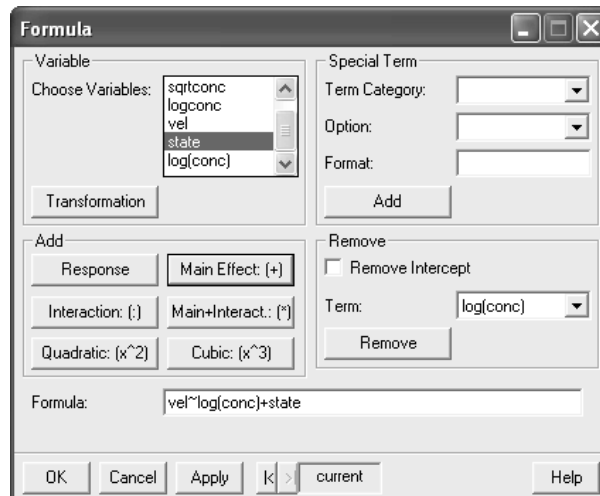
Figure 2.44. Regression output using log(conc).

Note To run a linear regression model, we could choose **Apply** instead of **OK**. Choosing **Apply** has the effect of producing the same output windows (Report and Graph sheet), but now the Linear Regression window remains

Figure 2.45. Residual plots using $\log(\text{conc})$.

open and we can modify the formula rather than restarting our definition of it. ◀

Reenter the commands from above to set up the model with ***log(conc)*** as the predictor variable. Add ***state*** as a second predictor variable (as a Main Effect), see Figure 2.46, and choose ***OK*** for the formula.

Figure 2.46. Regression formula using *state*.

This time, choose **Apply** for the regression. A new graph sheet is opened, but the results are printed in the same Report window. The graphs still look good, and we notice that **state** is significant on its own as a main effect. The output from this regression model appears in Figure 2.47, but the residual plots have not been shown.

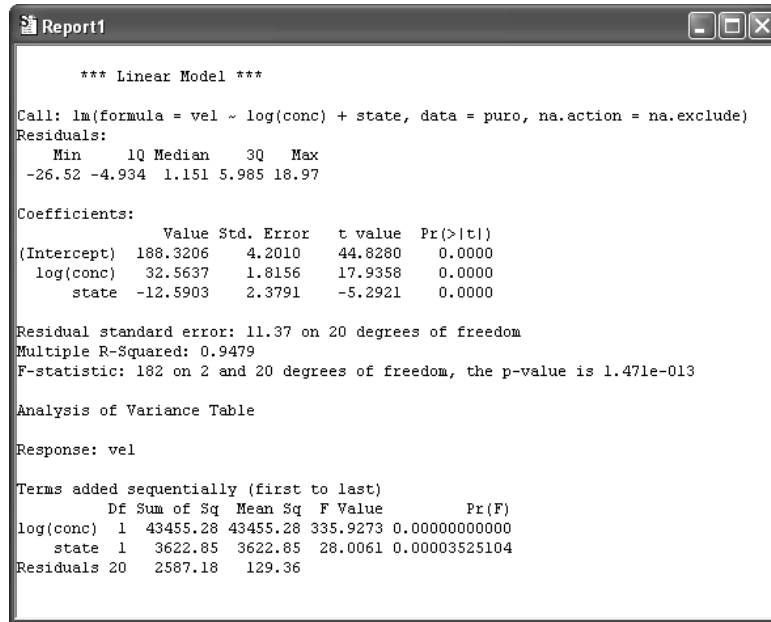


Figure 2.47. Regression output using state.

The last point we will cover is the interaction between **log(conc)** and **state**. We will leave that to the reader to perform (i.e., the solution stops here, but you can refer to Figures 2.48 and 2.49 to see if you're on track). We will simply note that the interaction term is significant, the interpretation of which is also left up to the reader (do scatterplots by **state**).

Solution to Exercise 2.2

The first task to be completed is to locate the Geyser data.

Copying Data to Another Folder

- In the left pane of the Object Explorer, click on the + to the left of the SearchPath icon

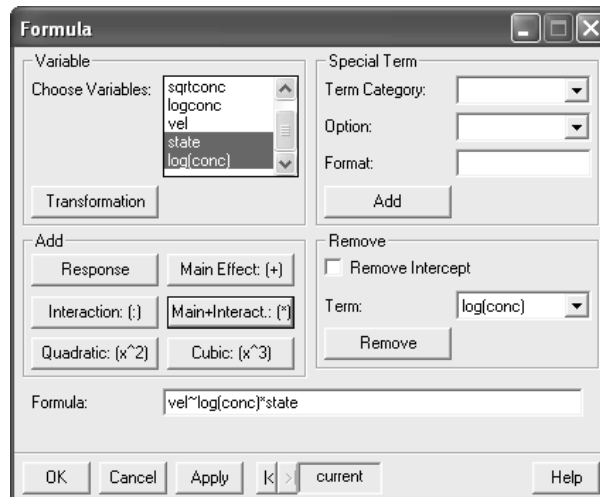


Figure 2.48. Regression formula using interaction term.

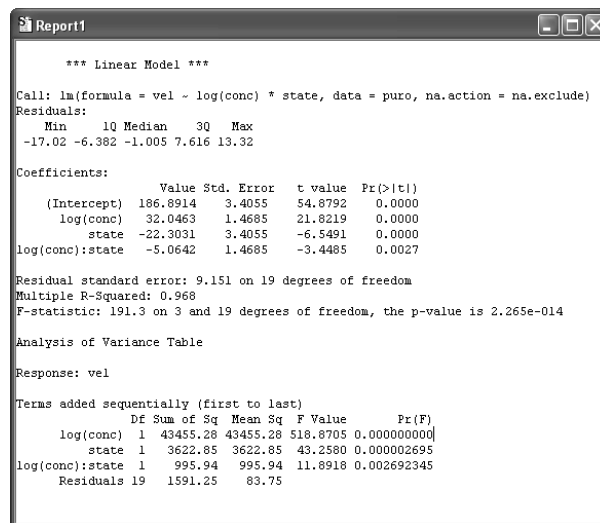


Figure 2.49. Regression output using interaction term.

- Click on the **+** to the left of the icon for **data** and all the built-in S-PLUS data sets will be listed
- Scroll down to find the **Geyser** data set, right-click on its icon, and choose **Copy**
- At the very top of the Object Explorer, right-click on the **Data** icon and choose **Paste**

- You have now copied the Geyser data set to your working data folder.

Using the Object Explorer, you can see that it has two variables: ***waiting*** and ***duration*** (see Figure 2.50). To find out what these two variables contain, invoke the help system.

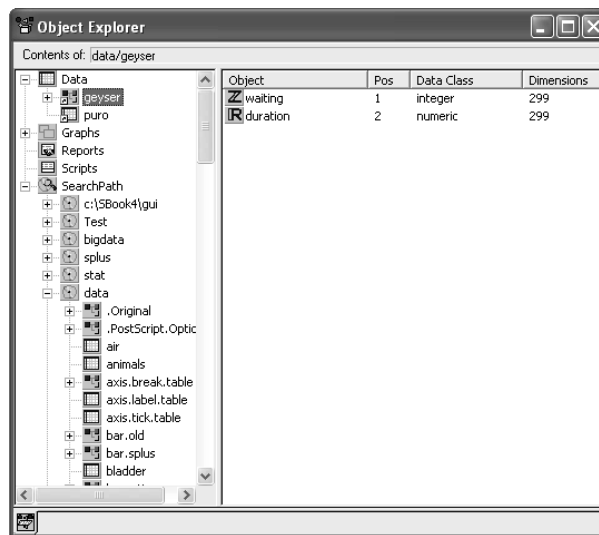


Figure 2.50. Object Explorer.

Help Entry for Geyser Data Set

- Click on the ***Help*** button on the toolbar
- Click on ***Available Help*** and then on ***Language Reference***
- Choose the ***Index*** tab (default)
- Type `gey` and hit RETURN or click the ***Display*** button
 - The help entry now appears.

The help entry states that these are the successive waiting times and durations of eruptions for the Old Faithful Geyser in Yellowstone National Park.

With scatterplots, it is necessary to choose one variable for the x -axis and one for the y -axis. For this kind of exploratory analysis where no hypothesis

is being examined, it doesn't really matter which variable goes where, but we chose waiting time for the horizontal axis and duration for the vertical.

Scatterplot for the Geyser Data

- Click on **waiting** in the right-hand pane of the Object Explorer
- CTRLclick on duration
- Click on the **2D Plots** button in the toolbar to open the palette
- Choose Scatter Plot (upper left-hand button)
- The resulting plot is a straight line with **waiting** on both axes
- This is NOT what we wanted!

The problem is that the plotting routines are expecting that the data are a data frame, but **Geyser** is saved as a list. Notice that in the Object Explorer, the icon is different from that of **puro**. To see the data type of a data object, right-click on its icon and choose **Properties**. Under **Class**, you can see that **Geyser** has a type (or class) list. If you double-click on the **Geyser** icon, it will not open a data sheet and display the data; a Report window is opened instead. Getting around the problem is not direct but can be done as follows.

Creating a Data Frame

- Right-click on the **Data** icon in the left pane of the Object Explorer
- Choose **Insert data.frame**
 - A data sheet is opened with the name **dfl**.
- Click on the **Geyser** icon in the left pane of the Object Explorer
 - **waiting** and **duration** appear in the right pane of the Object Explorer.
- Double-click on the **waiting** icon in the right pane of the Object Explorer
 - A read-only data sheet appears with the **waiting** data.
- Click on the top of the column with the **waiting** data to highlight it
- Press CTRL-C to copy the **waiting** data
- Click on the new **dfl** data sheet to open it, click on the top of the first column to highlight it, and press CTRL-V to paste the **waiting** data here
 - The new column name is simply **X**.

- Change the column name by right-clicking the top of the column, choosing **Properties**, changing the **Name:** to **Waiting**, and clicking **OK**
- Repeat the necessary steps to copy and rename the **duration** data into the new data frame
- Close the new data frame
- Rename the new data frame by right-clicking on the **dfl** icon in the left pane of the Object Explorer, clicking on **Properties**, changing the **Name:** to **Geyser**, and clicking **OK**

Now that the **Geyser** data have been saved as a data frame, we can create our scatterplot as before.

The axis labels are not ideal, and the plot symbol is a bit hard to see on the printed page, so we will change these features on our graph.

Changing the Axis Labels on the Geyser Scatterplot

- Click on the x -axis label
- Right-click and choose **Edit In-Place...**
- Change text to Waiting Time
- Click outside of the box to finish the text
- Repeat to change the font or font size as desired

Repeat the procedure above to change the y -axis label to Duration of Eruption.

The points are quite large so that they overlap and are blue (but the print in the book is black and white).

Changing the Plot Symbol

- Click on any point
- Double-click on green knob
- Click on **Symbol** tab
- Change **Style** to • **Circle, Solid**
- Change **Color** to black
- Change **Height** to 0.10 (see Figure 2.51)

- Choose **OK**

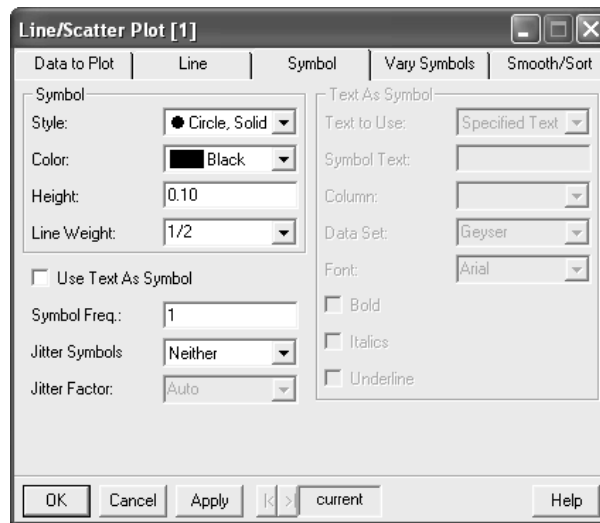


Figure 2.51. Plot Symbol dialog.

The result of these modifications appears in Figure 2.52.

There are many ways to “group” the data here. Speaking in terms of the way the data are distributed, you could easily say that the distribution of durations changes with waiting time. Consider, for example, a waiting time of 65. The durations corresponding to these shorter waiting times (< 65) range from about 3.7 to 5.5, whereas the durations for the longer waiting times (> 65) range from roughly 1 to 4.8. You could also group the data by looking at the distribution of waiting times for short durations as opposed to long durations.

We chose to split the data according to a waiting time of 60 minutes such that all points to the left of this defining line have short waiting times and those to the right have long waiting times. The remaining task is to record all of this onto the scatterplot using the **Annotation** palette in Figure 2.53.

Adding Annotations to a Graph

- Click on the **Annotation** palette in the toolbar
- Using the **Line Tool**, draw a vertical line starting at the waiting time of 60 minutes. Click on **Select Tool** to turn off the **Line Tool**.

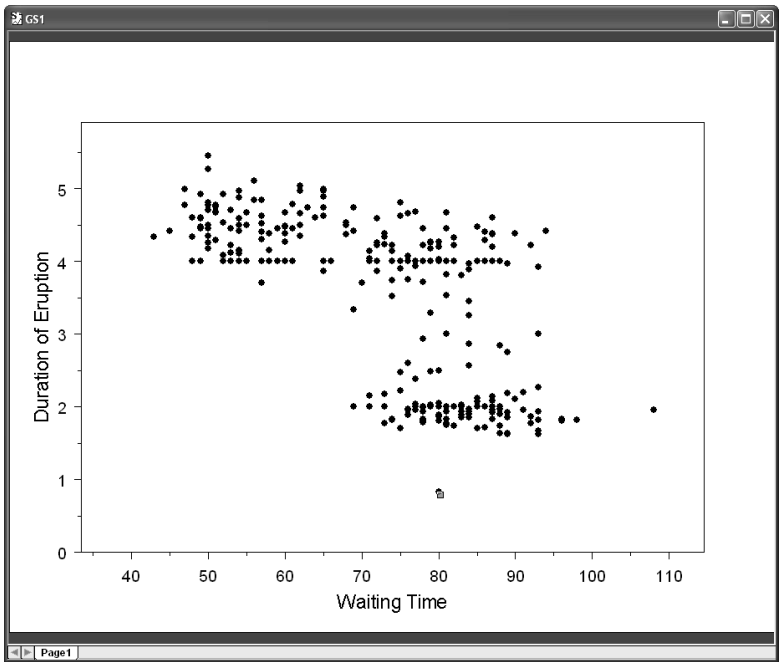


Figure 2.52. Geyser scatterplot.



Figure 2.53. The Annotation palette.

- Double-click on the line and change the **Style**, **Color**, and **Weight**, as in Figure 2.54, and then choose **OK**

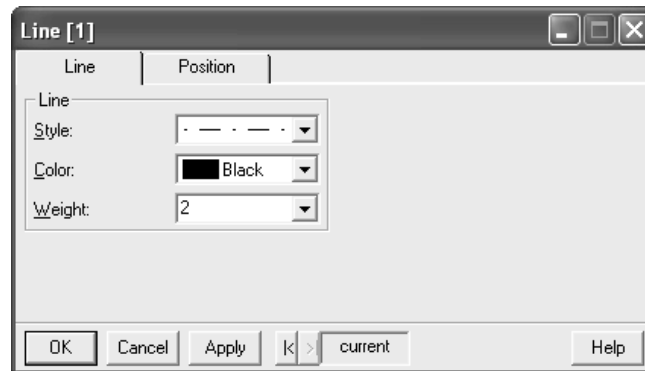


Figure 2.54. Plot Line dialog.

- Use the **Rectangle Tool** to draw a box around the points on either side of the reference (vertical) line. What do you notice? Click on **Select Tool** to turn off **Rectangle Tool**.
- Use the **Comment Tool** to label the rectangular boxes just drawn. Click on **Select Tool** and then change the font, text, and scale as desired. Experiment with the various fonts (choosing **Apply**, rather than **OK**, simplifies the experimentation process). Choose **OK** when you are satisfied with the results.
- Use the **Arrow Tool** to connect a label to its corresponding box and click on **Select Tool** to end. Change the color of the arrow (and arrowhead) to black and choose **OK**.
- Save the graph sheet

If you drew and labeled in a manner similar to us, you should now have a graph that looks something like the one we obtained in Figure 2.55.

The take-home message here is that if the waiting time is short (< 60 minutes), then the eruption should be fairly long (4 minutes or more). If, however, the waiting time is longer (> 60 minutes), there is no saying how long the duration might be. We leave it to the reader to write the group means on the plot as well as a linear regression line. Saving the graph sheet into a PowerPoint presentation was covered earlier.

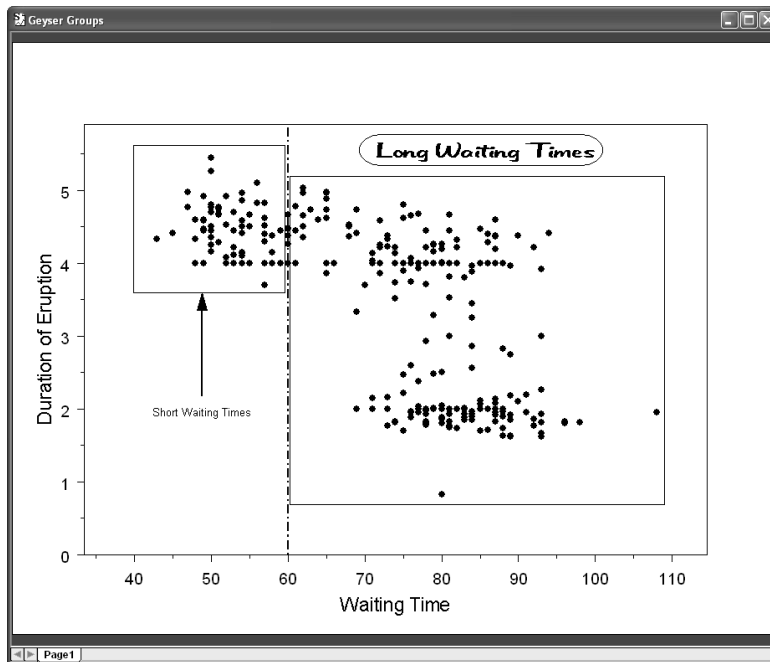


Figure 2.55. Geyser scatterplot with annotation.

The Basics of S-PLUS

Krause, A.; Olson, M.

2005, XXII, 444 p. 31 illus., Softcover

ISBN: 978-0-387-26109-6