STATGRAPHICS® *Sigma Express* for Microsoft Excel®

Getting Started

STATGRAPHICS® SIGMA EXPRESS GETTING STARTED

© 2011 by StatPoint Technologies, Inc.
www.STATGRAPHICS.com

Table of Contents

Preface

This book introduces users of STATGRAPHICS *Sigma Express* to installation of the program, basic system operation, and methods for analyzing data. The program functions as an "Add-In" for Microsoft Excel. All data entry and management is done in Excel. Statistical calculations are performed by STATGRAPHICS and the output presented in a separate window. Users may then copy and paste the results back to Excel or to Microsoft Word, PowerPoint or other applications.

Additional PDF files with detailed descriptions of each statistical procedure may be accessed from within Microsoft Excel by selecting *Help - Procedure Documentation* from the STATGRAPHICS Add-In menu.

NOTE: The data files referenced in the manual and in the other PDF files are provided with the program. They may be found in the DATA subdirectory where STATGRAPHICS is installed, which is usually *c:\Program Files\Statgraphics\STATGRAPHICS Sigma Express*.

StatPoint Technologies, Inc.
July, 2011

Installation

Installing STATGRAPHICS Sigma Express, accessing it from within Microsoft Excel, and first-time activation.

1.1 Installing STATGRAPHICS Sigma Express

STATGRAPHICS Sigma Express is distributed in two ways: over the Internet in a file that may downloaded to your computer, or on a CD-ROM. To use the program, it must first be installed on your hard disk. As with most Windows programs, installation is extremely simple:

Step 1: If you received the program on a CD, insert the CD into your CD-ROM drive. After a few moments, the setup program should begin automatically. If it does not, open Windows Explorer and execute the file **sgexpressinstall.exe** in the root directory on the CD-ROM.

If you downloaded the program over the Internet, locate the file that you downloaded and double-click on it to begin the installation process.

Step 2: If you are running the program from a CD, a dialog box will be displayed asking you to specify the edition of Microsoft Excel with which STATGRAPHICS will be used and the desired language:

Figure 1-1. Installation Selection Dialog Box
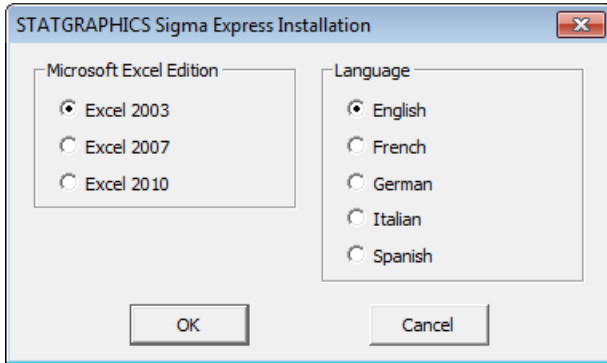
Make the desired selections. This will execute the proper MSI file.

If you downloaded the program from the Internet, you can simply execute the EXE or MSI file that you downloaded.

Step 3: Once installation begins, you will be guided through a series of dialog boxes. The first dialog box welcomes you to STATGRAPHICS *Sigma Express*:
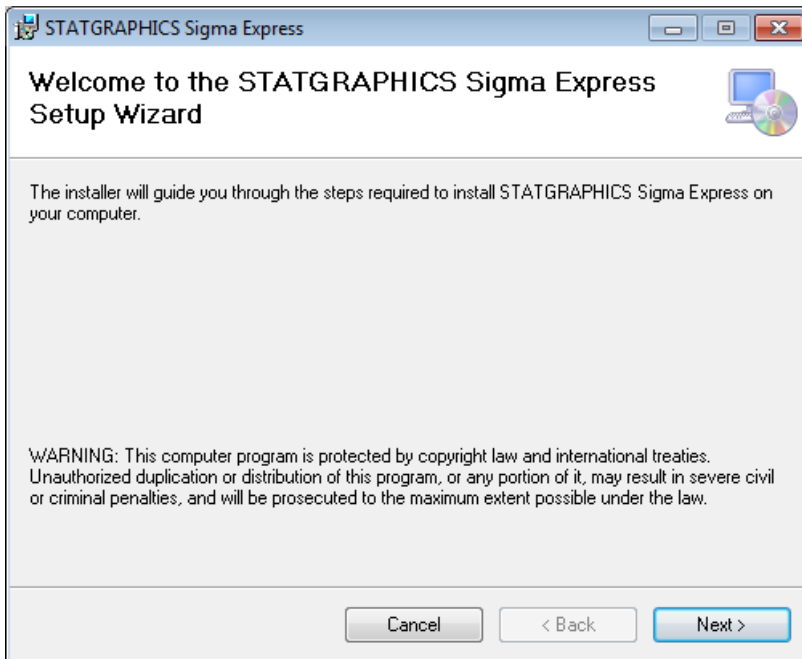
Figure 1-2. Welcome Dialog Box

Just press the *Next* button.

NOTE: In order to install and activate STATGRAPHICS *Sigma Express* you must have administrator rights to your computer. In the event that you need to have a system administrator present during the installation process, we highly recommend installing and activating the software while they are present.

Step 4: The second dialog box displays the license agreement for the software:
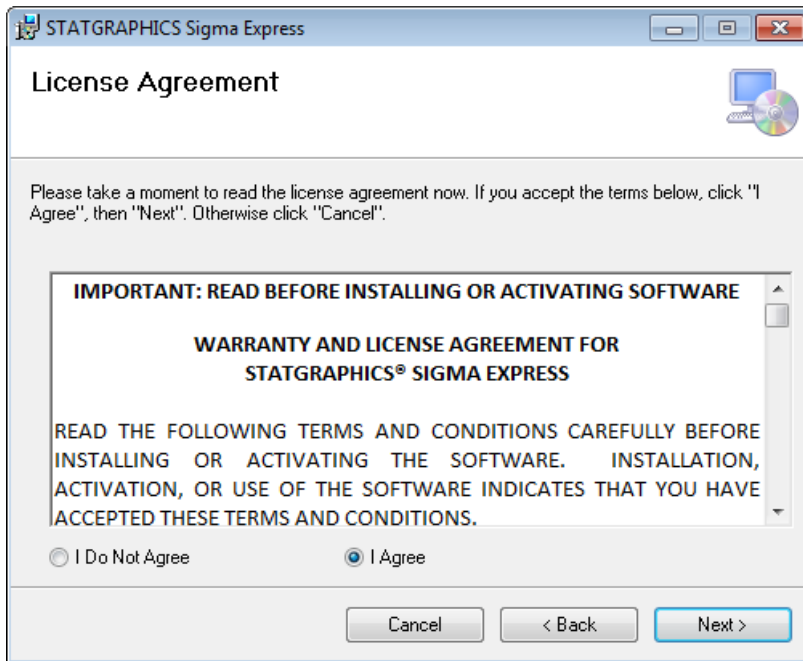
Figure 1-3. License Agreement Dialog Box

Read the license agreement carefully. If you accept the terms, click on the indicated radio button and press *Next* to continue. If you do not agree, press *Cancel*. If you do not agree with the terms, you may not use the program.

Step 5: The next dialog box requests information about the location where the program files will be installed:
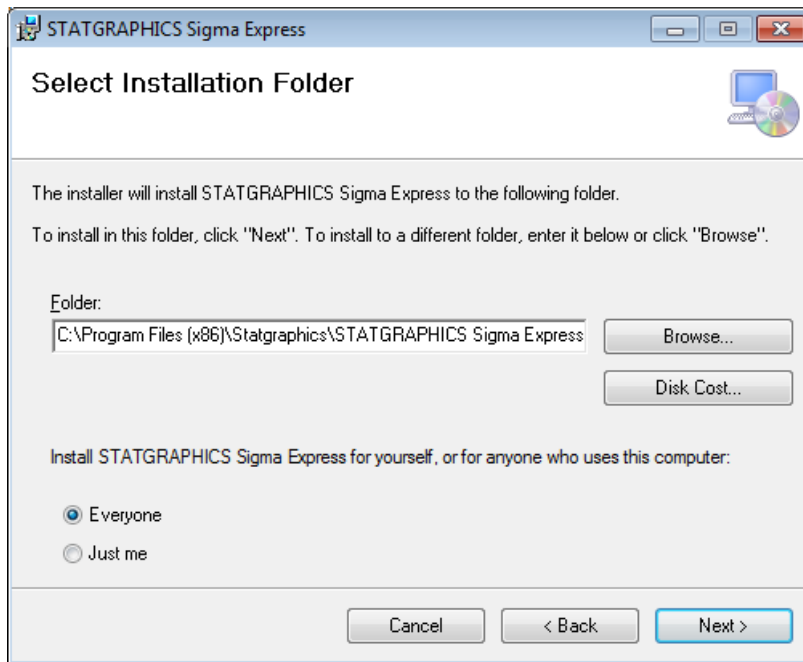
Figure 1-4. Customer Information Dialog Box

By default, STATGRAPHICS is installed in a subdirectory of *Program Files* if you are running a 32-bit version of Windows, or in a subdirectory of *Program Files (x86)* if you are running a 64-bit version of Windows. You may also allow anyone who uses the computer to have access to STATGRAPHICS by selecting the "Everyone" radio button, or restrict access to only the current user.

Step 6: The final dialog box asks you to confirm that you are ready to proceed with the installation:
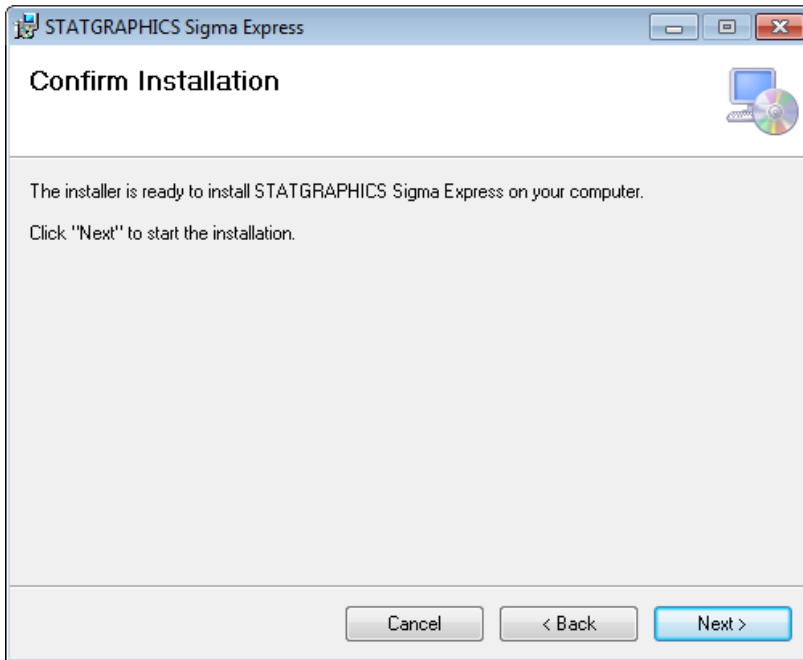
Figure 1-5. Confirmation Dialog Box

Press *Next* to begin installing the program files.

Step 7: When the installation is complete, a final dialog box will be displayed:
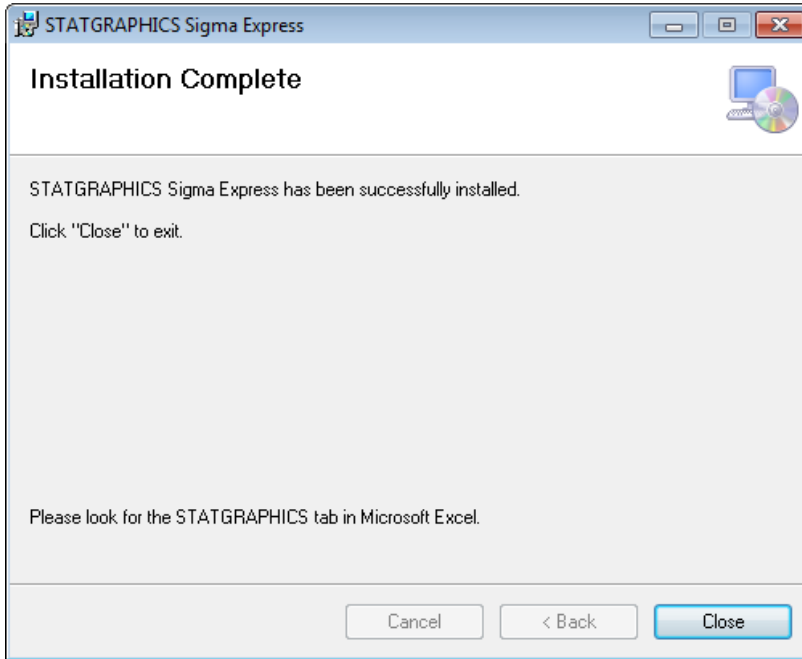
Figure 1-6. Final Installation Dialog Box

Click on *Close* to complete the installation.

1.2 Evaluation Period

The installation program will install STATGRAPHICS *Sigma Express* and add it to the Excel menu. When you next launch Excel, you will see a menu option labeled STATGRAPHICS:
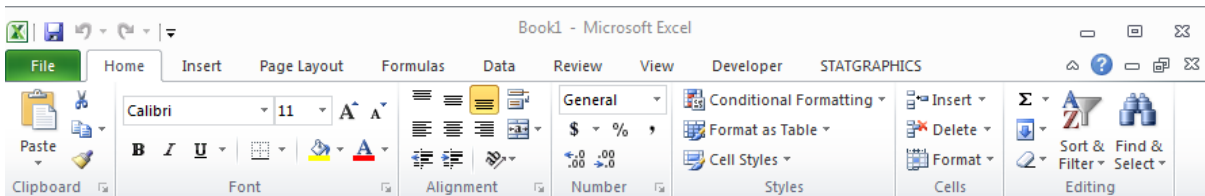
Figure 1-7. Excel 2010 Menu with STATGRAPHICS Menu Item

NOTE: In some cases, STATGRAPHICS will not automatically be added to the Excel menu. This happens for example when Microsoft Office 14 is installed by downloading it from the Internet, since it is then installed in a virtual location. If "STATGRAPHICS"

does not automatically appear on the Excel menu, it can be easily added manually by following the instructions at www.statgraphics.com/express/manualactivation.htm.

If you click on STATGRAPHICS, a menu similar to that shown below will appear:
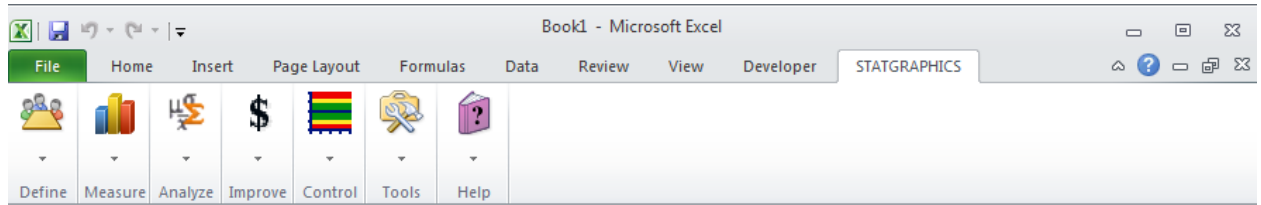
Figure 1-8. STATGRAPHICS Menu as Displayed in Excel 2010

The STATGRAPHICS procedures are divided into 7 categories, including the 5 categories of the Six Sigma DMAIC paradigm (*Define*, *Measure*, *Analyze*, *Improve*, *Control*) plus additional sections for *Tools* (which includes Monte Carlo Simulation) and *Help*.

When you first install STATGRAPHICS, you will be given 15 days to evaluate the program before having to purchase a license. To display the status of your evaluation, select *Help - About* from the STATGRAPHICS menu to display the dialog box shown below:
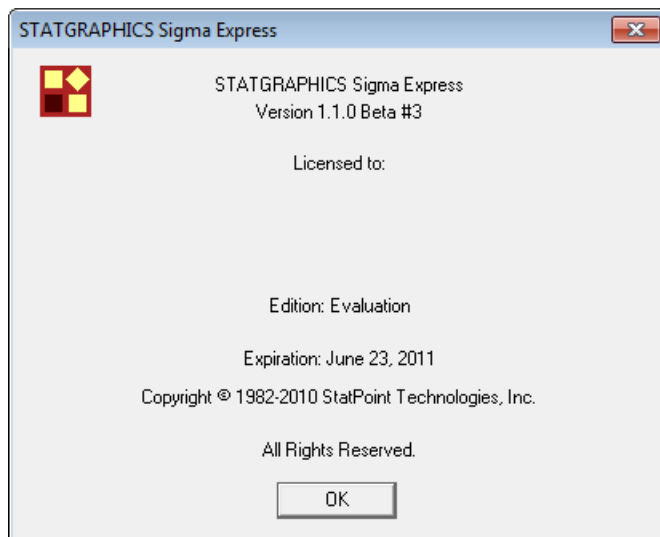
Figure 1-9. Help - About Dialog Box Showing Expiration Date

The dialog box shows the current version number, the edition, and the expiration date.

1.3 Entering a Serial Number and Activation Code

NOTE: To perform the steps described below, you must execute Excel with administrator rights. If you are using Windows Vista or Windows 7, be sure to click on the Excel icon with the RIGHT button mouse (not the left button) and select *Run as administrator* from the popup menu.

In order to continue using STATGRAPHICS after the end of the evaluation period, you must purchase a license. When you purchase a license, you will receive a serial number that must be entered into the license manager. To do so, select *License Manager* from the *Help* menu. This will display the dialog box shown below:
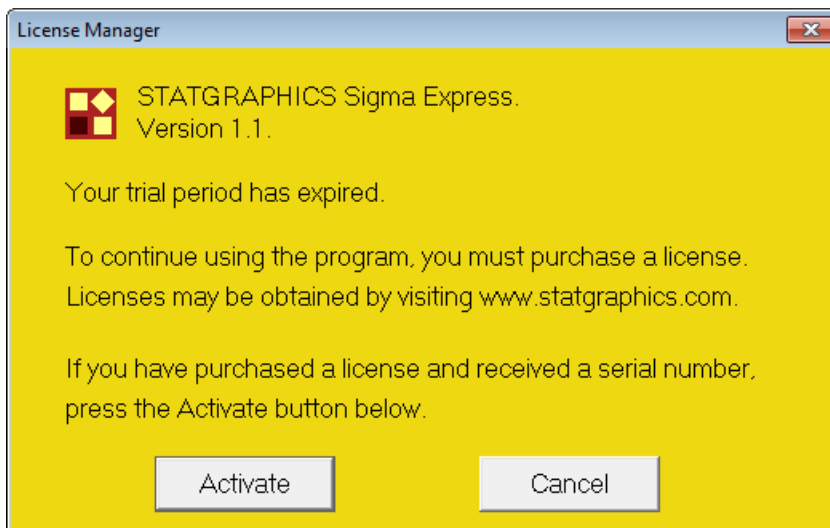
Figure 1-10. License Manager Dialog Box

Press the *Activate* button to display the following activation dialog box:

Figure 1-11. STATGRAPHICS Sigma Express Activation Dialog Box

Follow the three steps below to activate the program.

Step 1 - Enter personal information and serial number

Enter the required information, including your serial number. As soon as you enter a valid serial number, the *Product key* field will be filled with a 16-character key that is unique to your computer. To activate the program, you will need to enter an *Activation Code* that matches that product key.

Step 2 - Request activation code

To receive an Activation Code, press either of the two buttons under Step 2:

1. The button labeled *1. Press Here* automatically sends your Product Key to StatPoint Technologies over the Internet requesting an Activation Code. A web service will respond to that request immediately, sending the Activation Code to the e-mail address that you supply.

2. The button labeled *2. Press Here* accesses your default e-mail program, placing the information in a new e-mail that you can send to StatPoint. E-mail requests will be processed during normal business hours.

To avoid delay, use the first method whenever possible.

> **NOTE: Users who are activating copies obtained under an institutional site license must use the first method. Activation codes will only be sent to e-mail addresses issued by the institution that purchased the site license. Institutional IT personnel should advise users of this requirement.**

Step 3 - Enter activation code

Once your request is processed, an e-mail will be sent to you containing the Activation Code. Enter the code in the field provided under *Step 3* and press the *Activate* button. If the code matches your product key, you will see the following message:
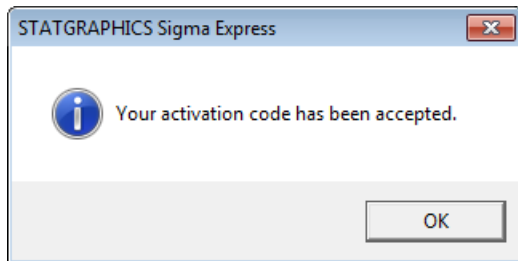

Figure 1-12. Activation Message

Press *OK* to continue. STATGRAPHICS has now been activated and you can now use any of the procedures.

> **NOTE: If you install STATGRAPHICS on more than one computer, you will need a different activation code for each computer, since it must match the product key.**

Template Procedures

Using STATGRAPHICS template procedures.

There are two types of procedures available on the STATGRAPHICS menu:

- *Template procedures* - These procedures create a new Excel worksheet and load a template that you complete by editing the worksheet. The data required by these procedures is typed directly into the template.

- *Data analysis procedures* - These procedures operate on data that you have already entered into columns of your Excel WorkBook. They display several dialog boxes for selecting the data and other options and then create a new window with tables and graphs. After examining the output and making any desired changes, tables and graphs may be copied back to the Excel WorkBook or into other applications such as Microsoft Word or PowerPoint.

This chapter discusses use of the template procedures.

2.1 Selecting a Template Procedure

To use a template procedure, select the desired item from the STATGRAPHICS menu. For example, suppose you wish to create a priority matrix, which is used to help select between alternative solutions to a problem. To do so:

> **Step 1**: First load any Excel workbook.

> **Step 2**: Select *Priority Matrix* from the *Measure* menu. This will add a new worksheet to your WorkBook and load the template shown below:
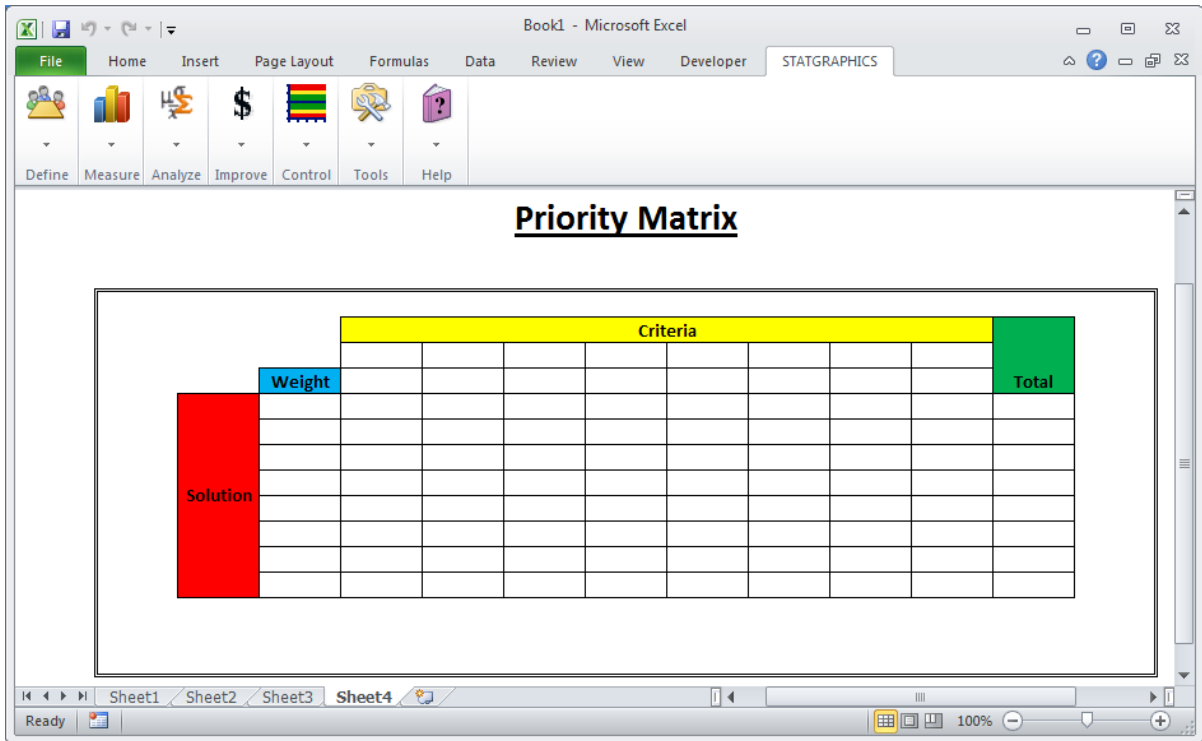
Figure 2-1. Priority Matrix Template

2.2 Completing the Template

To create the priority matrix, type entries directly into the cells of the worksheet containing the template. To do so:

1. Develop a list of s *solutions*. Each solution is assigned to one row of the matrix. Enter a label for each solution in the first open column.

2. Develop a list of c *criteria* on which each solution will be judged. Criteria are assigned to columns of the matrix. Enter a label for each criterion in the first open row.

3. Assign weights to the criteria, W_1, W_2, ..., W_c, the relative values of which express the importance of each criteria. Enter the values of the weights in the second open row.

4. Score each solution with respect to each criteria by placing a value $S_{i,j}$ in the appropriate cell of the matrix.

A typical completed priority matrix is shown below:

Priority Matrix

	Criteria								
	Safety	Time	Sigma	Legal	Cost	Benefit	People		
Weight	3	2	1	1	3	3	2		**Total**
A	6	8	8	10	6	7	4		99
B	9	6	9	10	7	5	5		104
C	5	7	5	9	6	4	1		75
D	4	9	7	8	3	4	2		70

Figure 2-2. Completed Priority Matrix

The worksheet will automatically calculate a total for each solution according to:

$$T_s = \sum_{j=1}^{c} S_{s,j} W_j \qquad (2.1)$$
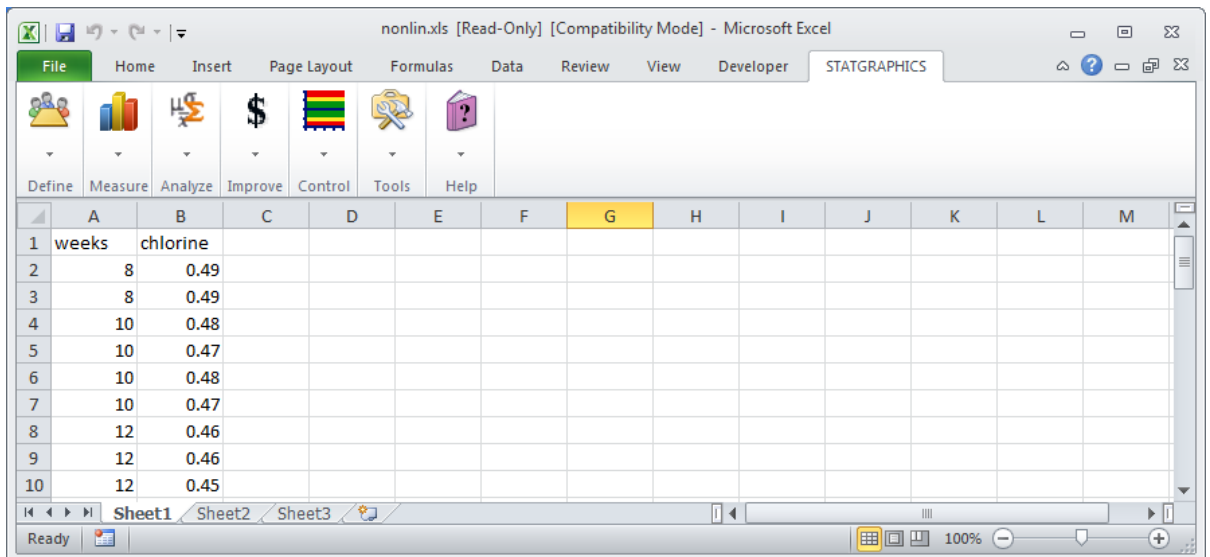
In the above template, Solution B has the highest total.

To save the results of a template procedure, simply save the current Excel WorkBook.

Data Analysis Procedures

Generating an analysis, selecting tables and graphs, specifying options, and saving the results.

Data analysis procedures analyze data that you have already placed in an Excel worksheet. As an example, load the file named *nonlin.xls* which was copied to the Program Files directory when STATGRAPHICS Sigma Express was installed. The data show the measurements made on the available chlorine in samples of a product taken at different lengths of time after production (from *Applied Regression Analysis*, third edition, by Draper and Smith, 1998):

Figure 3-1. Sample Data File from a Stability Study

To analyze data using *STATGRAPHICS Sigma Express*, it is usually placed in columns beginning with a header row containing names for each column. A second header row may be included if desired containing comments for each column. The comment row is often used to display the units of measurement. Subsequent rows contain the data to be analyzed.

Note that:

1. Each column name should be unique and may contain up to 32 characters. Names may contain any characters except a quote (') or a double quote ("). Spaces are allowed.

2. Column names do not have to be placed in the first row. However, the data should follow in the rows immediately below the names.

The sample data set has *n* = 44 observations for *chlorine* together with its age measured in *weeks*. It contains a single header row with column names but no comments.

3.1 Selecting a Data Analysis Procedure

To execute a data analysis procedure:

Step 1: Select the data to be analyzed by highlighting it in Excel. If no data is selected, it is assumed that the entire contents of the current worksheet is desired.

Step 2: Select the desired procedure from the STATGRAPHICS menu. To fit a curve relating *chlorine* to *weeks*, select *Analyze - Regression Analysis - Simple Regression*.

Step 3: Most procedures begin by displaying 2 dialog boxes which are used to specify the columns containing the data to be analyzed. The first dialog box specifies the structure of the Excel worksheet:
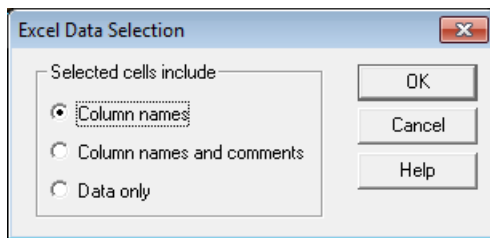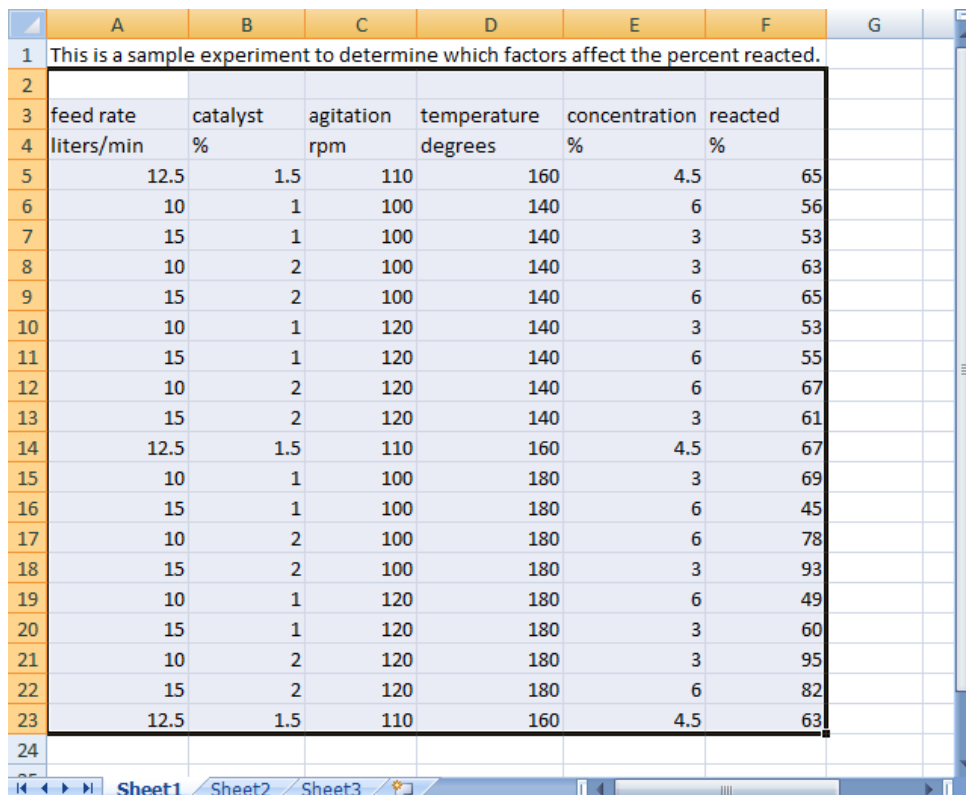

Figure 3-2. Excel Data Selection Dialog Box

There are 3 choices:

- *Column names* – indicates that the first non-empty row of the Excel selection contains names for each of the columns.

- *Column names and comments* - indicates that the first non-empty row of the Excel selection contains names for each of the columns, while the row immediately below the names contains a comment.

- *Data only* – indicates that all cells contain data. In this case, column names will be automatically generated to match the column indicators in Excel ("A", "B", etc.).

Note that column names do not have to be placed in row #1. However, they must be located in the first non-empty row of the selected range. As an example, the Excel worksheet below shows a valid structure containing both column names and comments:

	A	B	C	D	E	F	G
1	This is a sample experiment to determine which factors affect the percent reacted.						
2							
3	feed rate	catalyst	agitation	temperature	concentration	reacted	
4	liters/min	%	rpm	degrees	%	%	
5	12.5	1.5	110	160	4.5	65	
6	10	1	100	140	6	56	
7	15	1	100	140	3	53	
8	10	2	100	140	3	63	
9	15	2	100	140	6	65	
10	10	1	120	140	3	53	
11	15	1	120	140	6	55	
12	10	2	120	140	6	67	
13	15	2	120	140	3	61	
14	12.5	1.5	110	160	4.5	67	
15	10	1	100	180	3	69	
16	15	1	100	180	6	45	
17	10	2	100	180	6	78	
18	15	2	100	180	3	93	
19	10	1	120	180	6	49	
20	15	1	120	180	3	60	
21	10	2	120	180	3	95	
22	15	2	120	180	6	82	
23	12.5	1.5	110	160	4.5	63	
24							

Sheet1 / Sheet2 / Sheet3

Figure 3-3. Sample Excel Worksheet with Column Names and Comments

Note that the selected range begins in row #2, so that the column names are located in the first non-empty row of the selection.

The second dialog box requests the names of the columns to be used by the analysis procedure. When performing a *Simple Regression*, the dialog box appears as shown below:
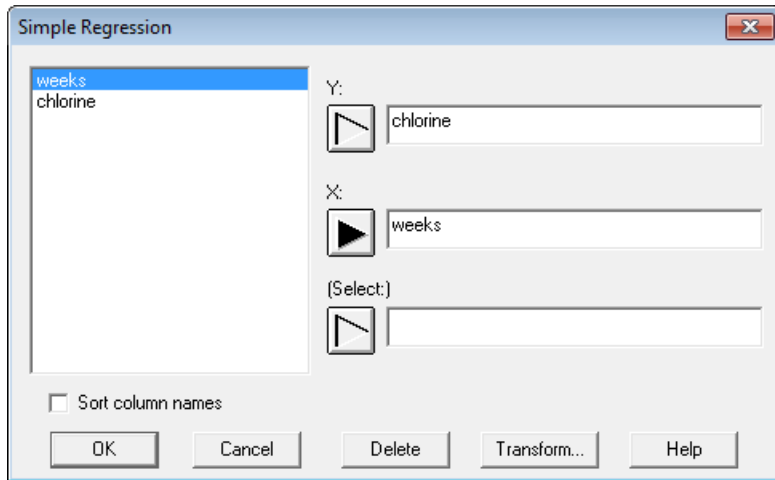

Figure 3-4. Data Input Dialog Box for Simple Regression Procedure

The first two input fields are required:

> **Y**: The dependent or response variable.

> **X**: The independent or predictor variable.

In data entry fields, you can enter either the name of a column such as *chlorine* or a STATGRAPHICS expression such as *LOG(chlorine)*. The *Select* field may be used to select a subset of the highlighted rows. For example, if you enter a statement such as *FIRST(20)* in that field, only the first 20 rows in the worksheet selection will be used. The *Select* field is discussed in more detail later.

Step 4: Most procedures will next display an *Analysis Options* dialog box which is used to specify options for the procedure to be run. For *Simple Regression*, it has the following form:
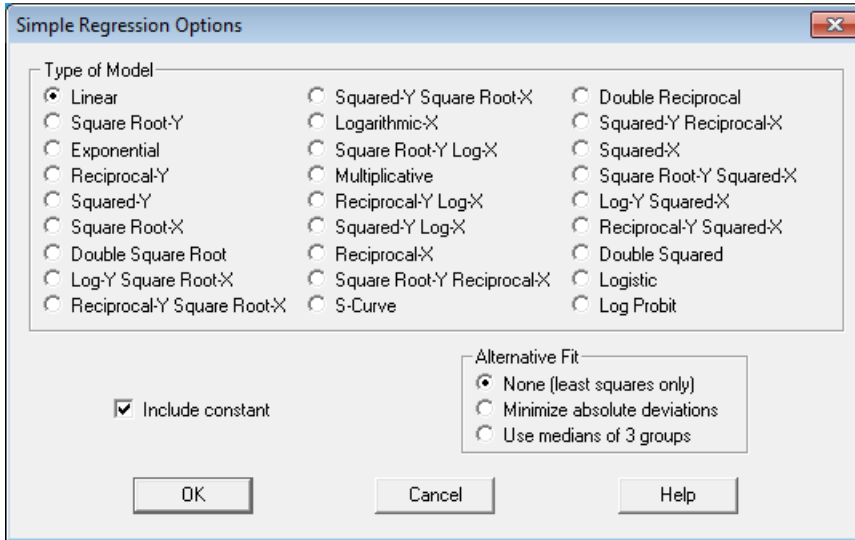
Figure 3-5. Analysis Options Dialog Box for Simple Regression Procedure

The above dialog box allows you to select a model other than a straight line if desired. Usually, the default selections will be adequate when first running the procedure. The options can be easily changed after the initial fit is examined.

Step 5: After the options are selected and *OK* is pressed, a final dialog box will be displayed showing the available tables and graphs:
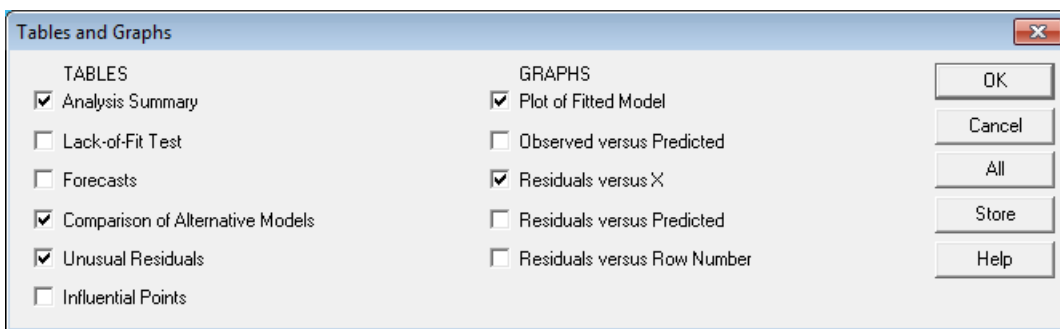
Figure 3-6. Tables and Graphs Dialog Box for Simple Regression Procedure

For this example, 3 tables and 2 graphs have been selected. This selection can also be easily changed after the initial analysis is examined.

Pressing *OK* will start the procedure.

3.2 Analysis Windows

Once the data and options have been specified, an *analysis window* will be created:
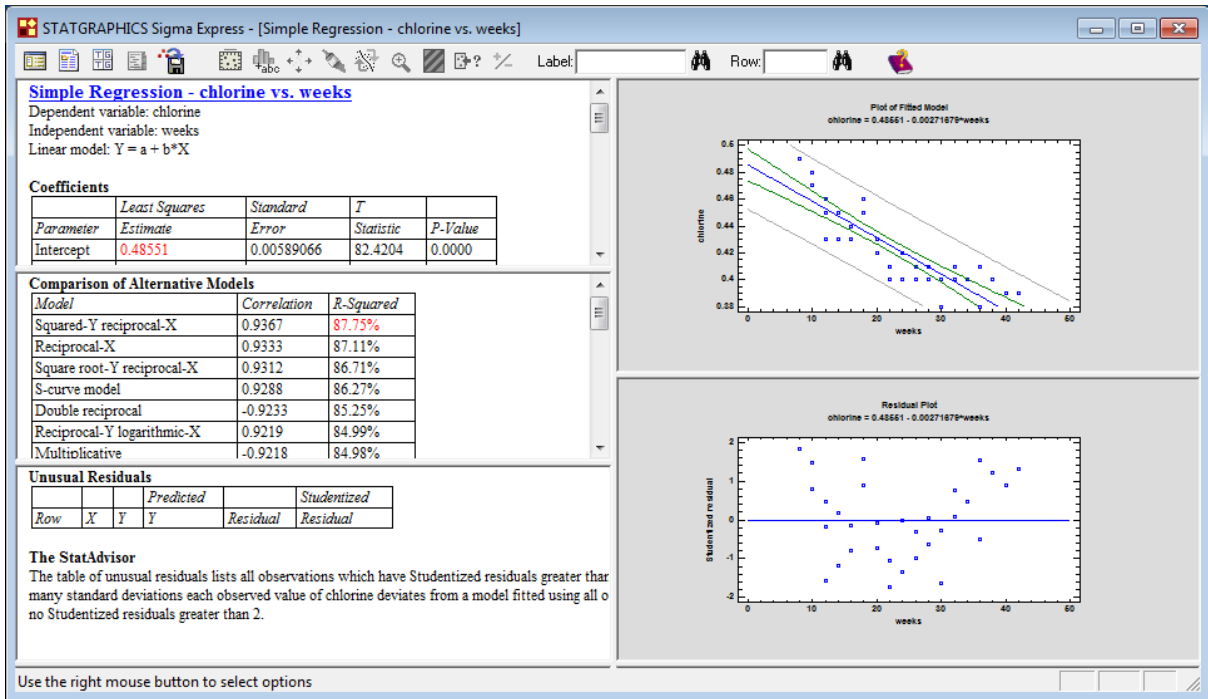
Figure 3-7. Analysis Window for Simple Regression Example

The analysis window is a "splitter window", with multiple *panes* divided by movable splitter bars. Tables are located along the left side of the window, while graphs are located along the right.

You can maximize any table or graph by double-clicking on it, in which case it will fill the window:
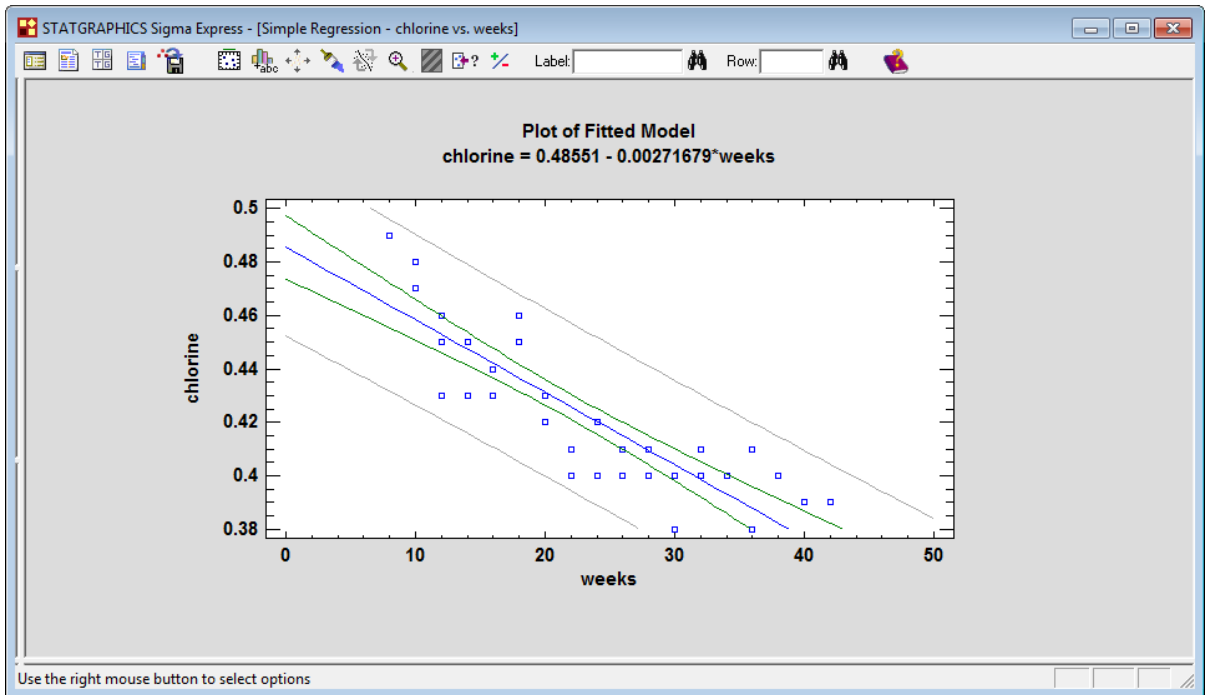
Figure 3-8. Simple Regression Analysis Window with Maximized Pane

Double-clicking on the pane a second time restores the multiple pane display.

3.2.1 Copying Graphs to Excel

Once a graph has been created, it can be copied to your Excel worksheet (or any other program) by pressing the right-mouse button anywhere on the graph and selecting *Copy* from the pop-up menu:

Figure 3-9. Graphics Pop-up Menu Accessed by Right Mouse Button

This places the graph on the Windows clipboard. If you return to Excel, select the cell where the graph should be pasted and then select *Paste*, it will be placed in the worksheet as shown below:

Figure 3-10. STATGRAPHICS Graph after Pasting into Excel Worksheet

Since the graph is pasted as a picture, it can be easily moved or resized.

3.2.2 Copying Tables to Excel

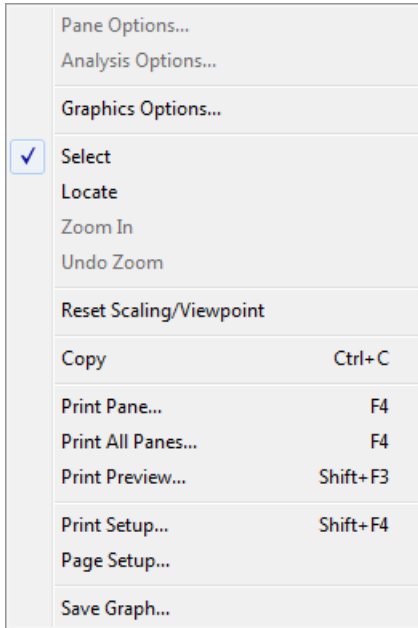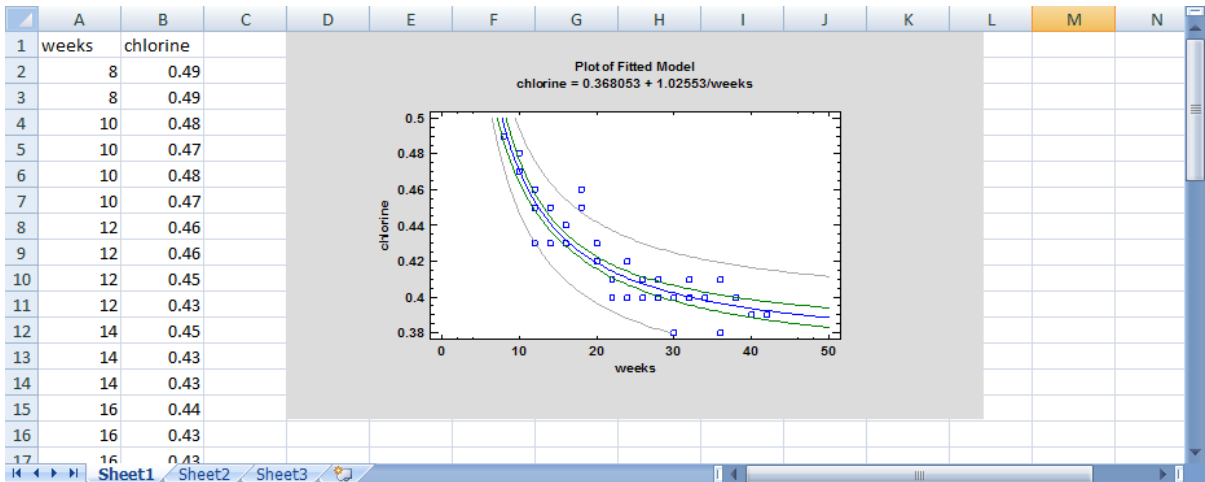Tables may be handled in a similar fashion. For example, double-click on the graph to minimize it and then double-click on the upper left pane to maximize the *Analysis Summary*:
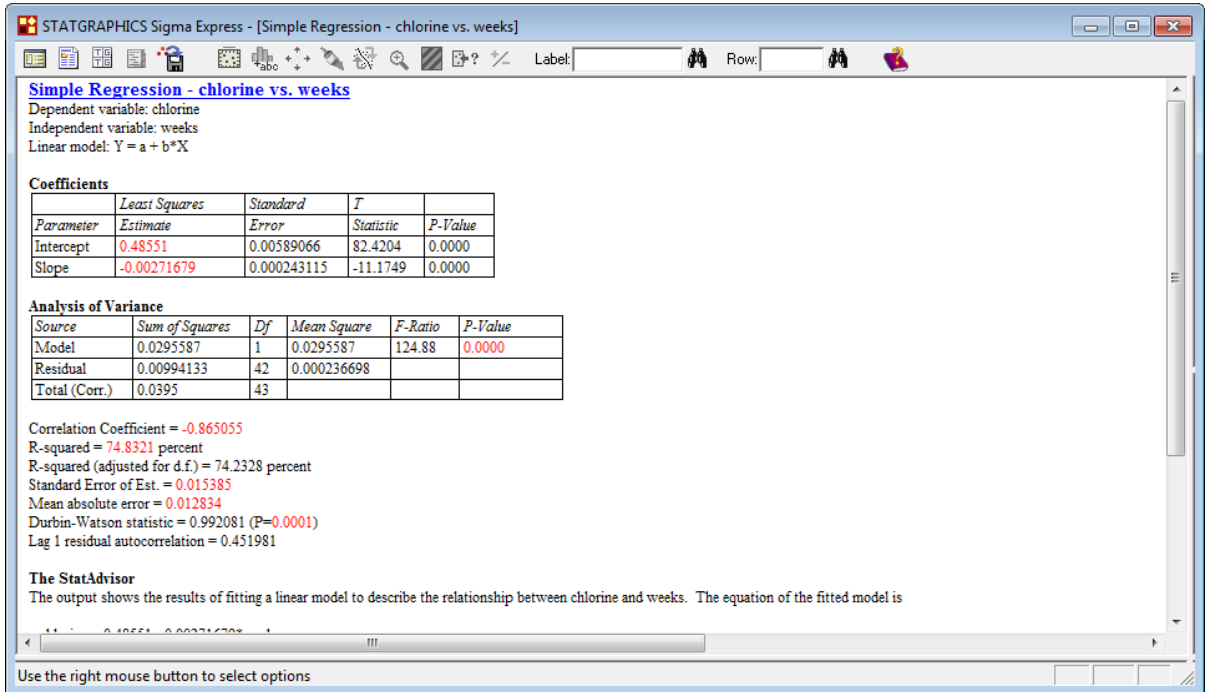
STATGRAPHICS Sigma Express - [Simple Regression - chlorine vs. weeks]

Simple Regression - chlorine vs. weeks
Dependent variable: chlorine
Independent variable: weeks
Linear model: Y = a + b*X

Coefficients

Parameter	Least Squares Estimate	Standard Error	T Statistic	P-Value
Intercept	0.48551	0.00589066	82.4204	0.0000
Slope	-0.00271679	0.000243115	-11.1749	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	0.0295587	1	0.0295587	124.88	0.0000
Residual	0.00994133	42	0.000236698		
Total (Corr.)	0.0395	43			

Correlation Coefficient = -0.865055
R-squared = 74.8321 percent
R-squared (adjusted for d.f.) = 74.2328 percent
Standard Error of Est. = 0.015385
Mean absolute error = 0.012834
Durbin-Watson statistic = 0.992081 (P=0.0001)
Lag 1 residual autocorrelation = 0.451981

The StatAdvisor
The output shows the results of fitting a linear model to describe the relationship between chlorine and weeks. The equation of the fitted model is

Use the right mouse button to select options

Figure 3.11. Simple Regression Analysis Summary Table

To copy this output to your Excel worksheet, highlight the output that you wish to copy and press the right-mouse button to display the following pop-up menu:

Figure 3-12. Tables Pop-up Menu Accessed by Right Mouse Button

There are 2 relevant options:

1. *Copy* - select this option to copy the data as numeric values with associated text. When pasted into Excel, the table structure will be preserved so that you can access individual values.

2. *Copy as Metafile* - copies the output as a picture. It will then appear exactly as in STATGRAPHICS, but you will not be able to access individual values.

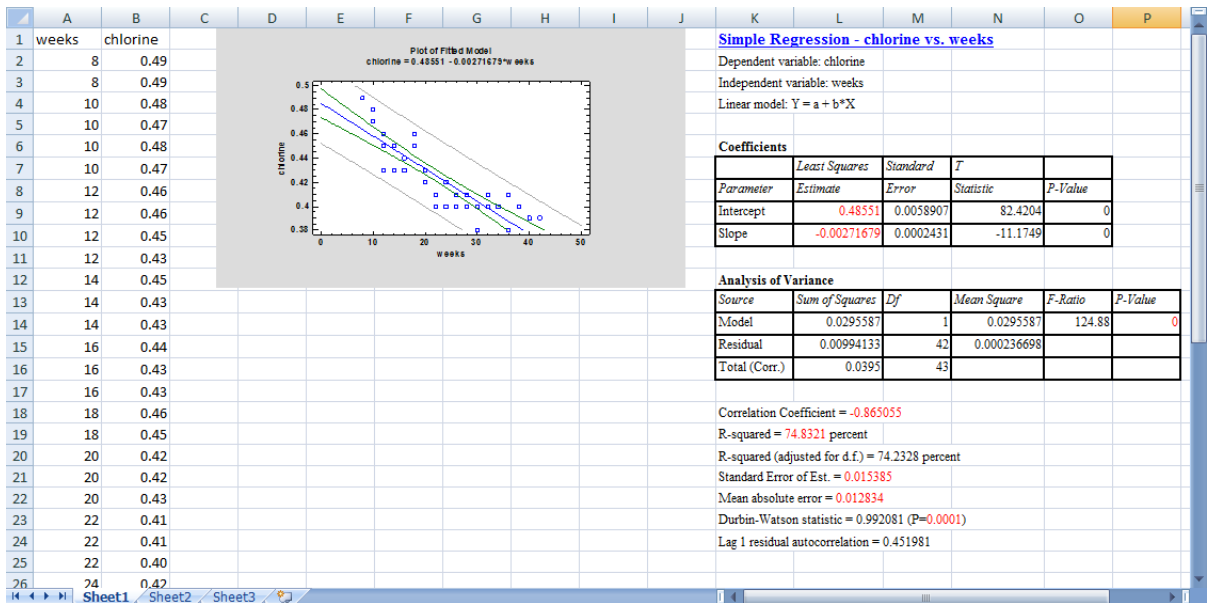The figure below shows the output pasted using the first method:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	weeks	chlorine									**Simple Regression - chlorine vs. weeks**					
2	8	0.49									Dependent variable: chlorine					
3	8	0.49									Independent variable: weeks					
4	10	0.48									Linear model: Y = a + b*X					
5	10	0.47														
6	10	0.48									**Coefficients**					
7	10	0.47										Least Squares	Standard	T		
8	12	0.46									Parameter	Estimate	Error	Statistic	P-Value	
9	12	0.46									Intercept	0.48551	0.0058907	82.4204	0	
10	12	0.45									Slope	-0.00271679	0.0002431	-11.1749	0	
11	12	0.43														
12	14	0.45									**Analysis of Variance**					
13	14	0.43									Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
14	14	0.43									Model	0.0295587	1	0.0295587	124.88	0
15	16	0.44									Residual	0.00994133	42	0.000236698		
16	16	0.43									Total (Corr.)	0.0395	43			
17	16	0.43														
18	18	0.46									Correlation Coefficient = -0.865055					
19	18	0.45									R-squared = 74.8321 percent					
20	20	0.42									R-squared (adjusted for d.f.) = 74.2328 percent					
21	20	0.42									Standard Error of Est. = 0.015385					
22	20	0.43									Mean absolute error = 0.012834					
23	22	0.41									Durbin-Watson statistic = 0.992081 (P=0.0001)					
24	22	0.41									Lag 1 residual autocorrelation = 0.451981					
25	22	0.40														
26	24	0.42														

Plot of Fitted Model
chlorine = 0.48551 - 0.00271679*weeks

Sheet1 / Sheet2 / Sheet3

Figure 3-13. STATGRAPHICS Table after Pasting into Excel Worksheet

Some resizing of rows and columns may be necessary.

3.2.3 Analysis Toolbar

Along the top of the STATGRAPHICS analysis window is an *analysis toolbar*, shown below:

Label: Row:
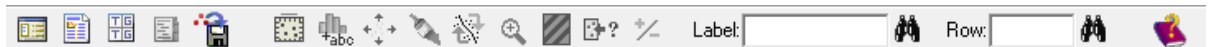
Each of the buttons on this toolbar performs an important operation.

Data Input Dialog Button

When pressed, this button displays the data input dialog box originally used to specify the data variables. If you change the data variables and press *OK*, the analysis will change to reflect the new selections. This enables you to try different combinations of data without having to start a new analysis.

Analysis Options Button

This button redisplays the *Analysis Options* dialog box shown in Figure 3.2. It could be used in this case to fit a different type of model, such as the *Reciprocal-X* model shown below:
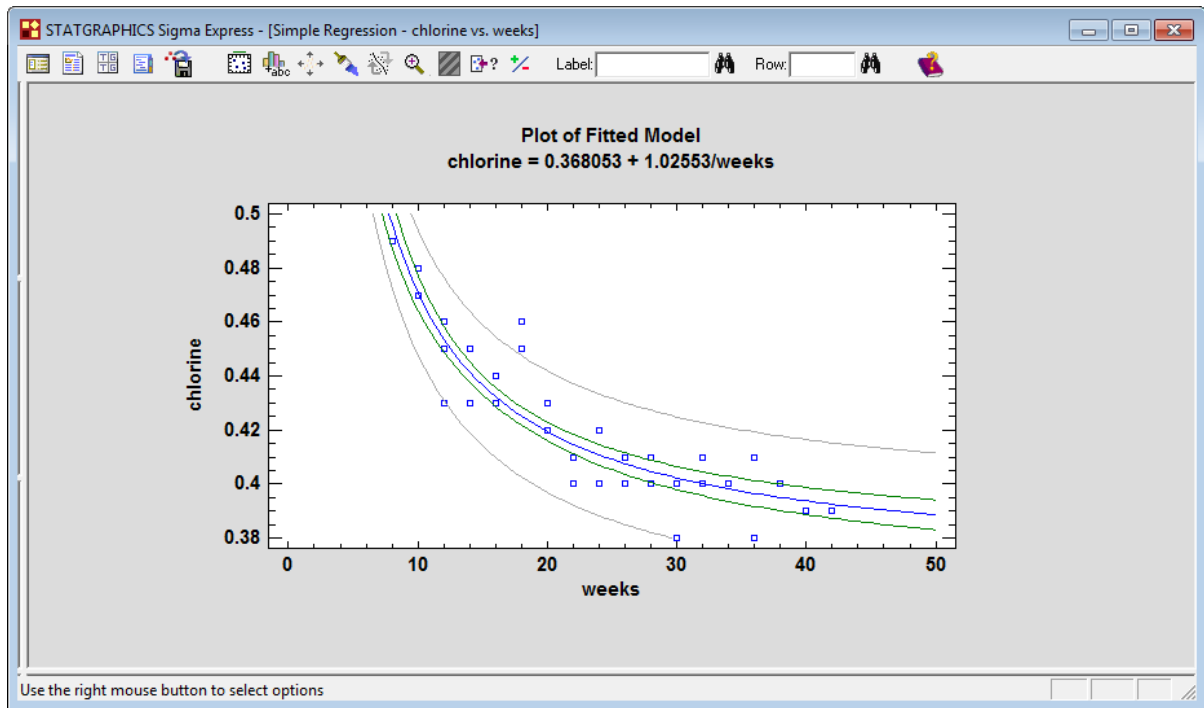

Figure 3-14. Fitted Reciprocal-X Model

Tables and Graphs Button

This button redisplays the list of tables and graphs that may be added to the analysis window.

Pane Options Button

In addition to options that apply to the entire analysis window, many individual tables and graphs have options that apply only to them. These options are accessible by first maximizing the selected table or graph and then pressing *Pane Options*. For the *Fitted Model Plot*, displayed above, the pane options are:
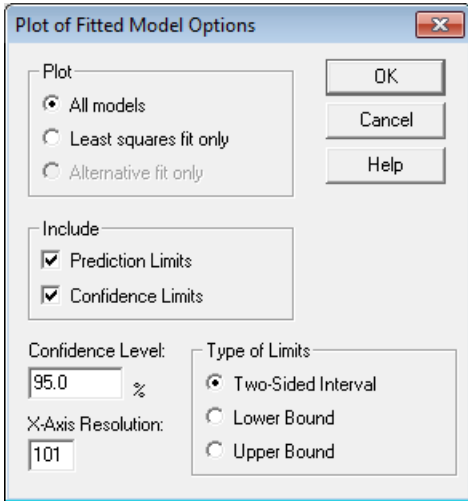
Figure 3-15. Pane Options Dialog Box for the Fitted Model Plot

For example, removing the check mark alongside *Confidence Limits* and pressing *OK* will replot the graph without the inner limits:
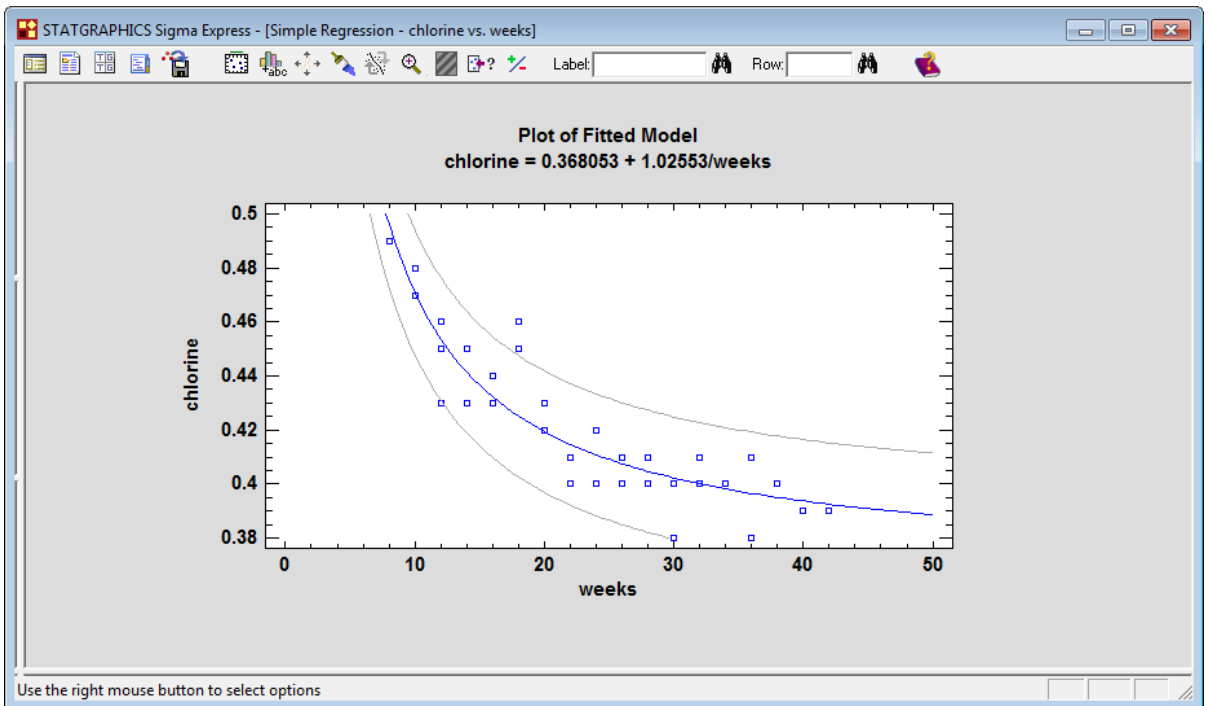
Figure 3-16. Fitted Model Plot Without Confidence Limits

Save Results Button

This button allows you to save numerical results calculated by the statistical analysis back to columns of your worksheet. For *Simple Regression*, it displays the following choices:
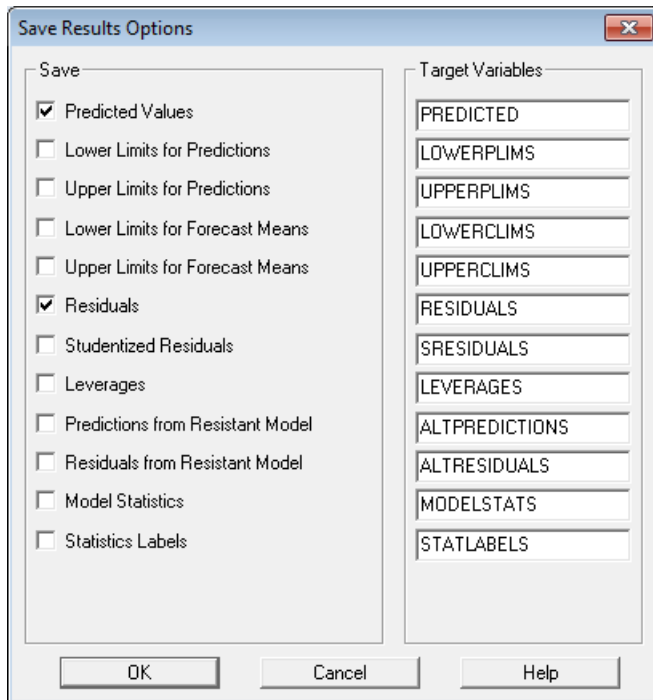
Figure 3-17. Simple Regression Save Results Dialog Box

To save information:

1. Check the items of interest in the *Save* field.

2. For each item to be saved, assign a column name under *Target Variables*.

3. Press *OK*. This will put the selected data on the Windows clipboard.

4. Return to your Excel worksheet, highlight the cell where you wish the results to be pasted, and select *Paste*.

This will place the data into the Excel worksheet, as shown below:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	weeks	chlorine	PREDICTED	RESIDUALS									
2	8	0.49	0.495428	-0.0054282									
3	8	0.49	0.495428	-0.0054282									
4	10	0.48	0.469791	0.0102093									
5	10	0.47	0.469791	0.0002093									
6	10	0.48	0.469791	0.0102093									
7	10	0.47	0.469791	0.0002093									
8	12	0.46	0.452699	0.00730095									
9	12	0.46	0.452699	0.00730095									
10	12	0.45	0.452699	-0.0026991									
11	12	0.43	0.452699	-0.022699									
12	14	0.45	0.440491	0.00950927									
13	14	0.43	0.440491	-0.0104907									
14	14	0.43	0.440491	-0.0104907									
15	16	0.44	0.431334	0.00866551									
16	16	0.43	0.431334	-0.0013345									
17	16	0.43	0.431334	-0.0013345									
18	18	0.46	0.424213	0.035787									
19	18	0.45	0.424213	0.025787									
20	20	0.42	0.418516	0.00148424									

Sheet1 / Sheet2 / Sheet3

Figure 3-18. Excel Worksheet with Saved Results

Graphics Buttons

Whenever a graph is maximized within the analysis window, several additional buttons are enabled. These buttons include:

Graphics options – displays a dialog box used to change colors, labels, axis scaling, and other similar features.

Add text – used to add additional text to the graph.

Jitter – used to offset points randomly in the horizontal or vertical direction to prevent their overplotting each other.

Brush – colors points on a scatterplot according to the value of a selected variable.

Smooth/Rotate – smoothes a 2-dimensional plot, or rotates a 3-dimensional plot.

 Pan or zoom – stretches or zooms the graph in the X-, Y-, or Z-direction.

 Explore – dynamically explores response surface and contour plots.

 Identify – displays a label identifying a point when clicked on with the mouse.

 Locate by name – highlights in red any points with values equal to that entered in the *Locate* field (used in conjunction with the *Identify* button).

 Locate by row – highlights in red any points corresponding to the row number entered in the *Row* field.

Each of these buttons is described in detail in Chapter 4.

Exclude Button

 Some statistical procedures allow you to interactively remove suspected outliers from an analysis by maximizing a graph, clicking on the suspect point, and pressing this button. For example, the plot in Figure 3-13 shows one point that is well outside the prediction limits. Clicking on that point and pressing the *Exclude* button causes the model to be refit without the point. The fitted model plot shows the new model, indicating the point (or points) that have been removed with an X:
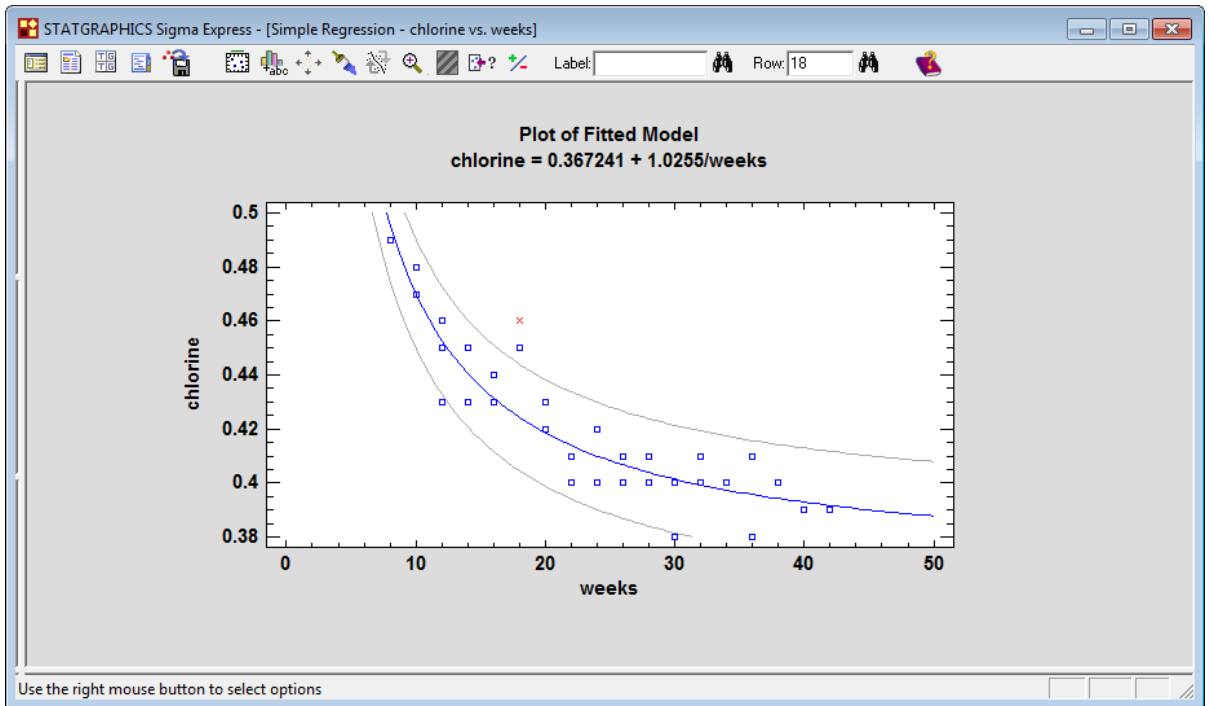
Figure 3-19. Fitted Model after Excluding a Suspected Outlier

All of the other tables and graphs in the analysis window will also change to reflect the new model.

Multiple points may be excluded from a model by clicking on them one at a time and pressing the *Exclude* button. Clicking on a point that has been removed will put it back into the model.

Help Button

Pressing this button will open a PDF file describing the selected analysis. These PDF files contain a description of all dialog boxes and options, provide formulas for important statistical calculations, and illustrate the use of the analysis with a typical example.

3.2.4 Statlets®

There are a number of special menu selections called *Statlets* that are somewhat different than standard analyses. Rather than creating multiple panes with various tables and graphs, they contain only a single graph. In addition, instead of using the *Analysis Options* and *Pane Options* dialog boxes, they put all options on a control bar at the top of the analysis window.

As an example, select *Analyze – Regression Analysis – Curve-Fitting Statlet* from the STATGRAPHICS menu. Complete the data input dialog box as shown below:
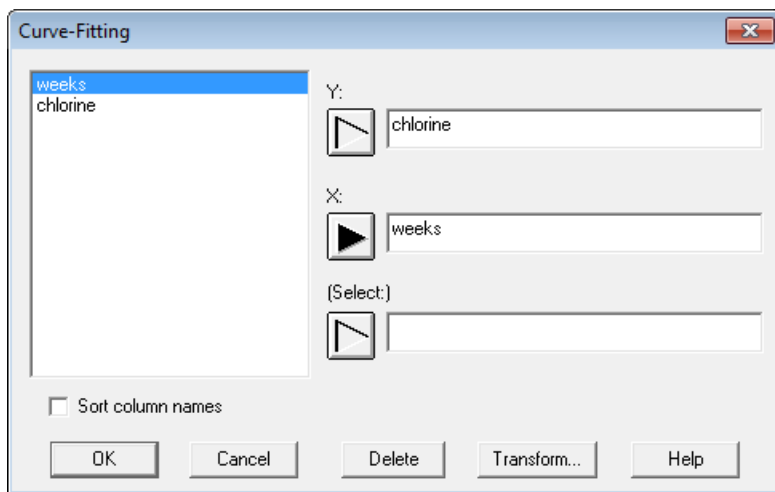
Figure 3-20. Data Input Dialog Box for Curve-Fitting Statlet

The analysis window that appears when *OK* is pressed is shown below:
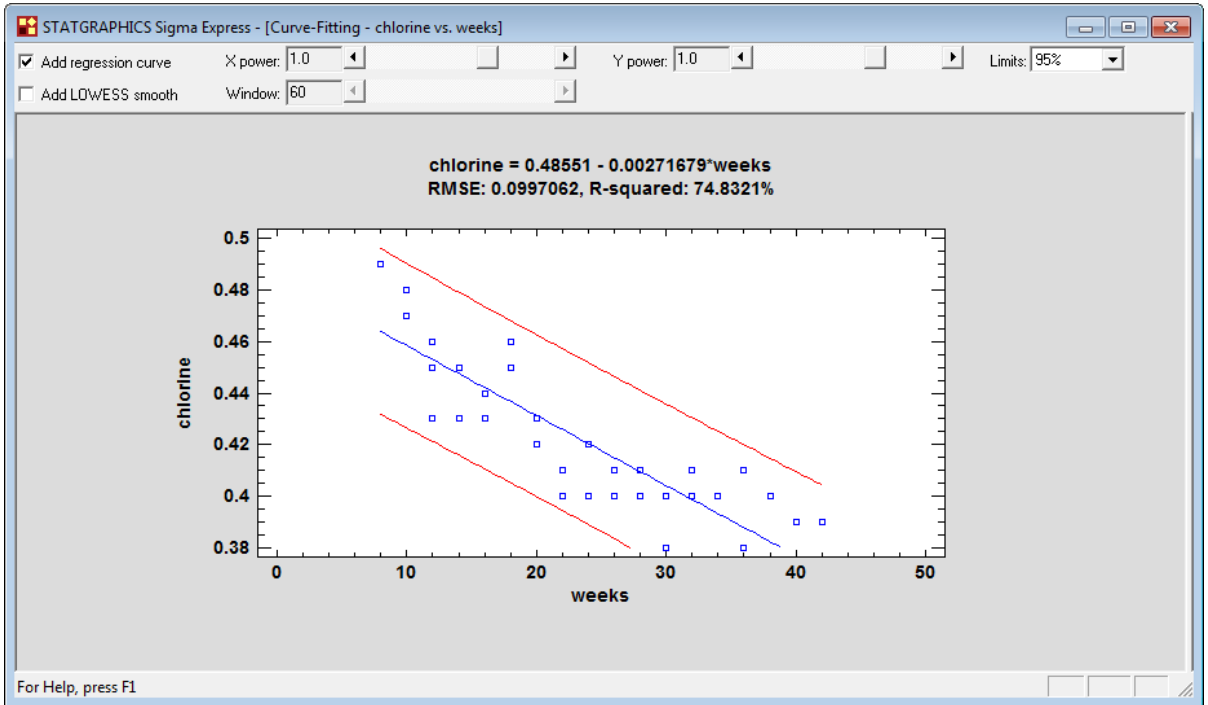
Figure 3-21. Initial Analysis Window for Curve-Fitting Statlet

A linear regression model has been fit to the data. Notice the extra control bar with the following options:

- *Add regression* – draws the fitted regression line.

- *Add LOWESS smooth* – adds a scatterplot smoother to the display.

- *X power* and *Y power* – fits a linear model relating a power of Y to a power of X, according to the equation $Y^p = a + b\, X^q$.

- *Limits* – sets the percentage for the prediction limits.

As changes are made on the dialog box, the graph updates immediately. For example, changing the power of X gives a fit with a much better R-squared:
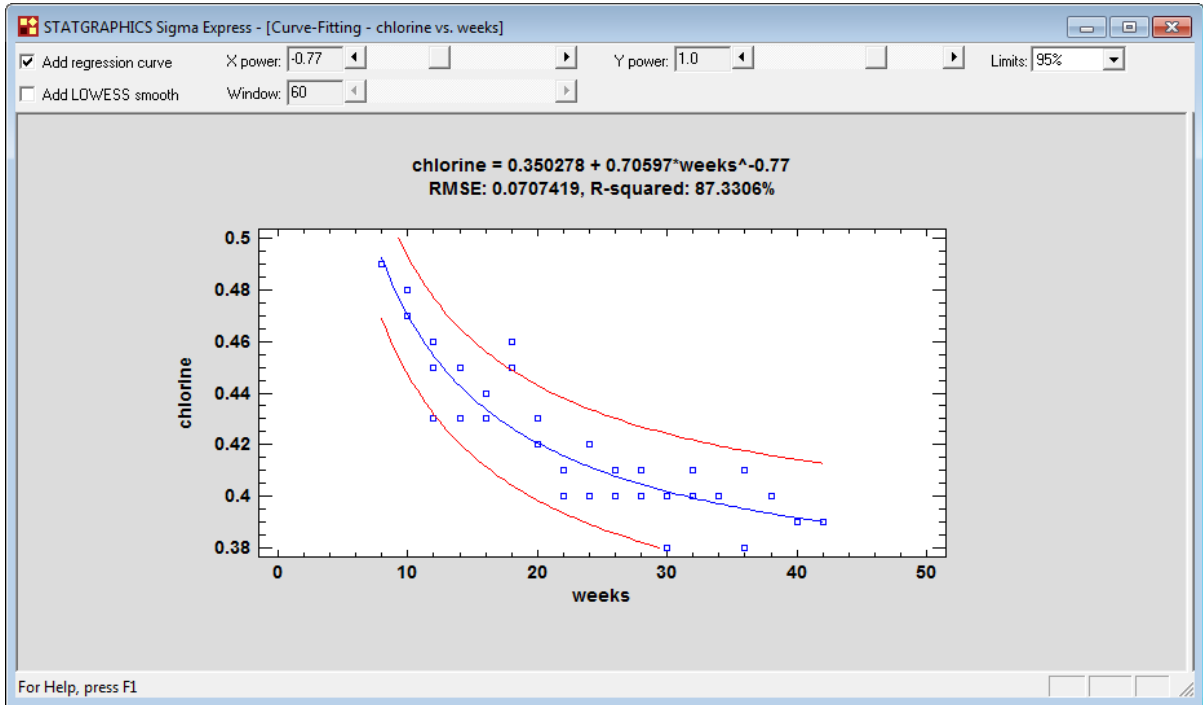
Figure 3-22. Model Fit After Transforming X

Statlets allow you to interact with your data in interesting ways.

3.3 Transformations

As mentioned previously, much flexibility is gained by allowing users to enter algebraic and other expressions on data input dialog boxes. For example, a user wishing to fit a model involving a logarithm or square root may be so directly on a data input box without changing the data in their worksheet. For example, the same reciprocal-X model shown in the previous section could have been fit to the stability study data by fitting a linear model to *chlorine* versus the reciprocal of *weeks*. To do so, the data input dialog box would have been completed as shown below:
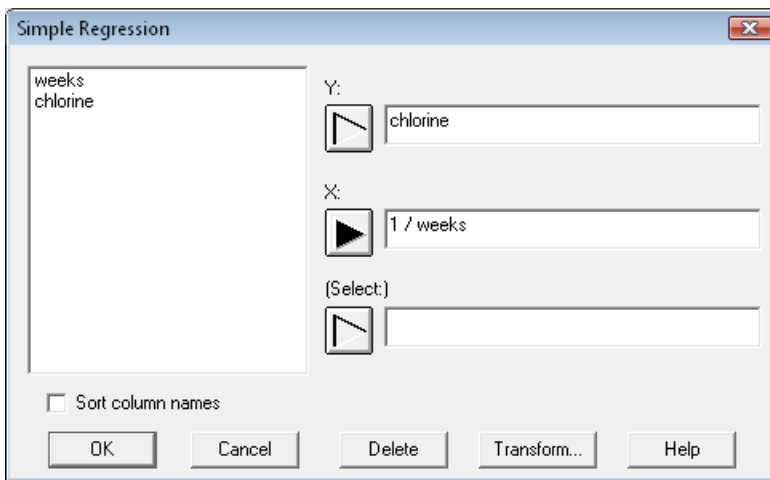
Figure 3-23. Data Input Dialog Box Showing Algebraic Expression for X

Any algebraic expression can be entered in the data input fields, using the following symbols:

Operation	Symbol	Example
Addition	+	weeks+8
Subtraction	-	weeks-2
Multiplication	*	weeks*10
Division	/	weeks/100
Exponentiation	^	weeks^2

Figure 3-24. Algebraic Symbols

When more than one operation is included, normal algebraic precedence rules apply (such as multiplication and division before addition and subtraction).

Expressions may also involve special operators, such as LOG, SQRT or ABS, as in the example below:
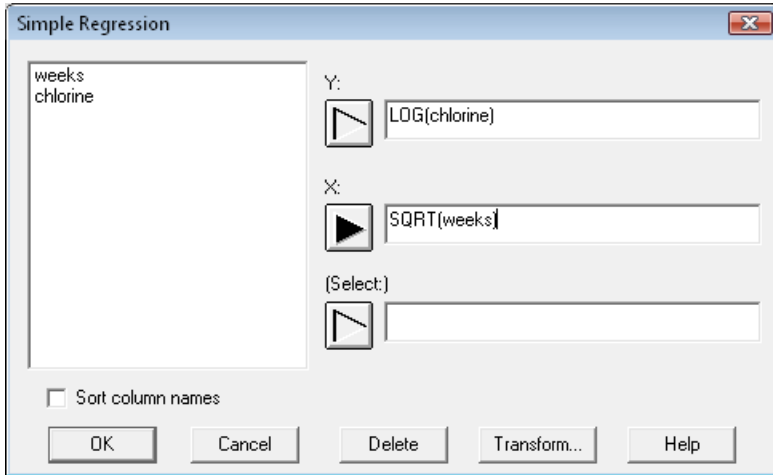
Figure 3-25. Data Input Dialog Box Showing Use of Special Operators

The following table shows the most common operators:

Operator	Use	Example
ABS	Absolute value	ABS(X)
AVG	Average	AVG(X)
DIFF	Backward differencing	DIFF(X)
EXP	Exponential function	EXP(10)
LAG	Lag by k periods	LAG(X,1)
LOG	Natural logarithm	LOG(X)
LOG10	Log base 10	LOG10(X)
MAX	Maximum	MAX(X)
MIN	Minimum	MIN(X)
SD	Standard deviation	SD(X)
SQRT	Square root	SQRT(X)
STANDARDIZE	Conversion to Z-scores	STANDARDIZE(X)

Figure 3-26. Common STATGRAPHICS Operators

3.4 Subset Selections

Every data input dialog box contains an optional input field labeled *(Select:)*. If not blank, this field is used to select a subset of the rows that contain data. For example, to restrict the fit to the first 25 rows of data, enter:
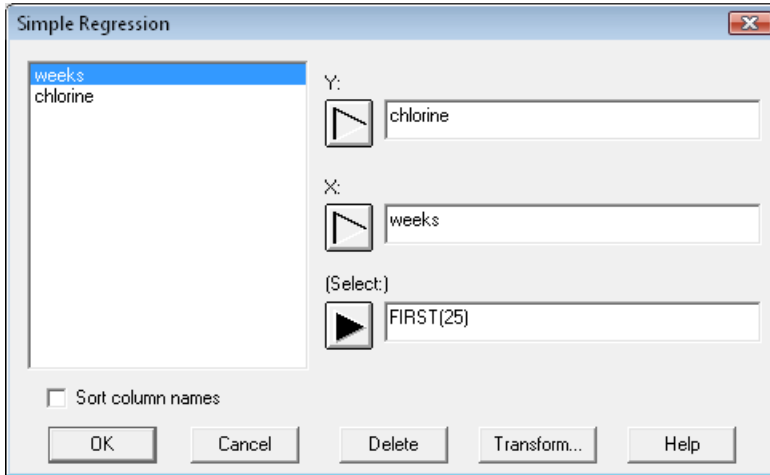
Figure 3-27. Data Input Dialog Box Using Select Field to Restrict Rows Used

Operators that may be placed in the *Select* field are:

Entry	Use	Example
FIRST(k)	Selects the first *k* rows.	FIRST(25)
LAST(k)	Selects the last *k* rows.	LAST(25)
ROWS(start,end)	Selects rows between *start* and *end*, inclusive.	ROWS(11,30)
EXCLUDE(m)	Selects all rows except row *m*.	EXCLUDE(7)
RANDOM(k)	Selects a random set of *k* rows.	RANDOM(25)

Figure 3-26. Operators Used in Select Field to Select Subset of Data Rows

Note that row numbers correspond to the location within the selected data, not the row numbers in the Excel worksheet. For example, if column names are located in row 1 of your Excel file, then a statement such as ROWS(11,30) will select the data in Excel rows numbered 12

through 31. You may find it convenient to add an additional column to your Excel file with observation numbers, although such a column is not required.

The *Select* field may also contain Boolean expressions to control the selection of rows, as in the dialog box shown below:
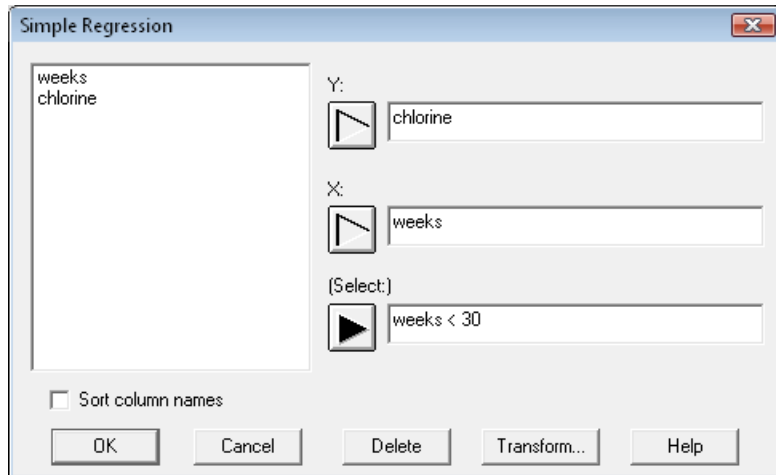
Figure 3-28. Data Input Dialog Box Using Boolean Expression in Select Field

Valid Boolean expressions are shown in the following table:

Entry	Use	Example
column < value	Selects only rows for which *column* is less than *value*.	weeks < 30
column <= value	Selects only rows for which *column* is less than or equal to *value*.	weeks <= 30
column > value	Selects only rows for which *column* is greater than *value*.	weeks > 30
column >= value	Selects only rows for which *column* is greater than or equal to *value*.	weeks >= 30
column = value	Selects only rows for which *column* equals *value*.	weeks = 30
column <> value	Selects only rows for which *column* does not equal *value*.	weeks <> 30
condition1 & condition2	Selects only rows that meet both conditions.	weeks <= 30 & chlorine > 0.4
condition1 | condition2	Selects only rows that meet at least	weeks <= 30| chlorine > 0.4

	one of the conditions.	

Figure 3-28. Common Boolean Expressions Used to Select Subset of Rows

Boolean expressions may involve any data variable, whether part of the current analysis or not. Expressions used with numeric variables are written without quotes. Expressions used with categorical variables require double quotes, as in *Color = "Red"*.

Graphics

Modifying graphs, saving graphics profiles, interacting with graphs, saving graphs in image files, and copying graphs to other applications.

The statistical procedures in STATGRAPHICS *Sigma Express* create hundreds of different types of graphs. To facilitate the data analysis process, default titles, scaling, and other attributes are selected whenever a new graph is created. For analysis purposes, the defaults often suffice. But when it comes time to publish the final results, creating a publication-quality graph is important.

This chapter describes everything you need to know to work with graphs in STATGRAPHICS Sigma Express. It shows you how to dress them up for publication. It shows you how to copy them to applications such as Microsoft Word and PowerPoint. It also shows you how to interact with graphs. For example, you might see an interesting point and wish to know more about it. Or you might want to spin a 3D plot around to get a sense of any relationship that might be present between the variables portrayed on the X, Y and Z axes.

As an example, we will consider again the data in the *nonlin.xls* file. To begin, the fitted model plot relating the amount of chlorine to the age of the product will serve to illustrate some of the important graphics operations.

4.1 Modifying Graphs

The *Simple Regression* procedure is commonly used to fit curves relating a response variable Y and an explanatory variable X. As illustrated in the last chapter, a Reciprocal-X model provides a good fit to the relationship between the *chlorine* column and the *weeks* column.

When first created, a plot of the fitted Reciprocal X model is displayed as shown below:
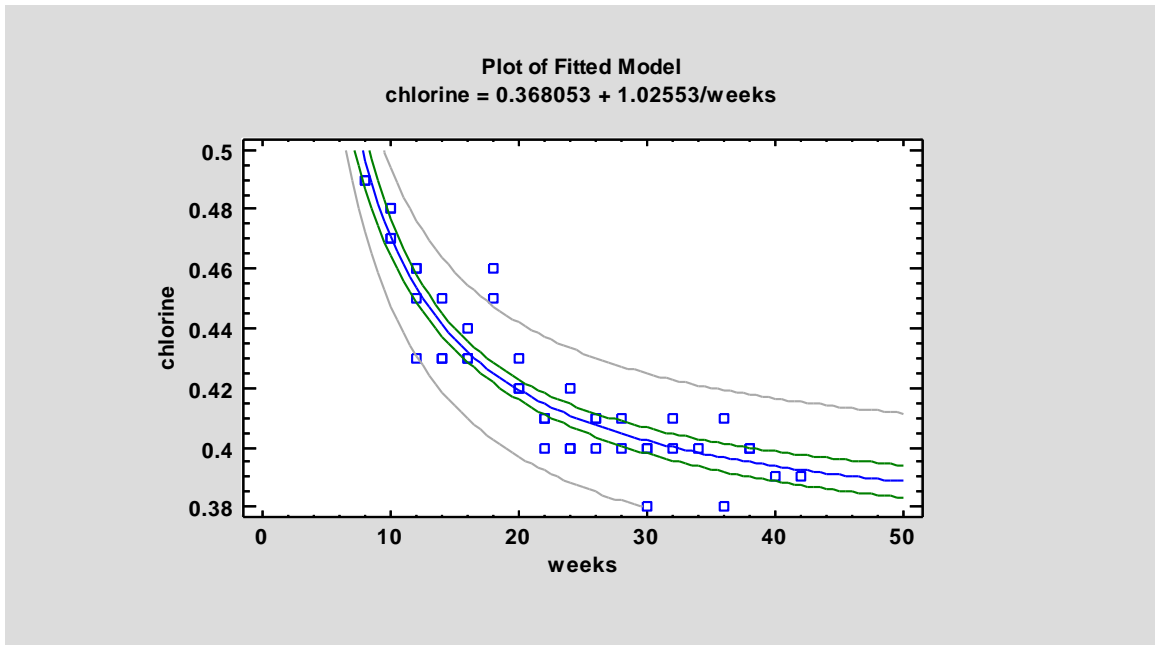
Figure 4-1. Fitted Model Plot with Default Titles and Scaling

The titles, scaling, point and line types, colors, and other graphics attributes were automatically selected.

4.1.1 Layout Options

To modify a graph once it has been created, first double-click on it so that it fills the analysis window. Then click on the *Graphics Options* button ▦ located on the analysis toolbar. A tabbed dialog box will be displayed, with tabs corresponding to different graphics elements.

The *Layout* tab of the *Graphics Options* dialog box is used to change some of the basic features of the graph:
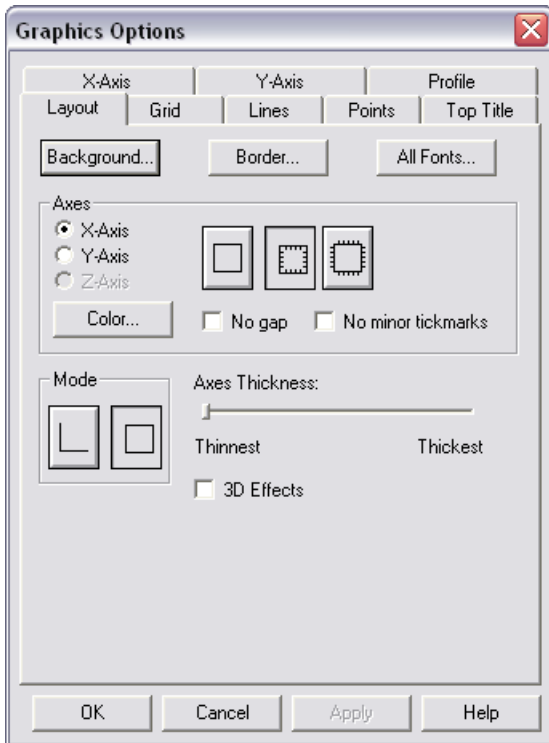
Figure 4-2. Layout Tab on Graphics Options Dialog Box

This includes the orientation of the axis tickmarks, the thickness of the axes, and the color of the graph's background and border. For example, changing the *Background* color to yellow and adding *3D Effects* modifies the plot as shown below:
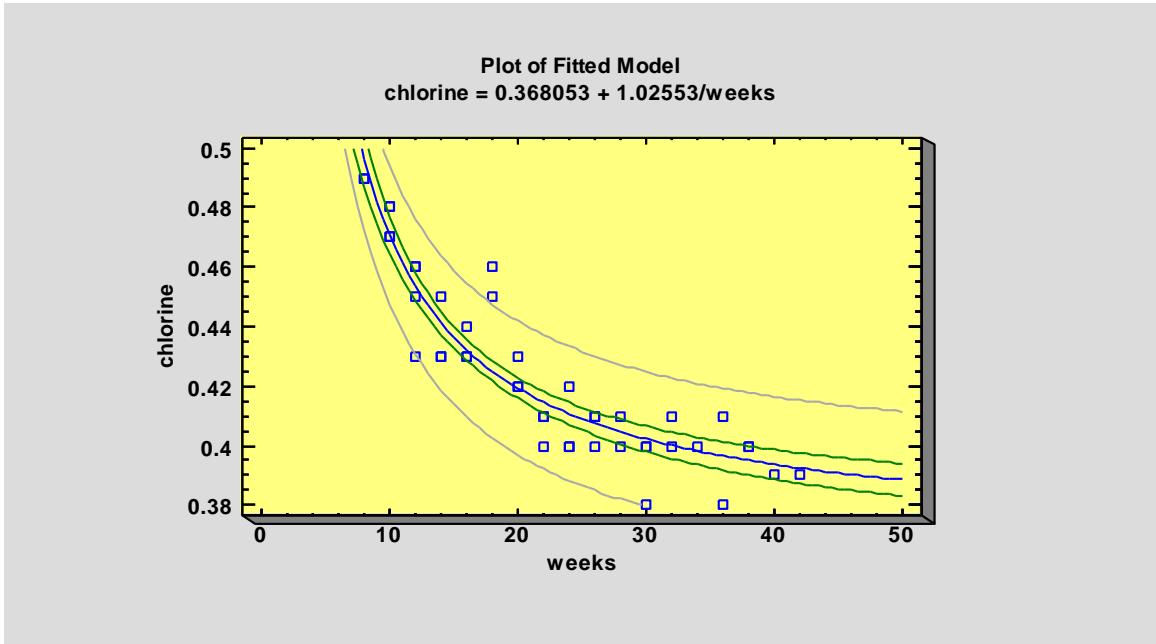
Figure 4-3. Plot after Modifying Background Color and Selecting 3D Effects

NOTE: This color change can be seen in the help documentation provided with your software by clicking on *Help – Getting Started.*

4.1.2 Grid Options

The *Grid* tab is used to add a grid to the plot:
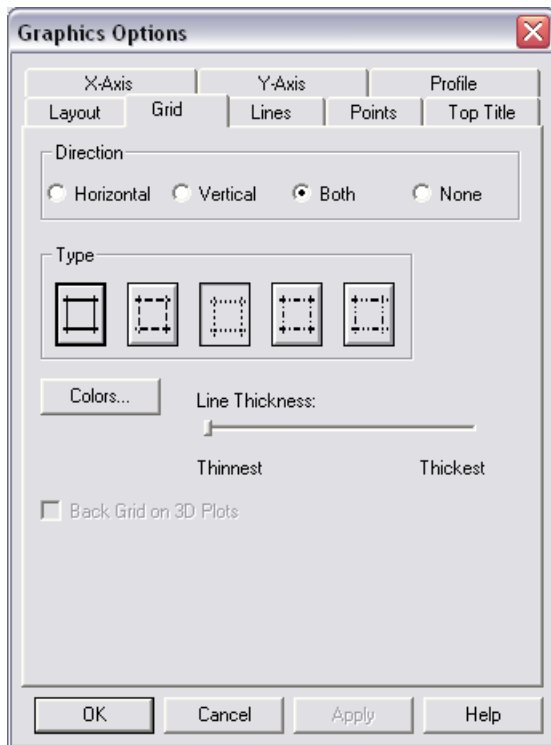
Figure 4-4. Grid Tab on Graphics Options Dialog Box

Adding a gray, dashed-line grid in *Both* directions produces the following graph:
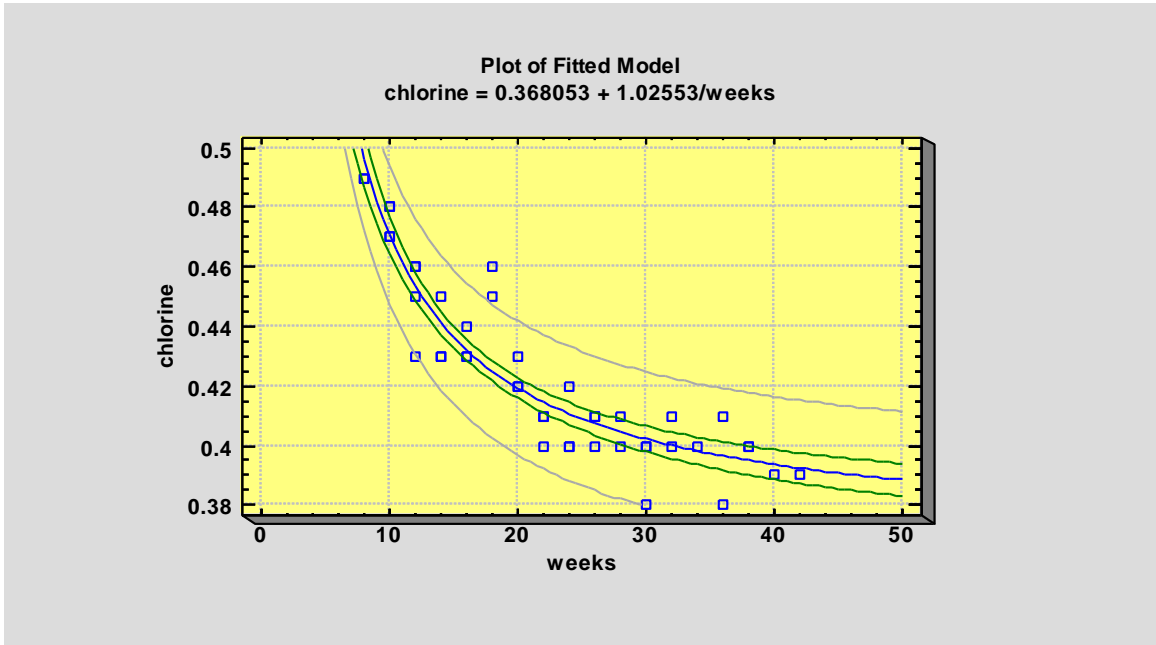
Figure 4-5. Plot after Adding a Grid

4.1.3 Lines Options

The *Lines* tab is used to specify the type, color and thickness of lines on a graph:
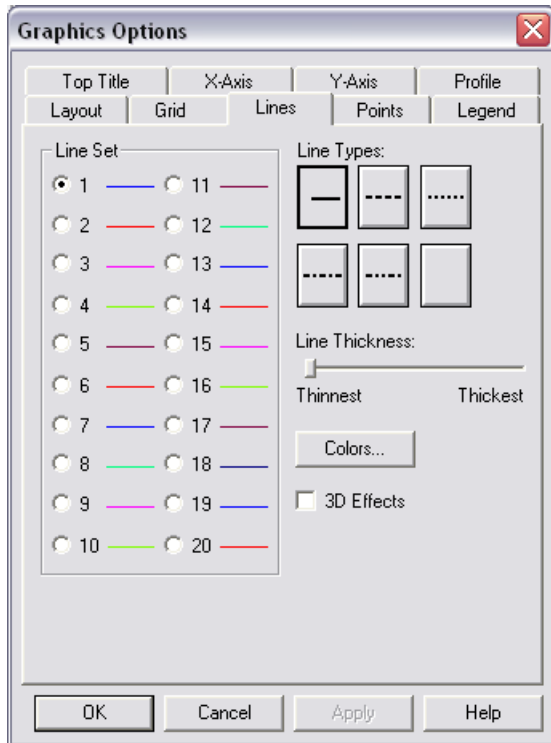
Figure 4-6. Lines Tab on Graphics Options Dialog Box

A plot such as that of the fitted model has three line sets: the line of best fit, the inner confidence limits, and the outer prediction limits. To change any of these types, click on radio button #1, #2 or #3 and then select the desired attributes. Increasing the thickness of the center line results in:
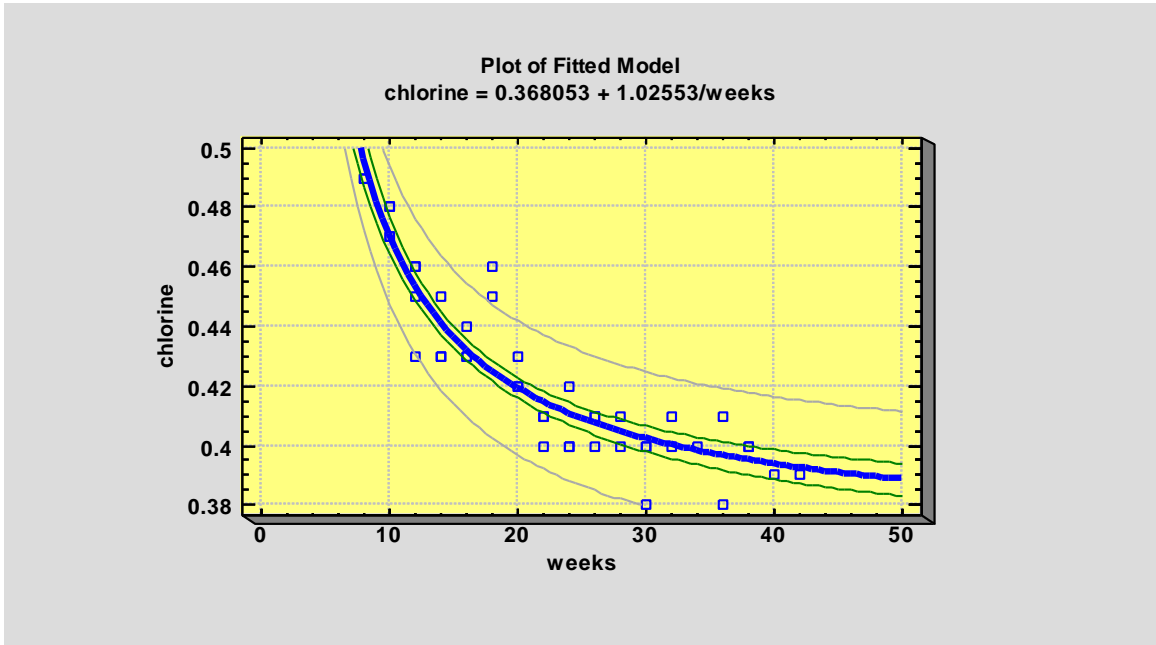
Figure 4-7. Plot after Modifying Thickness of Line Type #1

NOTE: You can only change the thickness of solid lines.

4.1.4 Points Options

The *Points* tab is used to specify the type, color and size of points on a graph:
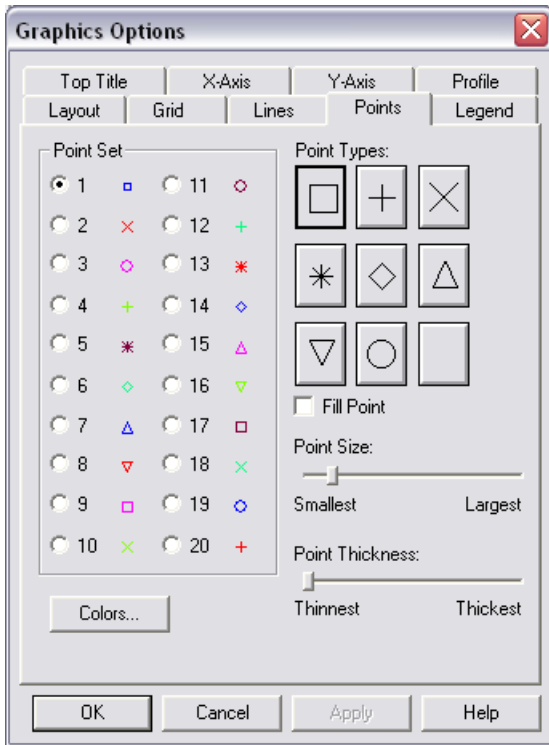
Figure 4-8. Points Tab on Graphics Options Dialog Box

Radio button #1 controls the attributes of the first set of points on a graph. In the current example, there is only one set. Changing the points to solid diamonds creates the following plot:
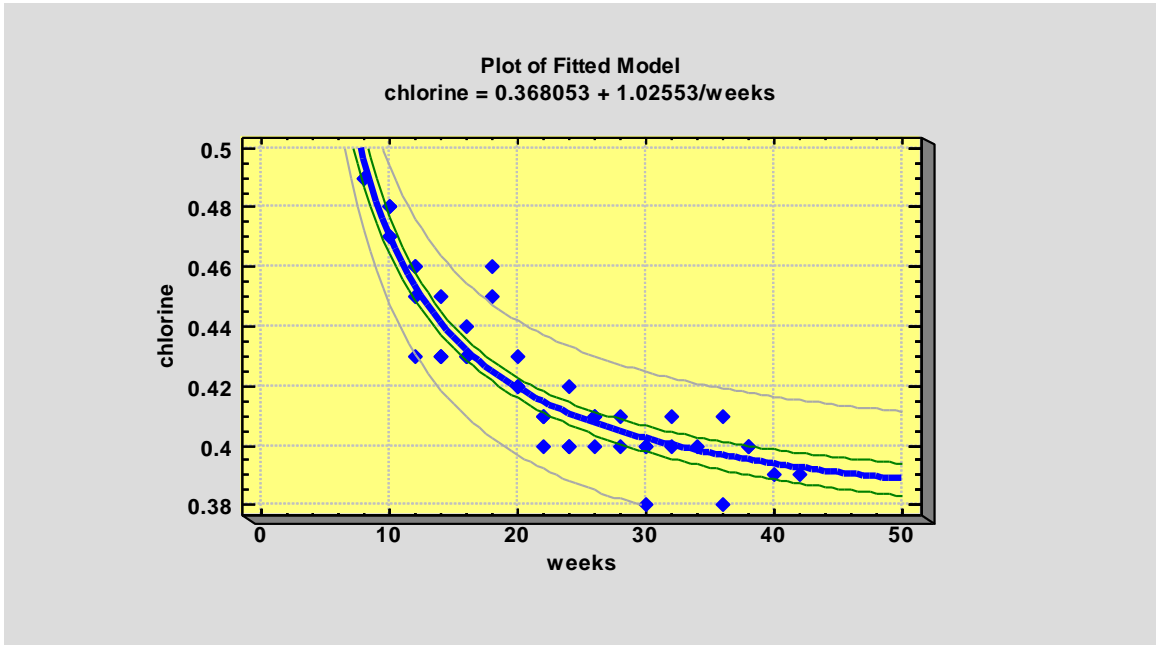
Figure 4-9. Plot after Modifying Point Type #1

4.1.5 Top Title Options

The *Top Title* tab is used to specify the text and font type for the information displayed above a graph:
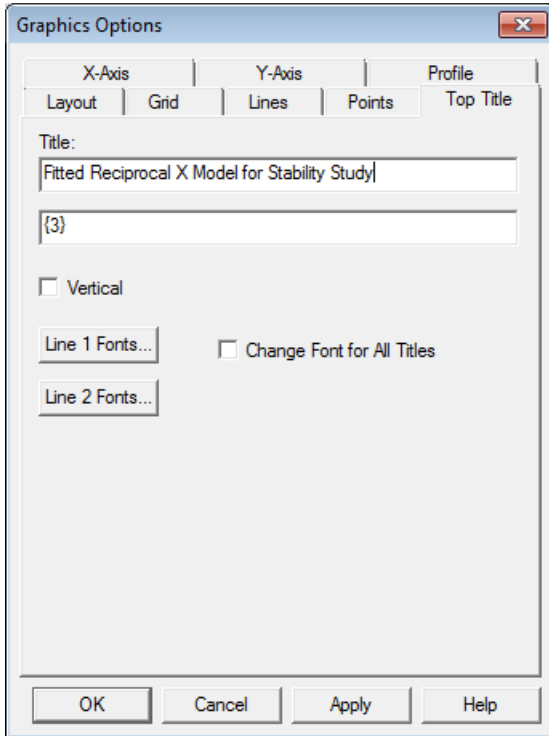
Figure 4-10. Top Title Tab on Graphics Options Dialog Box

Graphs have up to 2 title lines. An entry such as "{3}" in a title field indicates that the text is automatically generated by the analysis procedure, usually containing variable names or calculated statistics. You may change any title, including those that are automatically created. You may also drag the title to a new location with your mouse directly on the graph:
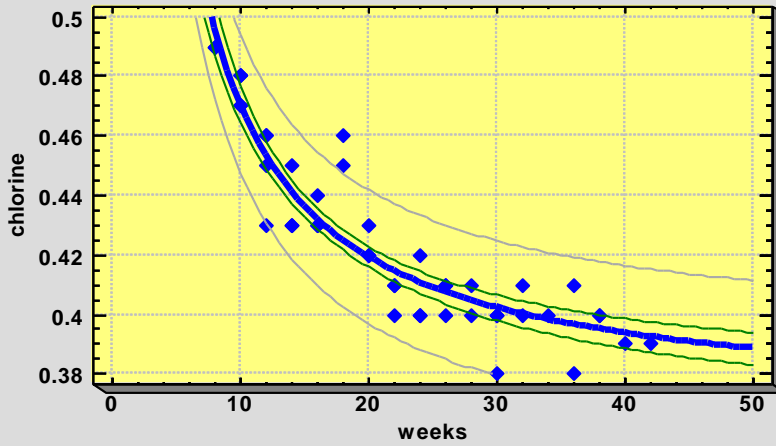
Figure 4-11. Plot after Modifying the Top Title

4.1.6 Axis Scaling Options

The *Graphics Options* dialog box also contains tabs that allow you to modify the axis titles and scaling:
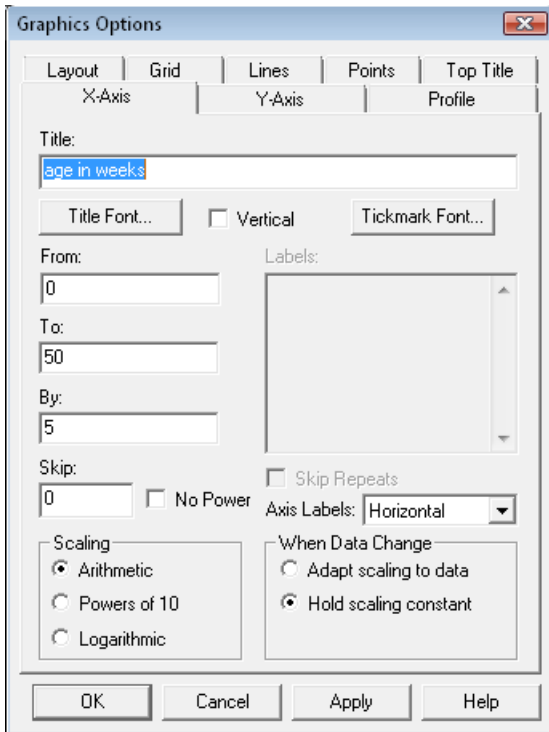
Figure 4-12. X-Axis Tab on Graphics Options Dialog Box

There are several important fields on this dialog box:

1. *Title*: title plotted along the axis.

2. *From*, *To*, *By*, and *Skip*: sets the tickmark scaling. The value in *Skip* is used to prevent displaying certain tickmarks if they run into each other. For example, a value of 1 in the *Skip* field would skip showing every other tickmark.

3. *No Power:* suppresses the display of large and small numbers using labels such as (X 1000).

4. *Axis Labels:* changes the orientation of the tickmark labels.

5. *Scaling:* draws the axis using two different base 10 logarithmic scales.

6. *When Data Change:* specifies whether the scaling should stay constant or change if new data are plotted.

7. *Title Font* and *Tickmark Font:* changes the color, size, or style of the title and tickmarks.

The output generated from the above dialog box changes is shown below:
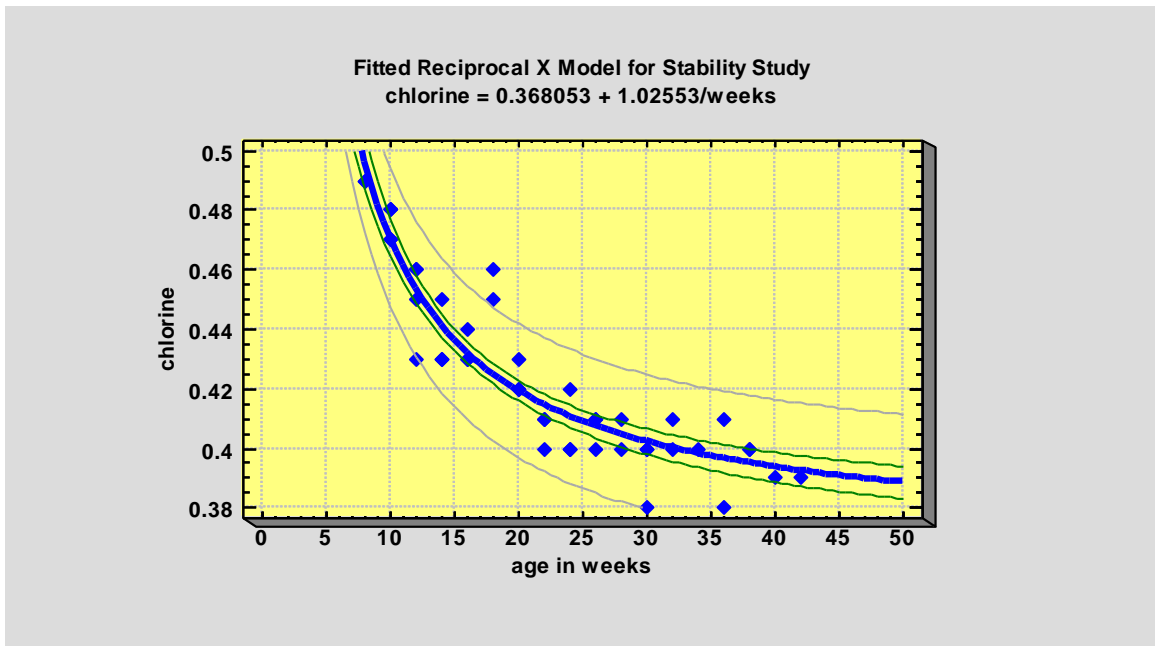
Figure 4-13. Plot after Modifying the Axis Titles and Scaling

4.1.7 Fill Options

Some plots, such as histograms, contain solid areas. The *Fills* tab on the *Graphics Options* dialog box controls the color and fill type of bars, polygons, and pie slices:
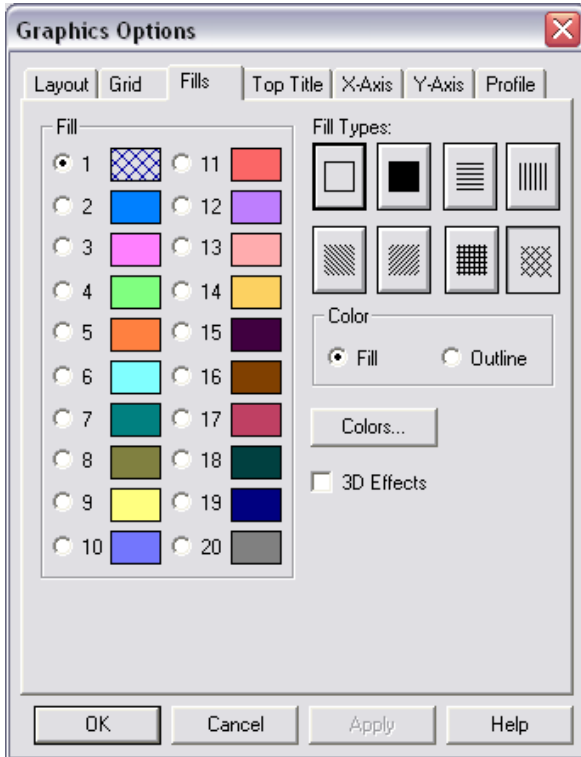
Figure 4-14. Fills Tab on Graphics Options Dialog Box

Radio button #1 controls the first fill type on a graph. In a histogram, all of the bars use the first fill type. On some graphs, such as piecharts, more than one fill type is used. In those cases, radio buttons #2 through #20 control the other fill types.

For plots such as histograms, setting a non-solid fill type is often a good idea when printing the results in black and white:
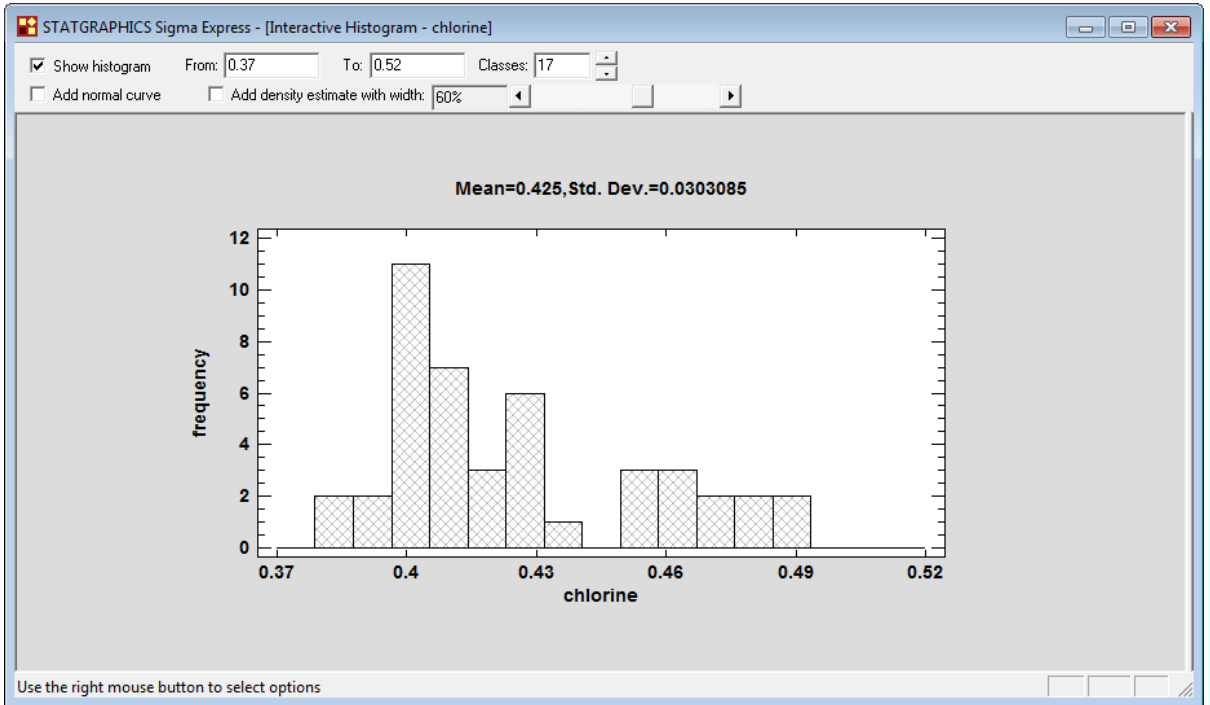
Figure 4-15. Frequency Histogram with Modified Fill Type

4.1.8 Text, Labels and Legends Options

For graphs containing additional legends or labels, tabs are included on the *Graphics Options* dialog box that allow you to change the text and fonts.

4.1.9 Adding New Text

Additional text may also be added to any graph by pressing the *Add text* button ![abc icon] on the analysis toolbar. A dialog box will be generated in which to enter the new text:

Text Options

Text:
< outlier

Direction
- Horizontal
- Vertical

[OK] [Cancel] [Fonts...] [Help]
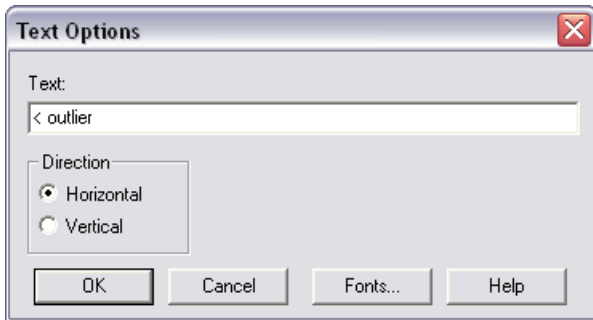
Figure 4-16. Dialog Box for Adding New Text

The text string will be initially positioned under the top title, but may be dragged to any location with the mouse:

Fitted Reciprocal X Model for Stability Study
chlorine = 0.368053 + 1.02553/weeks
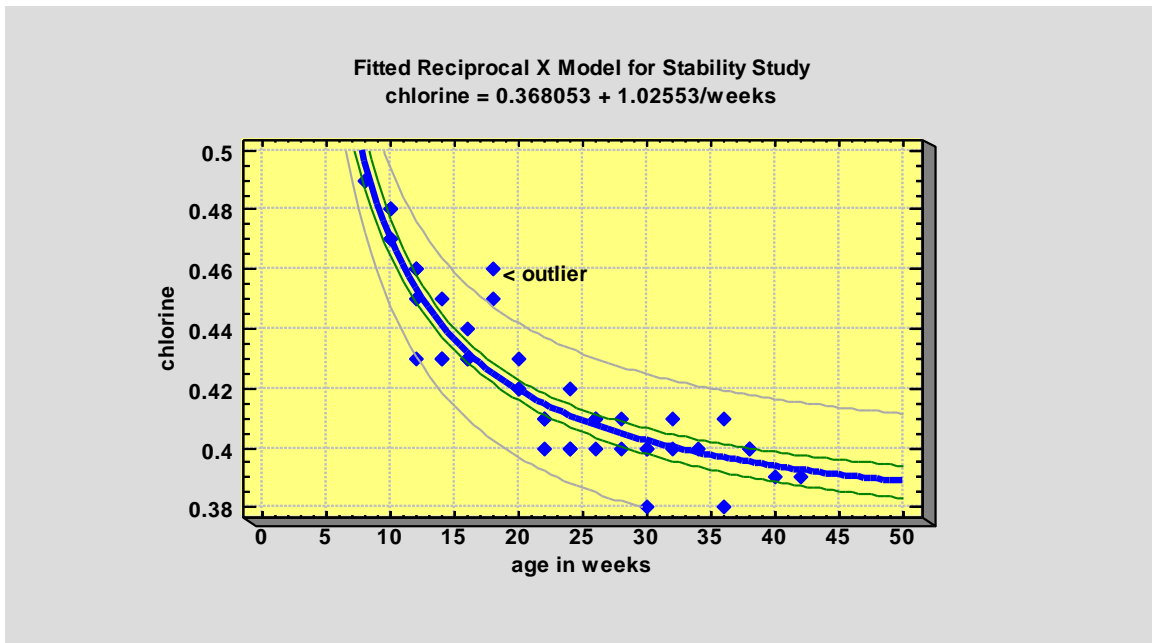
Figure 4-17. Plot after Adding New Text String

After text is added, click on it and then press the *Graphics Option* button if changes need to be made.

4.2 Jittering a Scatterplot

When one or both of the variables in a scatterplot are discrete, the chance of points being exactly in the same location and obscuring each other can be large. The analysis toolbar has a *Jitter* button that overcomes this problem by randomly offsetting points in the horizontal and/or vertical direction. For example, consider the following plot of data contained in the *93cars.xls* file:
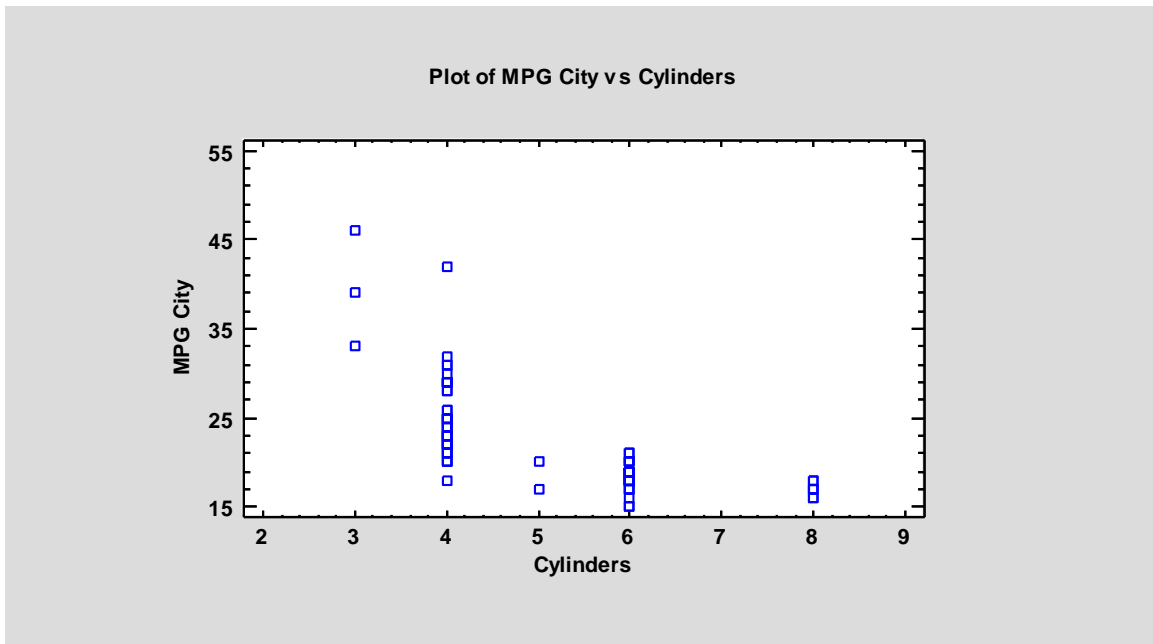
Figure 4-18. Scatterplot of Miles per Gallon versus Cylinders

The plot shows the miles per gallon obtained in city driving by 93 different makes and models of automobiles, plotted versus the number of cylinders in the car's engine. Although there are 93 rows in the worksheet, there are many less points than that on the plot. The small number of points is due to exact ties amongst the cases, which cause some points to plot exactly on top of others.

If you press the *Jitter* button on the analysis toolbar, a dialog box will appear allowing you to add a little jitter (random offset) to the points:
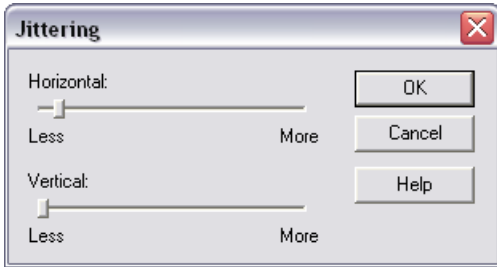
Figure 4-19. Jittering Dialog Box

In this case, adding a small amount of horizontal jitter gives a much better picture of the location of the points:
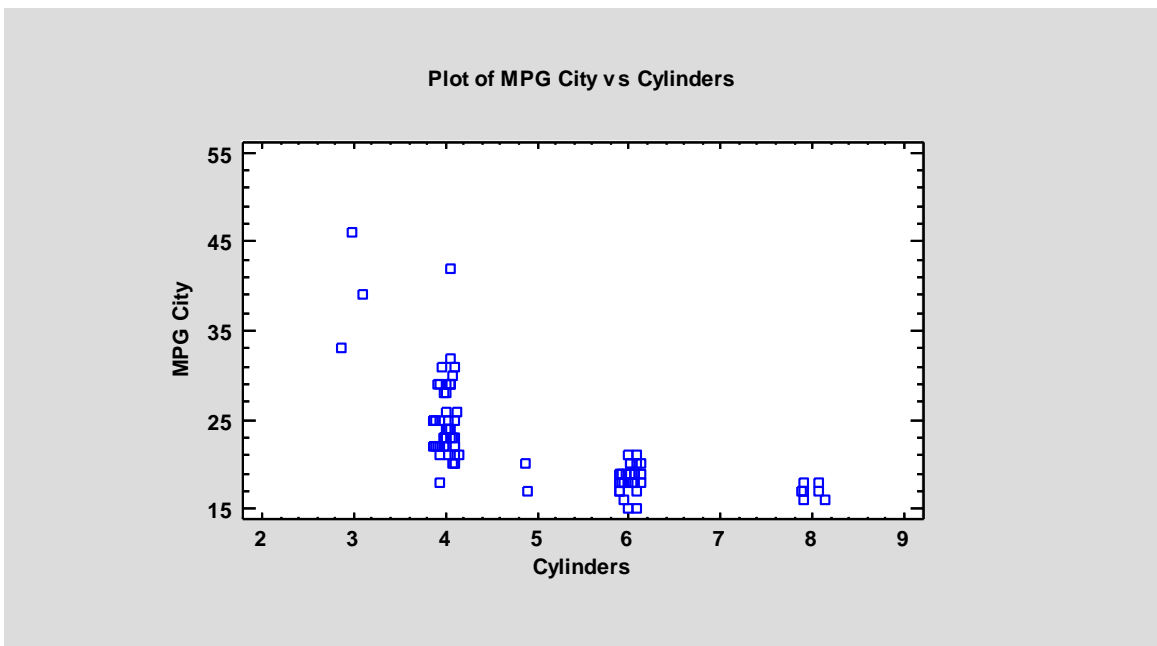
Figure 4-20. Scatterplot after Horizontal Jitter

Each point has been offset a small random amount along the horizontal axis. Jittering a plot affects only the display. It has no effect on the data in the datasheet or any calculations made with it.

4.3 Brushing a Scatterplot

An interesting method of visualizing relationships between variables is to color the points of a scatterplot according to the value of another variable. For example, consider the following *Matrix Plot* for selected variables from the *93cars.xls* file:
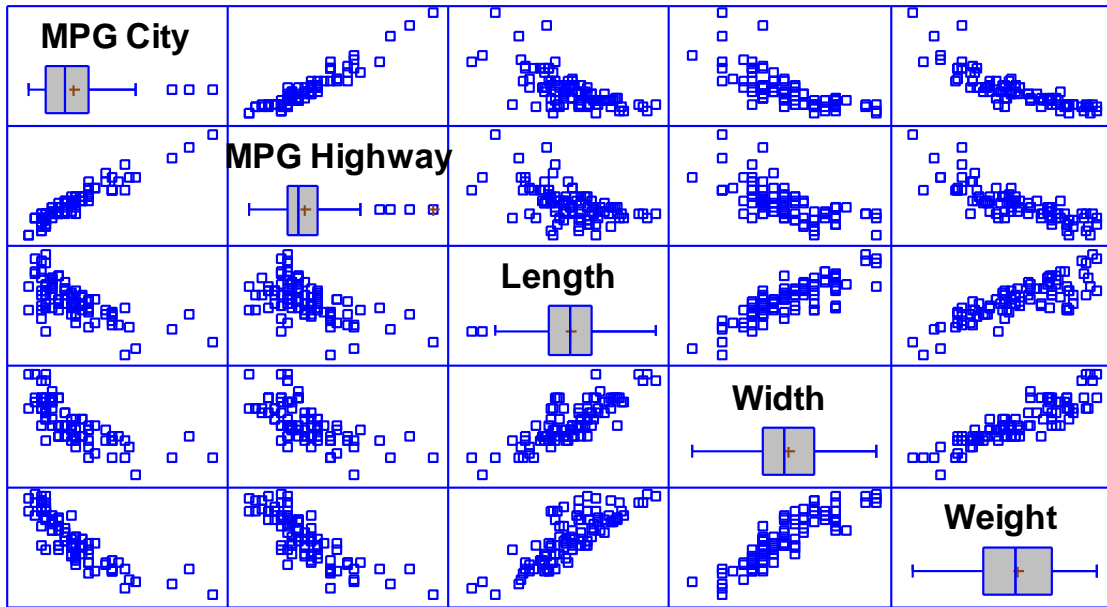
Figure 4-21. Matrix Plot for Data from the 93cars File

The scatterplot in each cell of the matrix plots the values of variables corresponding to its row and column identifiers.

Suppose you wish to visualize how the *horsepower* of the automobiles is related to the 5 plotted variables. If you press the *Brush* button on the analysis toolbar, the following dialog box will be displayed:
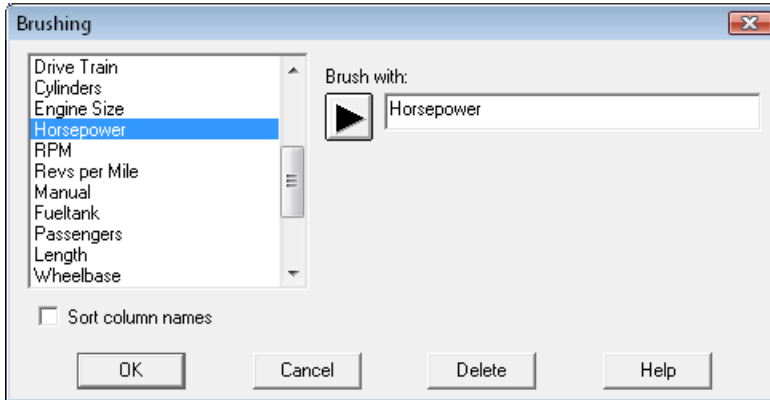
Figure 4-22. Dialog Box for Selecting Brushing Variable

Select a quantitative variable to use to code the points. After selecting the variable to brush with, a floating dialog box will appear:
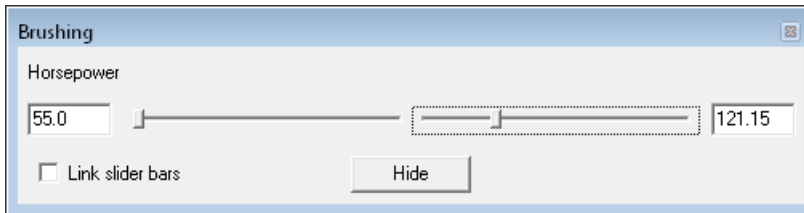
Figure 4-23. Floating Dialog Box for Selecting Brushing Interval

The two slider bars are used to specify lower and upper limits for the variable. All points in the plot are colored red if they fall within the specified interval. For example, in the plot below, all automobiles with horsepower between 55.0 and 121.15 are colored red:
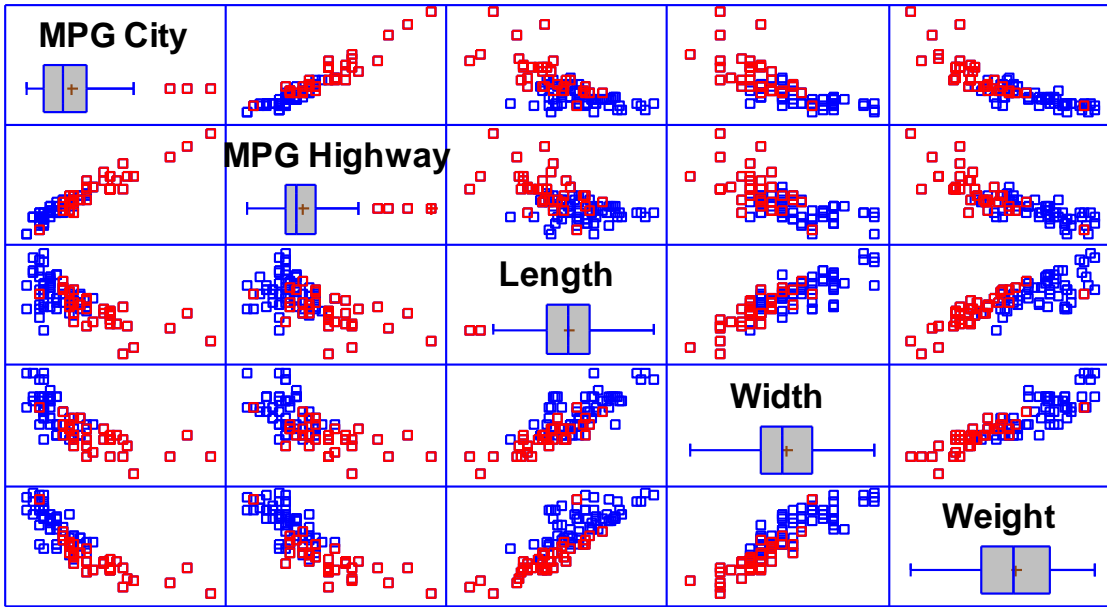
Figure 4-24. Matrix Plot after Brushing Points

It is evident from the above plot that *Horsepower* is strongly correlated with the other variables.

4.4 Smoothing a Scatterplot

To help visualize the relationships between the variables in a scatterplot, a smoother may be added. To smooth a scatterplot, press the *Smooth/Rotate* button on the analysis toolbar. This will display the following dialog box:
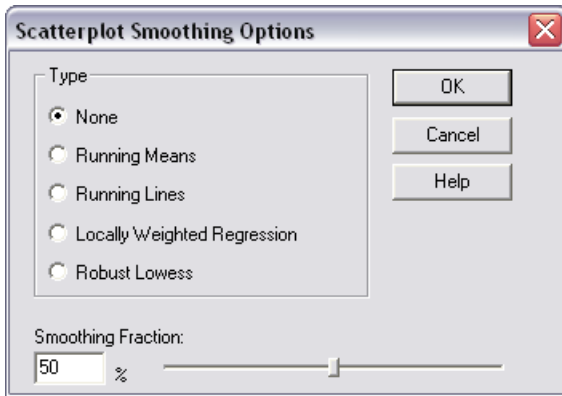
Figure 4-25. Scatterplot Smoothing Dialog Box

Smoothing a scatterplot is done by selecting a set of locations along the X-axis and at each location plotting a weighted average of the specified fraction of the points that are closest to that location. One of the best methods for smoothing is called LOWESS (LOcally WEighted Scatterplot Smoothing), usually with a smoothing fraction between 40% and 60%. The result of smoothing the *Matrix Plot* of the automobile data is shown below:
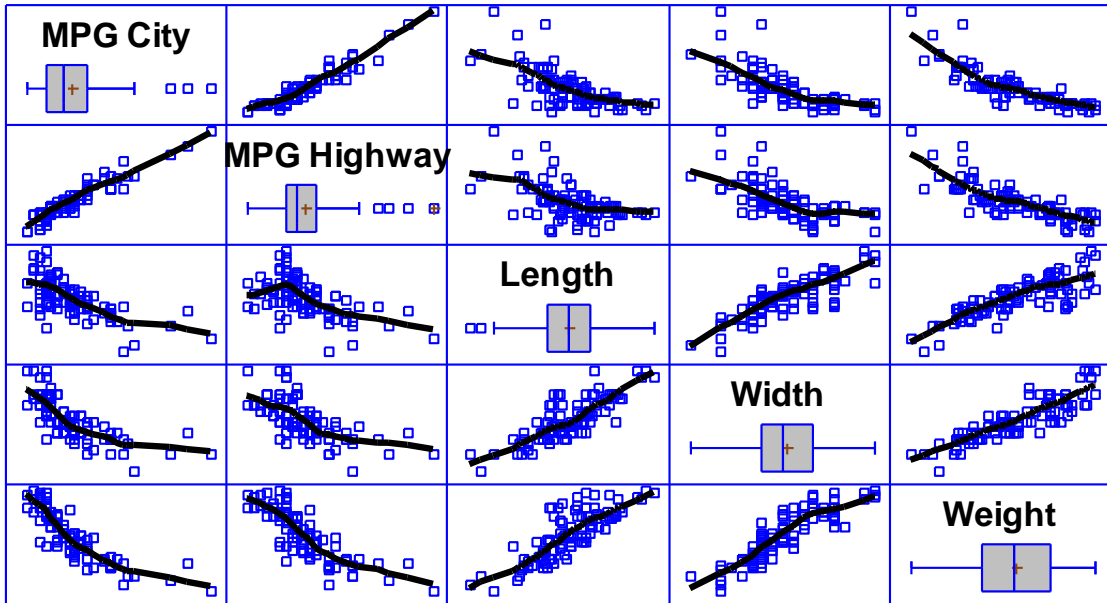
Figure 4-26. Smoothed Matrix Plot using Lowess with a 50% Smoothing Fraction

The smooth helps illustrate the type of relationships between the variables.

4.5 Identifying Points

To display the row number and coordinates corresponding to any point on a graph, you may hold the mouse button down on the point. A small box will be displayed in the upper right corner of the plot, showing the row number and coordinates of the point:
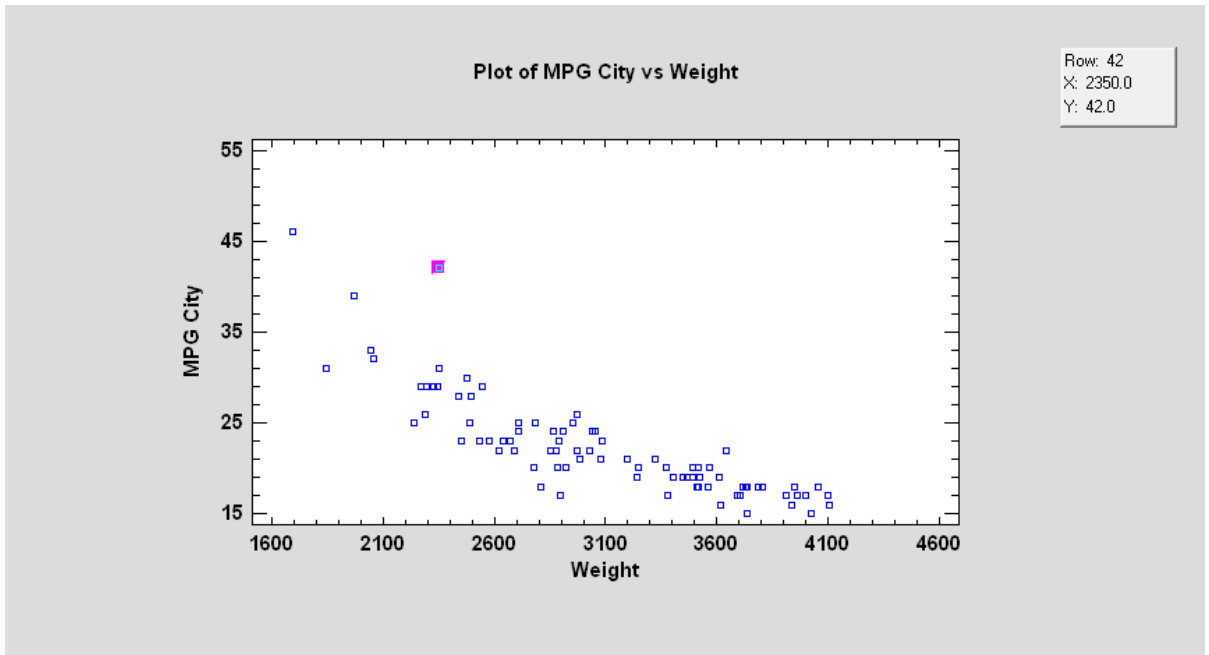
Figure 4-27. Displaying Information about a Selected Point

At the same time, the row number of the point will be placed into the *Row* field on the analysis toolbar:

Figure 4-28. Analysis Toolbar Showing Row Number of Selected Point

Additional information about the point may be obtained by pressing the *Identify* button and selecting a column from the Excel worksheet:
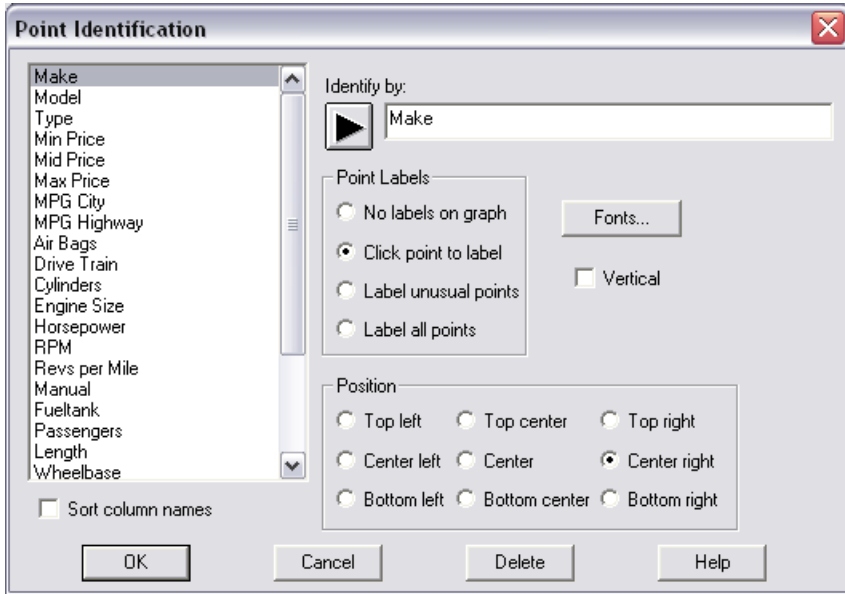
Figure 4-29. Point Identification Dialog Box

In the *Point Labels* field, select an option for which labels will be displayed. If you *select Click point to label* and press *OK*, clicking on any point will add the value of that variable to the *Label* field on the analysis toolbar:

Figure 4-30. Analysis Toolbar Showing Make of Selected Point

The binoculars buttons to the right of the *Label* and *Row* fields may be used to locate points on a graph. If you enter a value into either edit field and then press the corresponding *Locate* button, all points in the graph matching the entered value will be highlighted. For example, the plot below colors the points for all Hondas red:
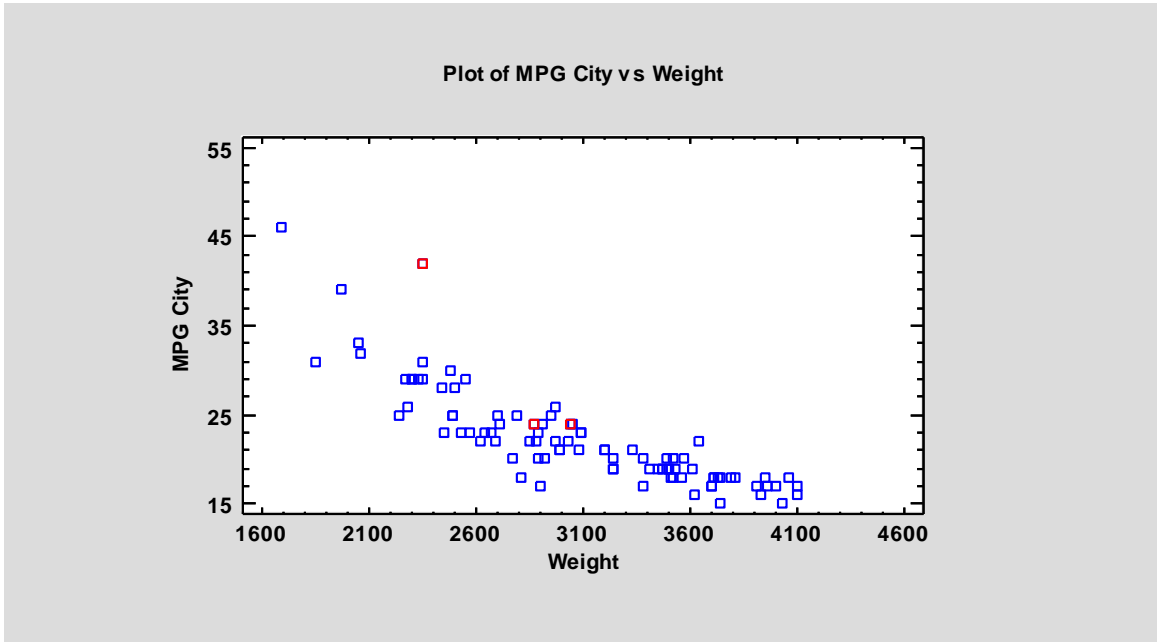
Figure 4-31. Plot Highlighting All Hondas

This technique is also quite effective on a *Matrix Plot*. In the following display, all points corresponding to row #42 have been highlighted:
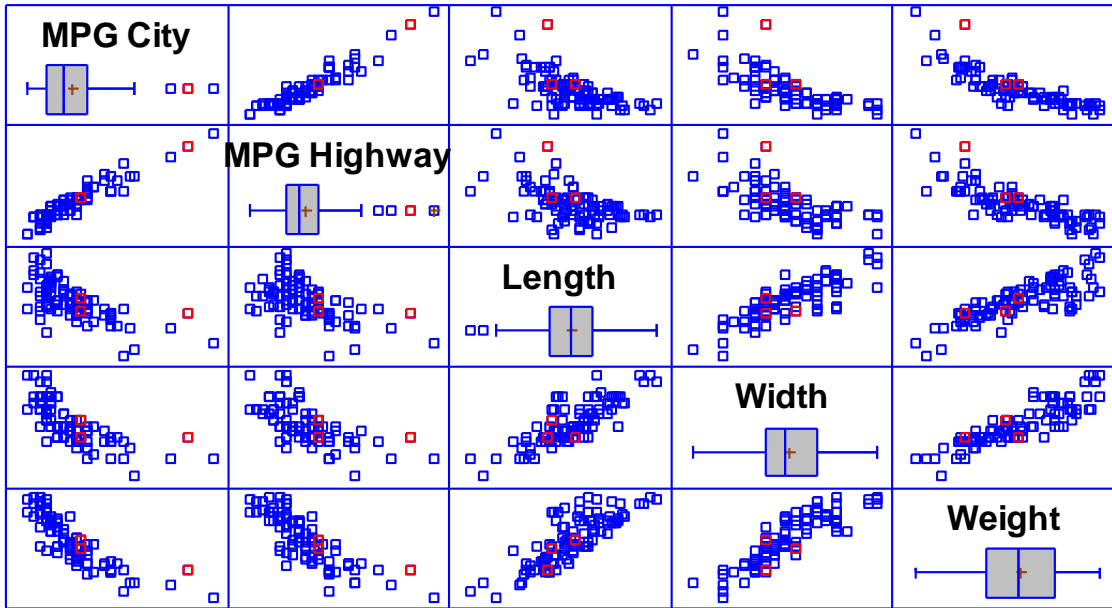
Figure 4-32. Matrix Plot Highlighting Row #42

Locating a point in a *Matrix Plot* can help identify whether it is an outlier with respect to more than one variable. The color used to highlight the points is specified on the *Graphics* tab of the *System Preferences* dialog box, accessible from the *Tools* menu.

4.6 Copying Graphs to Other Applications

Once a graph has been created in STATGRAPHICS Sigma Express, it can be easily copied to other programs such as Microsoft Excel, Word or PowerPoint by:

1. Maximizing the pane containing the graph.

2. Clicking on the graph with the right mouse button and selecting *Copy* from the popup menu.

3. Selecting *Paste* while in the other application.

By default, graphs are pasted in "Picture" format, which corresponds to a Windows metafile. In rare cases when you wish to paste the graph in some other format, you can select *Paste Special* instead of a simple *Paste*.

To copy both the graph and its enclosing window, a third-party screen capture tool is recommended. In producing this manual, a program called *SnagIt* has been used, which is available for purchase at *www.techsmith.com*. If you use *SnagIt*, we recommend that you set the *Input* option to "Window" and the *Output* option to "Clipboard". You can then paste images directly into any document.

4.7 Saving Graphs in Image Files

Individual graphs may also be saved in image files by maximizing a graph and then selecting *Save Graph* from the right-mouse-button popup menu. A dialog box will be displayed on which to specify a file name and image format:
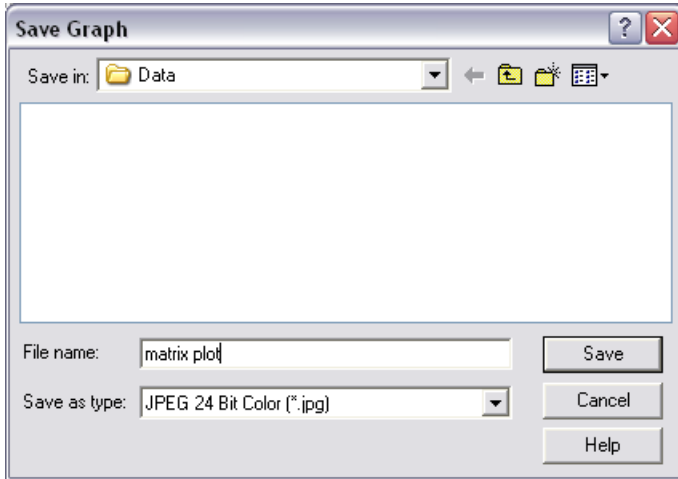
Figure 4-33. File Selection Dialog Box for Saving Graph in Image File

For saving graphs that are to be read into Word or PowerPoint, saving the graph as a Windows metafile gives the most flexibility. If the graph is to be displayed on a web page, saving it as a JPEG file is recommended.

System Preferences

Setting preferences for system behavior.

STATGRAPHICS Sigma Express contains many options, each of which has a default value that has been selected to meet most users' needs. If desired, you can set new defaults for many of these options. There are 2 main places in the program to do this:

1. **General system behavior**: set by *System Preferences,* accessible from the *Tools* menu.

2. **Graphs**: set by selecting *Graphics Options* while viewing any graph. The *Profile* tab of the *Graphics Options* dialog box allows you to save multiple sets of graphics attributes.

5.1 General System Behavior

The default values for general system behavior and selected statistical procedures may be changed by selecting *System Preferences* from the *Tools* menu. This displays a tabbed dialog box with a *General* tab for overall system behavior and other tabs for statistical analysis defaults:
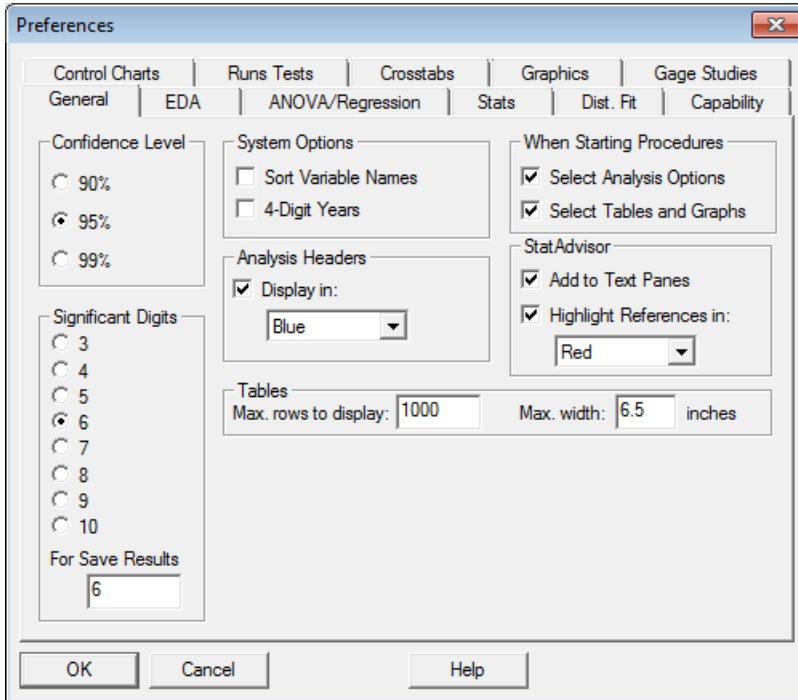
Figure 5-1. Preferences Dialog Box

Some of the most important options that may be set are:

- **Confidence Level**: default percentage used for confidence limits, prediction limits, hypothesis tests, and interpretation of *P*-values by the StatAdvisor.

- **Significant Digits**: number of significant digits used when displaying numerical results. The indicated number of digits will be displayed, except for trailing zeroes that will be dropped. A separate entry is provided for saving numerical results back to the Excel worksheet.

- **System Options**: options that apply system-wide.

  - **Sort Variable Names**: whether to list column names in alphabetic order on data input dialog boxes. Otherwise, column names will be listed in the same order as in the worksheet.

- o **4-Digit Years**: whether dates should be displayed with 4-digit years rather than 2-digit years. By default, 2-digit years such as 2/1/05 are assumed to represent dates between the years 1950-2049.

- **StatAdvisor**: sets the behavior of the StatAdvisor.

  - o **Add to Text Panes**: whether StatAdvisor output should be added to the bottom of text panes.

  - o **Highlight References in …**: whether to highlight in a special color values on text panes that are referred to by the StatAdvisor.

- **Analysis Headers**: whether to use a blue font to display the analysis title at the top of the *Analysis Summary* pane.

For a description of the options on the other tabs, refer to the PDF document titled *System Preferences*.

5.2 Graphics

Maximizing a pane containing a graph within any analysis window activates the *Graphics Options* button on the analysis toolbar. That button displays a tabbed dialog box that allows you to change the appearance of a graph, as described in detail in Chapter 4. Also included on that dialog box is a tab labeled *Profile*, which enables you to save sets of graphics attributes in user profiles and change the default profile that is used when a new graph is created:
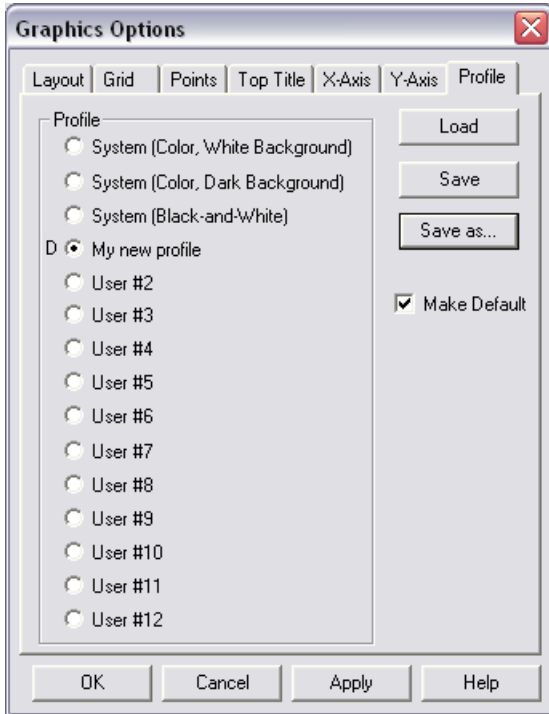
Figure 5-2. Profile Tab on Graphics Options Dialog Box

To change the system defaults:

1. Modify the features of a graph in any analysis window. Set the colors, fonts, and other options that you want future graphs to reflect.

2. Select *Graphics Options* from the analysis toolbar and go to the *Profile* tab.

3. Check *Make Default*.

4. Select any of the 12 user profiles and press the *Save as* button (the system profiles are read-only).
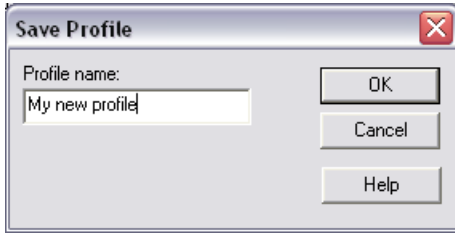
5. Enter a name for the profile to be saved:

Figure 5-3. Save Profile Dialog Box

6. Press *OK* to save the current set of graphics attributes (colors, fonts, point and line styles, etc.) in a new profile.

The next graph created will use the newly saved profile.

You can also apply other saved profiles to a new graph by creating the graph with default settings and then:

1. Select *Graphics Options* from the analysis toolbar and go to the *Profile* tab.

2. Select any of the 15 profiles and press the *Load* button.

The current graph will be immediately updated to reflect the settings in the selected profile.

You can also use the *System Profile Designer* on the *Tools* menu to help set all graphics attributes at one time. It displays sample 2D and 3D plots containing all attributes used thoughout the program:
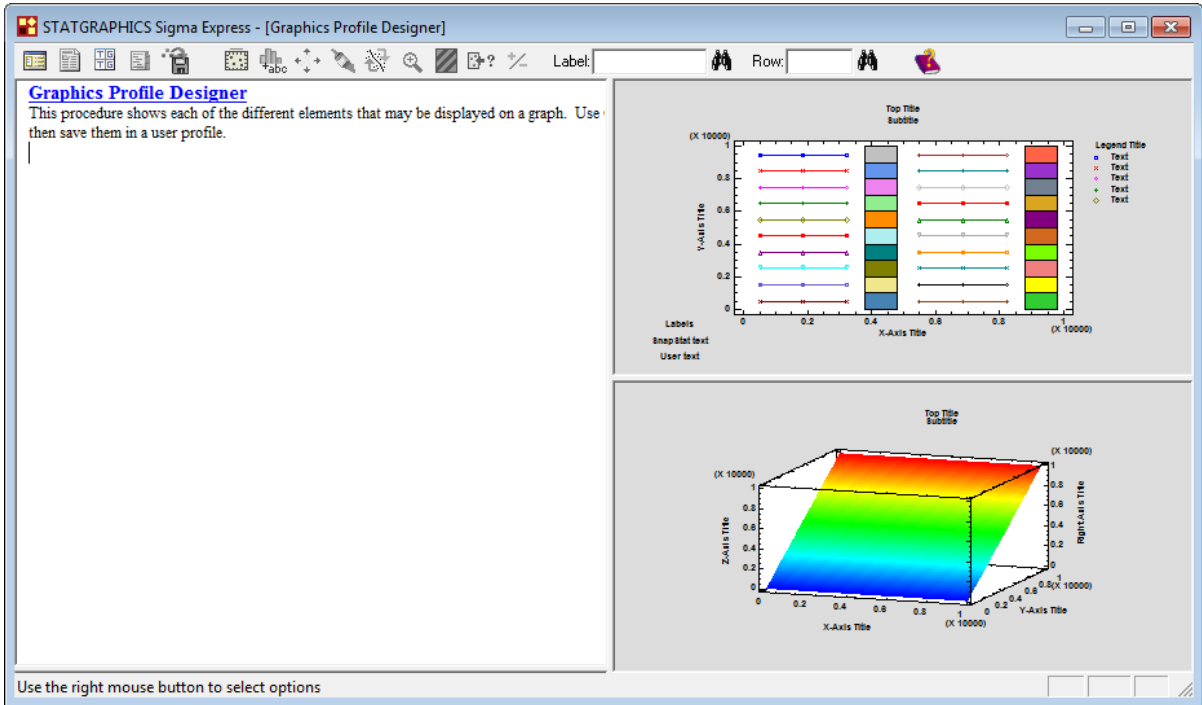
Figure 5-4. Graphics Profile Designer

Use *Graphics Options* to change all settings and then save a new graphics profile.

Index