GraphPad Prism

Version 2

Clearly

the fastest,

easiest way

to analyze

and graph

scientific data

Program design:	Dr. Harvey J. Motulsky Paul Stannard. Dr. Richard Neubig
Programming:	Paul Stannard (SoftEngine, Inc.) Mike Platt John Pilkington
Manual and help screens:	Dr. Harvey J. Motulsky Paige Searle

How to reach GraphPad:

Phone:	858-457-3909
Fax:	858-457-8141
Email:	support@graphpad.com or sales@graphpad.com
World Wide Web:	http://www.graphpad.com
Mail	GraphPad Software, Inc. 5755 Oberlin #110 San Diego, CA 92121 USA

Important

Please read the Software License Agreement and its statement of limited liability on the next page. GraphPad Software, Inc. does not guarantee that the program is error-free, and cannot be held liable for any damages or inconvenience caused by errors. Use of the program implies that you agree with the Agreement.

Although we have tested Prism carefully, the possibility of software errors exists in any complex computer program. You should check important results carefully before drawing conclusions. **GraphPad Prism is designed for research purposes only, and should not be used for the diagnosis or treatment of patients.**

If you encounter a problem, please notify us. We will try to correct all significant problems, and will send you a replacement disk without charge. Also contact us if you have ideas for enhancements or new types of scientific software.

If you find that Prism does not fit your needs, you may return it for a full refund (except shipping fees) within 90 days. You do not need to contact us first. Simply ship the package to us with a note explaining why you are returning the program. Include your phone and fax numbers, and a copy of the invoice.

If you purchased Prism through a dealer or catalog, please contact the company you purchased it from regarding returns.

SOFTWARE LICENSE AGREEMENT

By opening the sealed envelope of disk(s), you indicate acceptance of this license agreement. If you do not accept, return the unopened envelope immediately for a refund.

1. GraphPad Software Inc. (“GraphPad”) grants you a limited license to use the accompanying disks and manual (the “Licensed Software”). Title to the Licensed Software remains with GraphPad at all times. You may not sublicense, rent, lease, convey, or modify the software. You may transfer this license to another individual who agrees to these terms, provided that you keep no copies of the manual or disks.
2. You may use the Licensed Software on more than one computer, provided there is no chance it will be running simultaneously on more than one machine. You may make backup copies of the disks, provided they are labeled with the copyright notice. Unauthorized copying is an infringement of GraphPad's copyright.
3. We offer a 90 day money-back guarantee. If you find that the Licensed Software does not suit your needs, return the manual and disks and a copy of the invoice for a full refund.
4. GraphPad provides the Licensed Software “as is”, and does not warrant that the Licensed Software will meet customer requirements or be error-free. GraphPad shall not be liable for any consequential, incidental, special, punitive, exemplary or other damages, even if advised of the possibility of such damages. GraphPad expressly disclaims any and all warranties, expressed or implied. GraphPad's entire liability, and your exclusive remedy, shall be to replace the Licensed Software or refund the license fee.
5. The Licensed Software is designed for research purposes only, and should not be used for the diagnosis or treatment of patients.
6. This agreement shall be governed by the laws of the State of California, U.S.A. If you have any questions, please contact GraphPad Software: 5755 Oberlin #110, San Diego, CA 92121 USA; FAX 858-457-8141.

Contents

Overview

Introducing GraphPad Prism

GraphPad Software has provided scientists with tools to simplify data analysis and graphing for over a decade. Prism's unique design helps scientists efficiently manage, analyze and graph experimental data. Features unique to Prism include:

- Prism combines basic biostatistics, curve fitting, and scientific graphics in one program. You don't need to buy several programs.
- Prism automatically calculates and graphs error bars from replicate values.
- With Prism, one command analyzes and graphs all related data sets at once. You don't have to analyze one data set at a time.
- After you fix a data entry error, Prism automatically updates all analyses, graphs and page layouts. You don't have to repeat any commnds.
- Prism lets you analyze data from repeat experiments instantly. Simply enter the new data, click once, and Prism instantly graphs and analyzes the data. You don't have to retype commands, and you don't need to create complicated batch files or macros.
- Prism saves all related data, analyses, graphs and notes in one file to keep your work organized. The results of data analysis are part of the file, so you can review your analysis choices and results in the future.
- Prism makes it easy to arrange several graphs on a page. You can even include data tables and analysis results on the same page with graphs.
- Prism's comprehensive manual and help screens help you learn how to analyze your data and interpret the results. Prism doesn't assume that you are a statistician.

We plan to make Prism even better, and welcome your comments and suggestions.

System requirements

GraphPad Prism runs on any computer running Microsoft Windows version 3.1 or Windows 95 with at least 3 megabytes of space on the hard drive.

We recommend that you use a computer that has a fast 386 processor or better, at least 4 megabytes of RAM, and a mouse (or other pointing device).

If your computer has a math coprocessor, Prism will use it to accelerate mathematical calculations. Prism does not require a coprocessor.

Prism runs fine on a Macintosh running SoftWindows version 2, but not under SoftWindows version 1.

Prism uses Windows printer drivers. Microsoft Windows comes with printer drivers for most common printers. Other drivers are available from printer manufacturers. If you have installed the necessary Windows printer drivers, Prism can also "print" to a fax modem or slide maker.

Graphics programs such as Prism place heavy demands on the printer driver, more so than word processors. Since drivers are updated frequently, check that your driver is fairly recent. See "Printing problems" on page 364.

Prism only displays and prints text using TrueType fonts. Several TrueType fonts are provided with Windows, and thousands of additional fonts are available at software stores. Prism cannot use other types of fonts (i.e. ATM). Prism expects to find three standard fonts: Arial, Symbol, and Wingding.

GraphPad Software does not provide Windows printer drivers or TrueType fonts.

Installing GraphPad Prism

Before installing Prism, you may wish to make backup copies of the distribution disks. Prism must be installed from the original disks or copies of the original disks. It is not possible to copy Prism from one hard disk to another.

To install GraphPad Prism:

1. If Windows is not already running, start it.
2. Insert disk #1 into a drive. Do not write protect this disk.
3. If you are running Windows 3.1: From the Program Manager, pull down the File menu and select Run.

   If you are running Windows 95: Press the Start button at the lower left of the Windows screen, and select Run.
4. In the Run dialog box, type **A:INSTALL** or **B:INSTALL**.
5. Click OK.

6. Follow the instructions on screen.

Installation notes:

- If you purchased GraphPad Prism together with GraphPad StatMate, install Prism first.
- It is not possible to install Prism manually. You must use the installation program which decompresses and copies files from the original disks to your hard drive. The installation program also places the Prism icon onto your Program Manager desktop.
- The installation program does not make any changes to your system files (autoexec.bat, config.sys, win.ini, or system.ini).
- If you run Windows 3.1, the install program will install that version of Prism. If you run Windows 95 or NT, the install program will let you choose to install the Windows 3.1 (16 bit) or the Windows 95 (32 bit) version.
- After installing Prism, we recommend that you follow the tutorial in the next chapter.

If you are upgrading from version 1

If you are upgrading from version 1 to version 2, note the following:

- Do NOT uninstall that older version until after you have installed and run Prism 2. Otherwise you will lose your optional settings and user-defined equations.
- If you accept the default choices, Prism will be installed in a \PRISM2 directory, and Prism version 1 will remain in the \PRISM directory. When you are ready to remove version 1, delete the \PRISM directory. Be sure to keep your data files.
- Version 2 reads all your version 1 data files. The files saved by version 2, however, cannot be read by version 1.
- Version 1 ran under SoftWindows version 1 on a Power Macintosh. Prism version 2 does not, but runs fine under SoftWindows version 2.

What's new in version 2?

Statistical analyses

Prism now conveniently combines statistics and scientific graphics in one complete package. We've added all the basic statistics tests:

- Paired or unpaired t test, or nonparametric Mann-Whitney or Wilcoxon tests, to compare two columns.
- Ordinary or repeated measures ANOVA to compare three or more columns. Follow with the Tukey, Newman-Keuls, Dunnett's or Bonferroni post tests, or the post-test for trend.
- Kruskal-Wallis or Friedman nonparametric one-way ANOVA. Follow with Dunn's post test.
- Fisher's exact test or the chi-square test to analyze contingency tables. Also calculate the relative risk and odds ratio (with confidence intervals).
- Pearson or Spearman (nonparametric) correlation.
- Kolgoromov-Smirnov normality test to assess whether a distribution significantly deviates from a Gaussian distribution.
- One sample t test or Wilcoxon test to compare the mean (or median) of one column with a theoretical value.

Embed data and results tables on graphs or page layouts

Prism now lets you combine data tables, results tables, and graphs on one page. Embedded tables are linked to the original data and are automatically updated when you edit or replace the data. For example, you might want to include the results of nonlinear regression on the same page with a graph. Or, you might want to place the P value from a t test on a bar graph.

- As easy as copy and paste. Select a portion of a table, copy to the clipboard, and then paste onto a graph or layout.
- Customize the table's appearance. Add a border or grid lines. Include or exclude column titles. Change the font and point size.
- Embed an entire data or results table, a selected region, or a single value.
- Create text tables as well as numerical tables.
- Place any number of tables on one graph or layout page.

More control over every graphing detail

We've added many more options so you can create the graphs you need:

- Move graph title and axis titles closer or further from the graph.
- Start and stop log axes anywhere (not just on whole log values). Place more custom ticks on each axis.

- Choose larger symbol sizes. Control the line thickness of error bars, bar borders, and open symbols. Make open symbols transparent so you can distinguish overlapping symbols.
- Label bar graphs with vertical text so you can use longer bar labels. Control spacing between groups of bars. Choose more fill choices for histograms. Create column bar graphs in one step.
- Write text using 2 byte fonts (i.e. Japanese or Korean). Insert Wingding characters (i.e. ❶, ➔, ⇨, ✓,✹). Place a border around a block of text. Choose from three arrowhead sizes.
- Plot median on column scatter graph to accompany results of non-parametric comparisons.

More analysis choices

- When calculating unknown values from standard curves, you can now enter replicate unknown Y values and calculate X for each replicate.
- Calculate two-way ANOVA with up to 16 replicate values.
- Pick the number of significant digits you want to see in the results.
- Choose from 90%, 95% or 99% confidence intervals for mean (column statistics), fit variables (nonlinear regression) and r (correlation).
- Enter longer user-defined equations for curve fitting.
- New template file contains equations for more sophisticated analyses of radioligand binding data.
- Column statistics now includes coefficient of variation and geometric mean.
- When normalizing data to the first and last values in a data set, you can remove these values from the results.

Better tools to help you organize complex projects

- Rename a sheet easily. Click in the sheet name in the toolbar and edit.
- The streamlined GoTo button makes it easier to navigate a complex project.
- After you import data, Prism automatically gives the data table sheet the name of the imported file.
- When creating a new data table, choose whether to graph those data. For example, skip the graph if you'd rather graph the data after a transform.
- Enter up to 16 replicates per data set (up from 5).
- Easily combine projects with the new File merge command.

- New Analyze dialog makes it easier to choose precisely which data sets you wish to analyze and to change that choice later.
- New dialog makes it easier to add and replace data sets on a graph.
- View graph thumbnails when choosing a graph to place on a layout so you can quickly find the right graph. New layout template lets you place three graphs on a page.
- Import and export GIF files, the graphics format used on the World Wide Web.

Work more efficiently

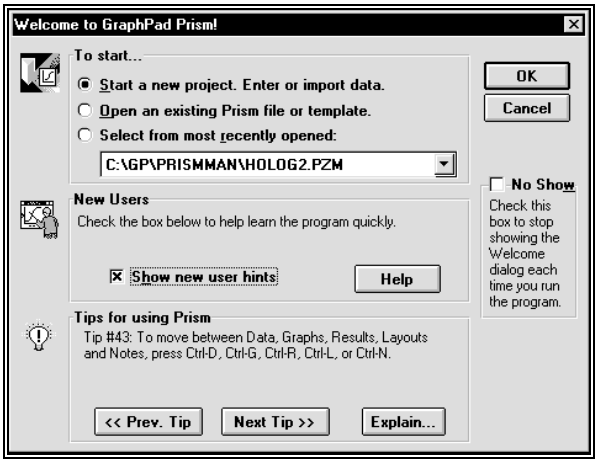
- Get to work quickly by opening one of the nine most recently used files directly from the Welcome dialog.
- Analyze data quickly by choosing one of the seven analyses you have used most recently from a shortcut list on the Analyze dialog.
- Right mouse shortcuts make common commands more accessible.
- Quickly locate the help you need by viewing a hierarchical list of help screens in the help navigator.
- Use Enter key to enter data in order. Moves the insertion point to the right within the same data set, otherwise moves down.
- Change analysis choices instantly by clicking the Analysis Parameters button.
- Automatic graphs are smarter, so you have to make fewer changes. If you change the X format from text to numbers, Prism changes the bar graph into an XY graph. If you don't enter X values, Prism creates a column scatter graph.
- Arrange objects on the page with new commands to center all objects on the page, align selected objects, equalize horizontal or vertical spacing, and display a grid on the screen.
- Group objects so they stay together while you move them and adjust their properties. Objects stay grouped as a unit until you choose the ungroup command.
- New compact file format saves disk space.

Tutorial

Before you begin using Prism with your own data, we recommend that you follow this brief tutorial. It will introduce you to most of Prism's main features in approximately 30-45 minutes.

Step 1. Start Prism

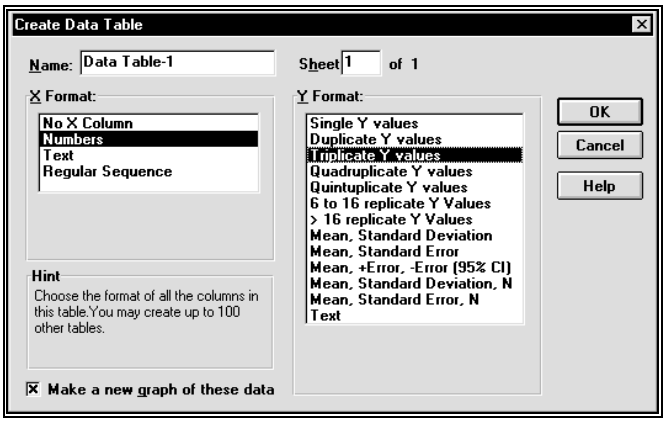
After installing and starting Prism, you will see the Welcome dialog.



Check that the dialog settings are set as shown above (to start a new project and to show new user hints) and press OK.

***If you don t see the Welcome dialog**: 1. Click the Cancel button on any dialogs you see. 2. Pull down the Edit menu, select Options, then select Program Options. 3. Choose to display the Welcome dialog on start up. 4. Quit Prism, and begin again.*

You will now see the Create Data Table dialog which lets you format the data tables to facilitate data analysis and graphing.

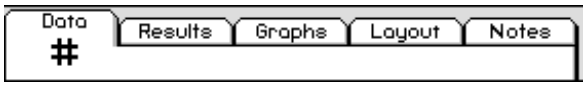


Select numbers for the X format, and triplicate Y values for the Y format. Then press OK.

Prism creates the data table and then superimposes a New user hint. These dialogs appear throughout the program to help beginners.

After reading the New User Hint, press OK.

Prism displays an empty data table. Note the five yellow tabs at the top of the window.



The tabs represent the five sections of every Prism project. The Data tab is on top because you are now in the data section.

Section	Explanation
Data	Enter or import data.
Results	Results of statistical analyses, curve fits, and data manipulations.
Graphs	Graphs automatically generated by Prism and graphs you create from scratch.
Layout	Arrange several graphs, drawings, tables and text on a page.
Notes	Text editor for recording experimental protocols, conclusions, etc.

Step 2. Enter data

To import data:

1. Position the insertion point in row 1 of the X column.
2. Pull down the File menu and select Import.

3. Read the new user hint, then press OK.
4. Set the file type to "Text".
5. Choose the file **tutorial.txt** in your Prism2 directory.
6. On the next dialog, confirm that you want to read all data from row 1 to row 13.

Your data table now looks like this:

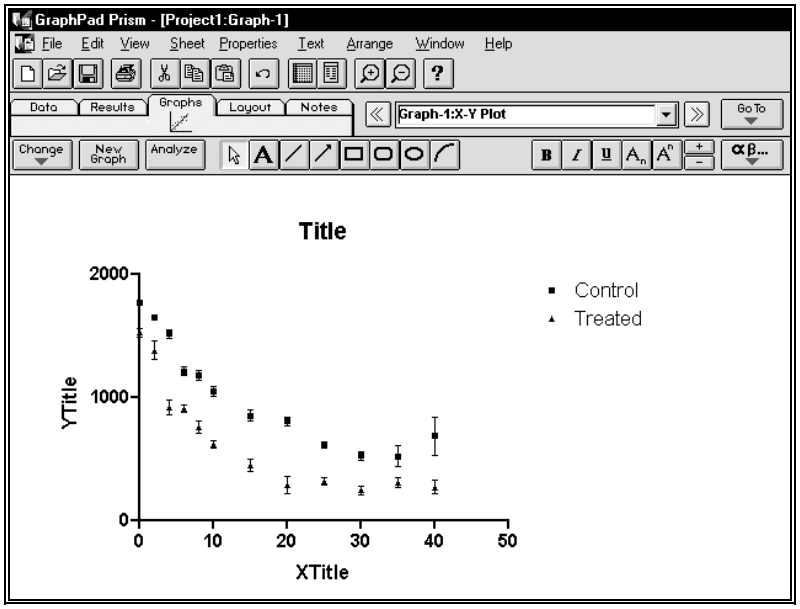
	X Values	A			B		
	X Title	Control			Treated		
	X	Y1	Y2	Y3	Y1	Y2	Y3
1	0.00	1750.00	1765.00	1773.00	1456.00	1543.00	1575.00
2	2.00	1654.00	1631.00		1235.00	1446.00	1462.00
3	4.00	1515.00	1454.00	1568.00	1020.00	924.00	804.00
4	6.00	1243.00	1175.00		961.00	874.00	875.00
5	8.00	1098.00	1187.00	1245.00	711.00		804.00
6	10.00	1032.00	987.00	1123.00	663.00	569.00	622.00
7	15.00	874.00	908.00	765.00	532.00	431.00	376.00
8	20.00	754.00	869.00	789.00		354.00	220.00
9	25.00	653.00	609.00	567.00	365.00	275.00	309.00
10	30.00	567.00	456.00	543.00	315.00	196.00	215.00
11	35.00	604.00	432.00		256.00	378.00	287.00
12	40.00	587.00	476.00	987.00	239.00	375.00	201.00

Notes:

- You imported data for two experimental conditions (data sets) placed side-by-side on the data table. Prism will analyze and graph these together.
- The empty cells indicate missing values. Prism handles missing data appropriately.
- Prism automatically named this data table TUTORIAL.TXT (the name of the file the data were imported from). To rename the sheet, click on the sheet name and change it. Every sheet in the project (data table, graph, etc.) has its own name.

Step 3. Make a graph

Click on the yellow Graphs tab to see the graph that Prism creates automatically.



Notes:

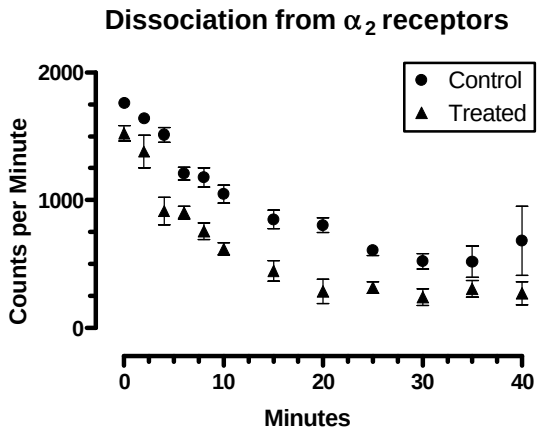
- Since you moved from the Data section to the Graphs section, the Graphs tab is now on top.
- You didn't have to specify any commands to calculate the error bars. Prism automatically averaged the replicate values and plotted the mean and standard error of the mean.
- You didn't have to tell Prism what data to graph. Prism automatically graphed all the data sets (columns) on the table. You may easily add or remove data sets from the graph.

Customize the graph by following these steps. If you wish, you may skip these steps and continue with the rest of the tutorial.

To	Do this
Offset the axes.	Double-click on an axis to bring up the Axes dialog. Drop down the Frame&Axes list and choose Offset X&Y axes.
Make the axes thicker.	Click on the Change button and choose Thickness of Axes and Frame. Select a thicker setting.
Change the symbols.	Double-click on a symbol to bring up the Symbols and

	Lines dialog. Change symbol size and shape.
Change the axis numbering.	Double-click on the X axis to bring up the Axes dialog. Click on the box labeled Auto to deselect it. Then change the range and tick intervals, and add minor ticks.
Move the legend.	Click on the Control legend to select it. Hold the Shift key and click on the Treated legend to select it also. Drag both legends on the graph to move them.
Frame the legend.	Click the rectangle tool and draw a box around the legends.
Add titles to the axes.	Click on the automatic titles and edit. Use the formatting buttons in the third row of the toolbar to enter Greek letters and create subscripts.
Convert the error bars from SEM to SD.	Double-click on a symbol to bring up the Symbols & Lines dialog. Hold the Control key to make a global change to all data sets on the graph, and click on SD error bars.

Your graph should now look like this:



Click the Change button at the beginning of the third row of the toolbar to see an alternative way to edit the graph.

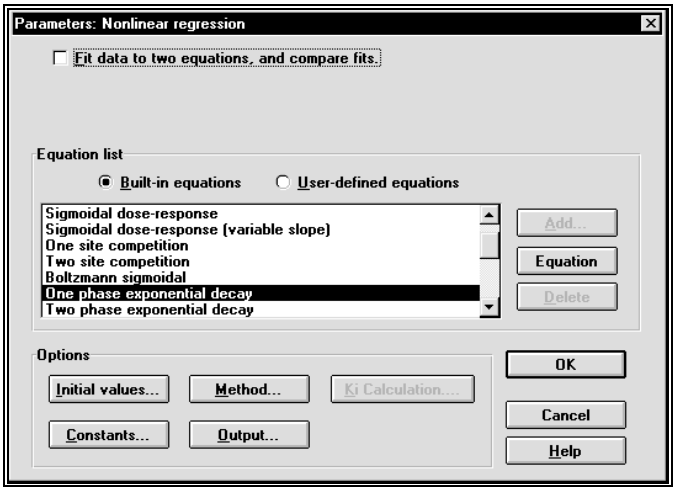
The Change button is probably Prism's most useful tool. It drops different menu choices in each of Prism's five sections.

Step 4. Fit a curve

Prism makes it very easy to fit curves with nonlinear regression. Even if you have no interest in curve fitting, follow these simple steps. Many of the same principles apply to other Prism analyses.

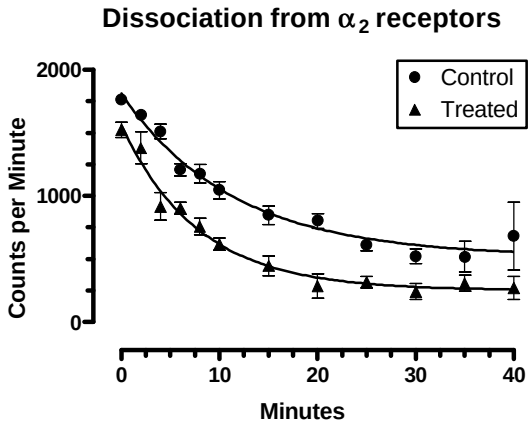
To fit a curve through your data:

1. Click the Analyze button.
2. Select Nonlinear regression. The Parameters dialog will appear.



3. Scroll through the list of equations, and choose "One phase exponential decay" (don't worry if you don't know what that means).
4. Press OK. After a few moments of calculations, Prism superimposes the curves on the graph.

Your graph now looks like this:



Notes:

- You didn't have to repeat the commands for each data set. Prism automatically analyzed the entire data table (both data sets).

- To view the numerical results, click on the Results folder tab.
- You have many more analysis choices when you start from a data table, rather than from a graph. To see these choices, go back to the data table and press Analyze. Press cancel once you have reviewed the options.

Step 5. Create a second graph

1. Click the Data folder tab to go back to the data table.
2. Press the New Table button. Press OK to create a new empty table. Note that you can create many (up to 100) data tables in one file.
3. Set the X format to text. Set the Y format to mean and standard error.
4. Enter these data:

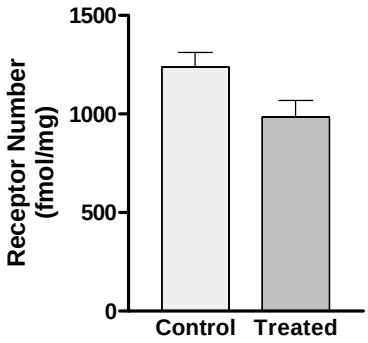
	X Labels	A		B	
	X Labels	Control		Treated	
	X	Y	SEM	Y	SEM
1	Receptor #	1237.0	76.0	984.0	87.0

5. Go to the Graphs section to see the bar graph that Prism made automatically. Your project now has two graphs. Click the << or >> buttons to the right of the folder tabs to switch between them.

6. Prism gave this graph sheet the simple name, Graph-2 Bar Chart. As your project grows, you'll find it easier to navigate if the sheets have more meaningful names. To rename, click on the sheet name and edit.
7. Customize the graph as follows.

To	Do this
Make the graph narrower.	Click on the X axis to select it. Point the mouse over the knob at the right of the axis. Hold down the left mouse button and drag to the left.
Enter two lines for the Y title.	Double-click on the default title to select it. It turns horizontal. Type the first line, press Return, and type the second line. Click elsewhere and the title becomes vertical again.
Delete the X axis title.	Double-click on the default title to select it, and press Del.
Delete the legends.	Click on each legend to select it. Then press Del.

Your graph now looks like this:



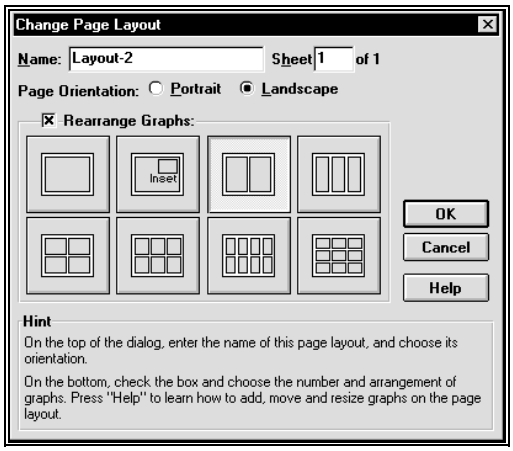
- Prism automatically created a bar chart because the X column was formatted as text. Prism creates an XY plot when the X column contains numbers, and a column scatter graph when the X column is empty.
- A file can contain many (up to 100) graphs.

Step 6. Arrange two graphs on a page layout

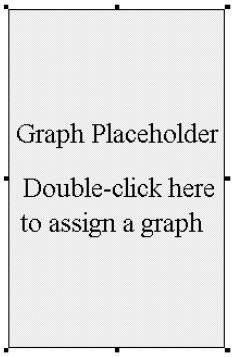
While working in the Graphs section of your project, you can only see one graph at a time. To place several graphs on a single page, create a page layout.

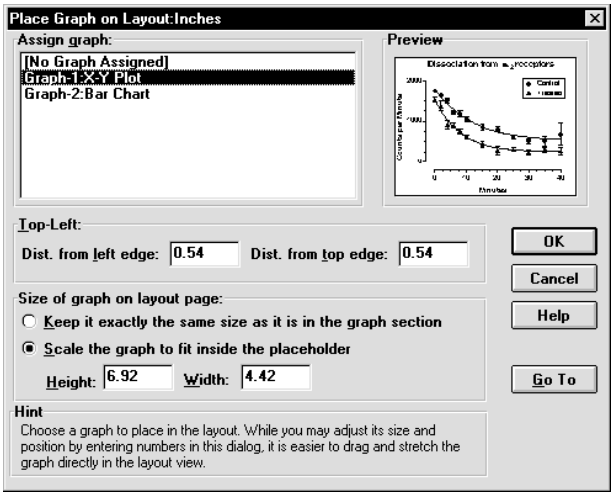
To create a page layout:

1. Press the Layout tab. Prism displays a page layout.
2. Press the Change button and choose Number and Arrangement of Graphs. Prism displays the Change Page Layout dialog.
3. Select a landscape orientation. Press the third button to create a page with two graphs, and press OK.



4. From the page layout, double-click on the left placeholder to bring up the Place Graphs on Layout dialog.





5. Choose Graph-1, then press OK.
6. Assign the other graph to the right placeholder.

Notes:

- Prism makes it easy to place 2, 3, 4, 6, 8 or 9 graphs on a single page. Use Prism's layout arrangements as they are, or customize them as needed. You can add or delete graphs, change their size, and move them to a different position.
- You can move graphs and change their size from the page layout. To edit the graph in any other way, you must go back to the Graphs section of your project. After you edit the graph there, Prism will automatically update the layout.
- In addition to graphs, layouts can contain imported pictures, drawings, text, and embedded tables.

Step 7. Embed a table

You can embed portions of data or results tables on either graphs or page layouts.

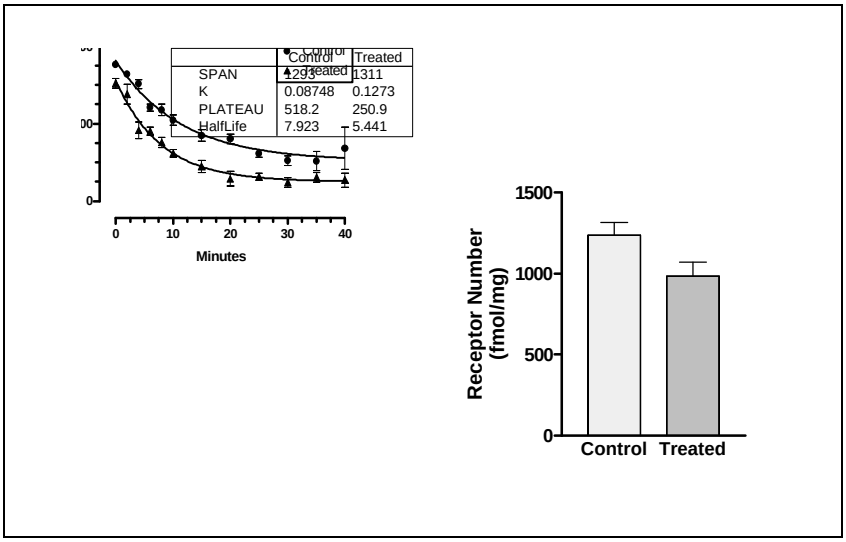
To place a portion of the curve fitting results table on the page layout:

1. Click on the Results folder tab to see the nonlinear regression results.
2. Click on the cell labeled SPAN. Hold down the left mouse button and drag down and to the right to select four rows and three columns.

	X Labels	A	B
		Control	Treated
	X	Y	Y
1	Equation 1		
2	Variables		
3	SPAN	1293	1311
4	K	0.08748	0.1273
5	PLATEAU	518.2	250.9
6	HalfLife	7.923	5.441
7	Std. Error		

3. Click the right mouse button, and select Copy from the shortcut menu.
4. Click on the Layout tab to return to the page layout.
5. Point the mouse above the left graph. Click the right mouse button and select Paste Table.
6. Double-click on the table. Make the border thicker. Select bold column titles and X titles. Press OK to leave the Table Properties dialog.
7. Move the table to align with the graph.
8. Move the mouse over a knob at the corner. Hold down the left mouse button and drag to change the table's size.

Your layout page now looks like this:



Notes:

- You can embed either results tables or data tables onto either graphs or layouts.
- The embedded tables are linked to your data and results. If you make any changes, Prism will automatically update the embedded table.
- You can embed a selected portion of the table (as above), the entire table, or just a single value.

Step 8. Exclude a value and update the graph

Note that the error bar on the last control point is much larger than the others because one of the replicate values is quite different than the others.

To exclude this outlier from the graph:

1. Click on the Data folder tab.
2. Go back to the first data table.
3. The Y3 replicate in row 12 for the controls (987) is much higher than its neighbors. Assume that you had noticed a problem with that tube during the experiment, and decide that you don't want to include the value in the graph or the analyses. Click once to select that number. Pull down the Edit menu and choose Exclude. The value 987 now appears in blue italics which indicates it is excluded. It will remain documented in the data table, but will now be excluded from all analyses and graphs.
4. Click on the Layout tab to return to the page layout. Prism automatically recalculates the error bar and the curve fit, omitting the excluded value, and updates the graph and embedded table.

Notes:

- Since Prism remembers the links between sheets in your project, it updates analyses, graphs, embedded tables and page layouts whenever you edit the data.
- You can use this feature to quickly analyze data from a repeated experiment. Replace one set of data with another and click once to go to the layout.
- If you don't want a graph or analysis to be updated, you can freeze that sheet. Pull down the Sheet menu and select Freeze.

Step 9. Save the file

To save the file, pull down the File menu and choose Save. Then enter a file name. The file will contain your entire project -- both data tables, both graphs, the page layout, and the analysis results.

Getting oriented

The Prism approach

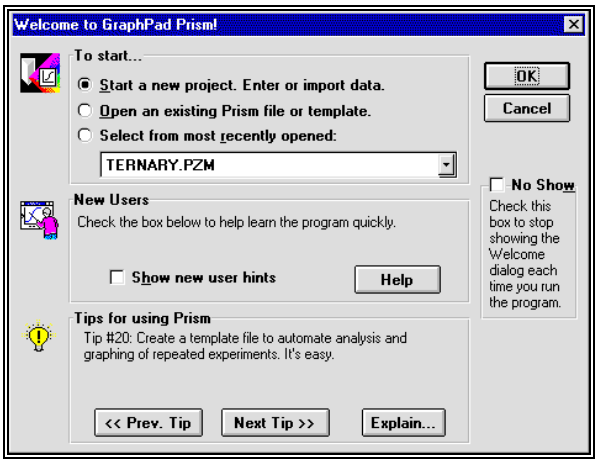
Prism makes it easy to enter, analyze and graph scientific data. Follow these steps:

1. Enter data onto data tables. The tables are structured to allow you to enter related data sets side-by-side. Columns are labeled to allow entry of replicate values or SD (or SEM) values. You may create several data tables in one file.
2. Analyze the data. Prism normally analyzes all related data sets at once. You don't need to repeat the commands for each data set. Results are presented in clear tables. Note that Prism uses the word *analyze* to include both data manipulation (i.e. transforms) and statistical analysis (i.e. regression).
3. Customize the graphs. Prism automatically creates graphs of all data tables and most analyses. You can change every detail, and can create additional graphs.
4. Place the graphs onto page layouts. Prism provides several layouts to start from. You can rearrange the graphs, add new graphs, and embed data or results tables.

Starting GraphPad Prism

To start Prism, double-click on the GraphPad Prism icon. 

Prism displays the Welcome dialog.



Note: If the Welcome dialog does not appear when you start Prism, someone has chosen another option. See "Program options" on page 134.

Start new project or open an existing one?

You can create a new project or open an existing project or template. Even though some template files are provided with Prism, we recommend that you create new projects first. Wait until you are familiar with the program before trying the supplied templates.

New user hints

While you are learning Prism, be sure to select new user hints.

As you use different features Prism will display special dialogs that explain how to use the program, and point out possible ambiguities. Each New user hint contains an option box “Stop showing this and other new user hints”. When the hints are no longer helpful, check that box. Also see "Program options" on page 134.

Tips

Prism displays tips at the bottom of the Welcome dialog. Each time you start Prism, it displays a new tip. Press the Previous and Next buttons to browse through the tips. Press Explain to learn more.

To look at tips while running Prism, pull down the Help menu and choose Tips.

Projects, sections and sheets

To use Prism effectively, you need to understand how it is organized.

Projects

When you work with Prism, you work on a *project*. A project contains all the data, analyses, and graphs that you want to keep together. You decide its scope. A project can contain a single data table with graph. A project can contain all parts of a complicated experiment. A project can contain all data and graphs from a series of repeated experiments.

If you use a Windows word processor, consider the analogy between Prism projects and word processing documents. While using a word processor, you work on a document. While using Prism, you work on a project. Each word processing document is saved into its own file. Each Prism project is saved into its own file. If you open several word processing documents, each appears in its own window. If you open several Prism projects, each appears in its own window. Within that window, you switch between looking at data, graphs, results, layouts and notes.

Sections

Each Prism project is divided into five sections: data tables, results of analyses, graphs, page layouts (containing one or more graphs), and notes. Some of the sections may be empty.

Sheets

Prism uses the term "sheet" to refer to each individual data table, each set of results, each graph, each page layout, and each page of notes. Each section may contain up to 100 sheets.

At any one time, you can only view a single sheet.

Data sets

Unlike most programs, Prism's data tables are very structured. Each data table has one column for X and many columns for Y values. You choose a format for each set of Y values: single values, up to sixteen replicate values, or mean with SD or SEM (and N). For example, here is a piece of a table formatted for entry of duplicate Y values in each data set.

	X Values	A		B	
	X Title	Title		Title	
	X	Y1	Y2	Y1	Y2
1					
2					
3					

Each set of Y values (with its X values) is called a *data set*. A data table can hold up to 52 related data sets. In most cases, each data set represents one experimental condition. Within a data set, you may enter replicate (i.e. duplicate or triplicate) values.

The Prism tool bar

Prism's tool bar consists of three rows of buttons and tools.

If you don't see buttons on top of the Prism window, someone may have selected the command to hide the tool bar. To bring it back, pull down the View menu and select Show Tool Bar.

Top row: Buttons

New file, Open, Save, Print

Cut, Copy, Paste, Undo

Embed table, Start StatMate

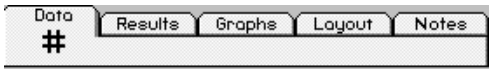
Increase zoom, Decrease zoom, Help

Tip: You don t have to remember which button does what. If you point at any button for more than a second, Prism will display a short explanation automatically.

Middle row: Navigation tools

Section tabs. A Prism project is organized into five sections shown as folder tabs. Click on a tab to go to that section.

Keyboard shortcut: Hold the Ctrl key and press D, R, G, L or N to go to the data, results, graphs, layout or notes section.



Select sheet. Each section may contain many sheets. (The term *sheet* refers to each data table, each graph, and so on). Press < < to see the previous sheet or > > to see the next. Or drop down the list to choose a particular sheet by name. To rename a sheet, click in this box and edit.

Keyboard shortcut: Hold the Ctrl key and press PgDn or PgUp to go to the next or previous sheet.

Go to a related sheet. A menu will drop listing all sheets linked to the one you are looking at. For example, if you are looking at a data table, the menu will list results of analyses of that data table, graphs of those data, and page layouts containing those graphs. Prism precedes the sheet names with [D], [R], [G] or [L] to denote data tables, results, graphs, and page layout sheets.

Third row: Other tools

The first button in the third row is always Change. The other buttons on this row come and go depending on which section you are in.

Change. This is the most useful button in Prism. Whenever you are stuck, try pressing the Change button. The Change button is dynamic. Depending on which section you are in, the choices on the Change menu differ.

New. Create a new sheet or duplicate a sheet. Each section has its own New button (i.e. New Graph, New Layout, etc.).

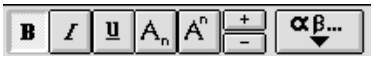
Analyze. A dialog lists all available analyses. Note that the analyses include data manipulations (i.e. transforms) as well as statistical analyses (i.e. regression). If you press Analyze while working with a data table, all these choices

are available. If you press Analyze while working with a graph, only regression choices are available.

Select view. The results of many analyses are presented on several pages called *results views*. For example, depending on which options you select, linear regression produces several views: tabular results, comparison of slopes, and residuals. While looking at results, drop down this list box to select the view you want to see.

Drawing tools. The first button is a pointer – press it to cancel from drawing. The other tools are used to draw lines, arrow, boxes, rounded boxes, ovals or circles, and arcs. When you are in the page layout section, an additional tool appears to the right. Use it to add additional graphs to the layout.

Text tools. Enter bold, italics, underline, subscript, superscript, increase font size, decrease font size, and add Greek letters.



Template files

Prism comes with 7 template files, designed to show you the power of Prism and to give you ideas for how to use Prism to analyze your data. Each template file opens to a page of notes that explains how to use it. You can test the template by importing a data file with the same name, but with the .txt extension. For example, to test the Scatchard template, import the file scatchard.txt.

Please note that these files are not designed to be complete solutions for data analysis. They are only examples to get you started. You'll probably need to change the template to meet your particular needs. You'll find the template files most useful after you have learned how to use Prism. The template files will not help you learn the basics.

Prism includes these template files:

File	Use
exams.pzt	Creates a histogram from a group of exam scores.

ria.pzt	Analyzes radioimmunoassay data.
scatchard.pzt	Uses nonlinear regression to fit saturation radioligand binding data to determine Bmax and Kd. Also draws a Scatchard plot.
simple.pzt	Creates a graph. The X values are contained in the template, because they never change. Enter Y values and click once to make a graph.
stdcurve.pzt	Analyzes a standard curve with linear regression, and then reads unknown X values from the curve.
sumtable.pzt	Shows the power of summary tables produced by nonlinear regression.
ttest.pzt	Compares two groups with an unpaired t test, and prepares a graph that includes the P value.

Prism also includes another template file, radiolig.pzt. For more information, see "Advanced radioligand binding template" on page 333.

Obtaining help

To bring up the on-line help press the Help button in any dialog or pull down the Help menu. While in the help system, you can:

- Press Contents to see the table of contents.
- Press Search to search for keywords.
- Press the < < and > > buttons to browse through help screens in sequence.
- Press Navigate to see the help navigator.

The navigator is a separate window that shows you the list of all help topics. Initially it shows only the chapter headings. Click on one of the book icons to open that chapter and see the subheadings.

To	Do this
Expand a main topic heading to see subtopics	Click on the closed book icon. 
Collapse subtopics to view only main topic.	Click on the open book icon. 

Expand all topics to see the entire tree.	Pull down the Navigator's Topic menu and choose Expand all.
Collapse all subtopics to see only chapter titles	Pull down the Navigator's Topic menu and choose Collapse all.
Go to a help topic	Click on the topic title

The contents of the Help screens are identical to the manual. When you jump to a help screen, you may wish to know how that screen fits in with the others. Click the navigate button and the navigator will show you an outline, and you'll be able to see the context of the help screen you are reading.

Entering data

Structured data tables

Prism's data tables are structured to aid data entry, analysis and graphing.

The first column in each data table is for X values which may be numbers or text (for bar graphs). Enter up to 52 related sets of Y values in the rest of the table. Each set of Y values may be entered as mean only, 2-16 replicate values, or as mean and SD or SEM.

Here is a portion of a data table formatted for triplicate Y values.

	X Values	A			B			C		
	X Title	Title			Title			Title		
	X	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3
1										
2										
3										
4										
5										
6										

Unless you specify otherwise, Prism analyzes and graphs all data sets on the data table at once. You don't need to repeat commands for each data set.

You may create up to 100 data tables in a single file. Enter related data sets that you want to analyze and graph together on one data table. Enter unrelated data on other tables.

Note that each column usually represents results from a different experimental condition. Unlike many statistical programs, Prism is not designed so that each column is a different variable.

When you create a new project file, Prism prompts you to create the first data table. You may create up to 99 additional tables in the same file.

To create a new data table:

1. If you are not already looking at a data table, press the Data tab.

2. Press the New Table button (on the left side of the bottom row of buttons).

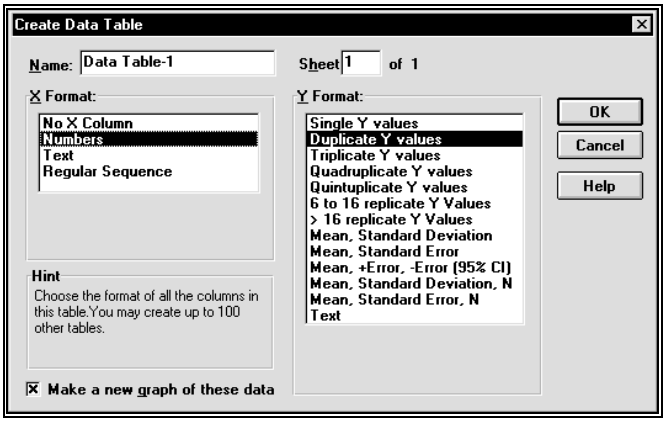
   The next dialog asks whether you want to create a new data table or duplicate the one you are looking at.

3. If you choose to create a new data table, Prism will prompt for the X and Y column format.

   If you choose to duplicate the data table you are looking at, you may then edit the copy. Duplicating is useful even if you only want to keep some of the values. For example, you might want to erase all the Y values but keep the duplicated X values.

Column formats

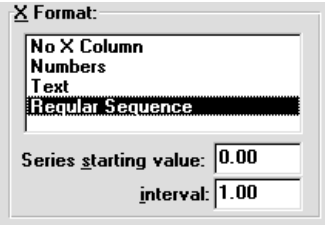
When you create a new data table, Prism prompts you to define its format.



You can change the format of an existing table (Press Change and choose Column format). Change column format only if the table is empty, or if the column headings don't match your data. When you change column format, Prism only changes the labels of the columns. It does not perform any calculations. Don't change the column format when you really want to create or change error bars on graphs. Prism can plot error bars directly from replicate Y values, and can convert between SD, SEM, and 95% CI.

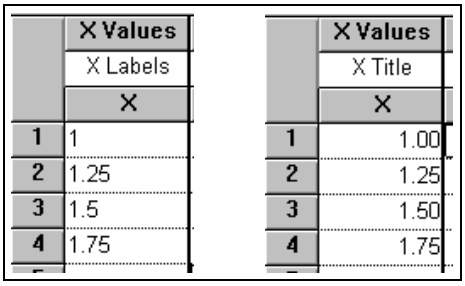
X format

Format	Explanation
No X column	You don't need to have an X column. Without one, you can create column scatter and box-and-whiskers graphs

	and can compare columns with a t test or ANOVA.
Numbers	You'll use this format most often. Create XY graphs and perform linear and nonlinear regression.
Series	Use when X values are regularly spaced – for example time points or fraction numbers. You enter the first X value and the interval between succeeding values. Prism calculates all the X values and displays them in bold. You may not edit individual values in the series. 
Text	Use to make a bar graph.

Notes:

- It is easy to forget that you have formatted the table for text X values, and mistakenly enter numbers. The X values on the left are formatted as text. Note that they are left justified. The X values on the right are formatted as numbers. Note that the numbers are right justified with aligned decimal points.



	X Values X Labels X
1	1
2	1.25
3	1.5
4	1.75

	X Values X Title X
1	1.00
2	1.25
3	1.50
4	1.75

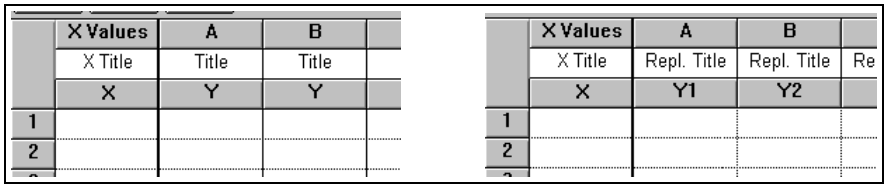
- If you choose a regular sequence, you don't specify the end of the sequence. Prism generates as many values as it can (until the bottom of the table). The range of the graph is determined by how many Y values you enter. Ignore the extra X values.

Y Format

Format	Comments
Single Y values	Use if you don't have replicate values, or if you want to only enter the mean value.
2-16 Replicate values	Choose duplicate, triplicate, quadruplicate or quintuplicate formats. Or choose 6-16 replicates, and enter the exact number. Allow for the maximum number of replicate values in the data sets you want to analyze and graph together. When you enter data, leave some cells blank when you have fewer replicates.
> 16 replicate values	The table will contain only a single data set, with 52 columns for replicate values. There will be only one type of symbol when the data are graphed, and all data will be analyzed together.
Mean with SD or SEM	Use if your data is already averaged.
Mean with SD or SEM and N	If you enter N, Prism can switch between SD, SEM and 95% CI error bars, and can calculate statistical comparisons (t tests, ANOVA).
Mean, - error, + error	Create asymmetrical error bars. Don't enter the high and low values, enter how far the error bar extends below and above the mean.
Text	Create a text table to embed in a graph or layout. You cannot make graphs from these tables.

Notes:

- You don't have to enter SD or SEM to graph error bars. Prism can graph error bars directly from replicate (i.e. triplicate) values.
- It is easy to mix up data tables with single Y values (left) and those with many replicate values (right). The table on the right below is formatted for replicate Y values. The entire table is for a single data set. Note that the Y columns are separated by dotted lines. Each row represents many replicate (repeated) measurements of the same outcome. The data table on the left is formatted for single Y values. Each column represents a different experimental condition. Note that the Y columns are separated by solid lines.



	X Values	A	B
	X Title	Title	Title
	X	Y	Y
1			
2			

	X Values	A	B	
	X Title	Repl. Title	Repl. Title	Re
	X	Y1	Y2	
1				
2				

Working with the data table

Moving the insertion point with the keyboard

Move the insertion point with the arrow keys or these shortcuts:

Key	Does this
Tab	Move to right.
Shift-Tab	Move to the left.
Enter	Move to the next Y value. If you have single Y values (mean only) move down to the next row. If you have multiple replicates (or mean, SD, etc.) then move to the right, within the same data set if possible. Otherwise move down to the next row.

Editing values

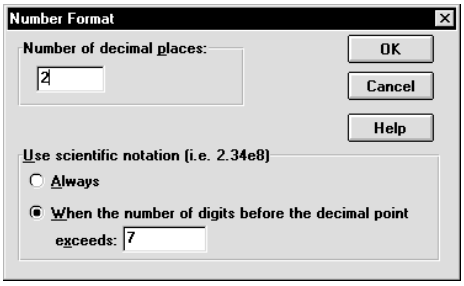
When you place the insertion point in a cell that contains a value, the number in that cell will be selected. When you start typing, you will erase the existing number and enter a new one. To edit an existing number, click once to go to the cell and then click again to place the insertion point inside the cell so you can edit.

Numerical format

Prism automatically chooses the number of decimal points to display, based on the first value you enter in each column. All numbers you enter later will be displayed in the same format unless you change it.

To change the number of decimal points displayed:

1. Select the column or columns you wish to change.
2. Click the Change button and choose Numerical format.
3. Enter the number of decimal points in the dialog.



Notes:

- You can only change numerical format for an entire column (or several columns). It is not possible to change the numerical format of selected cells.
- Altering the numerical format does not change the way Prism stores numbers. No matter how the numbers are displayed, they are stored internally with 8 significant figures.
- When you place the insertion point in a particular cell in the data table, Prism expands that value to show all digits. When you move to another cell, the value reverts back to the selected numerical format.

Column widths

Column widths are determined automatically. There is no way to manually adjust the width of selected columns.

Sorting data

If you choose to graph data with point-to-point lines connecting the points, the points will be connected in the order they appear on the data table. If the X values are not in order, the connecting lines will jump back and forth.

To sort your data by X values:

1. Make sure you are looking at the table you want to sort.
2. Pull down the Edit menu.
3. Choose Sort.

Excluding data

If a value is too high or too low to be believable, you can exclude it. Excluded values are shown in blue italics on the data table, but are not included in analyses and are not shown on graphs. From the point of view of analyses and graphs, it is just as if you had deleted the value. But the number remains on the data table to document its value.

Note: If you want to exclude entire data sets from a graph or analysis, this is not the best way to do it. You can select which data sets to analyze more directly. See "Analyzing data" on page 146.

To exclude data:

1. Select the cell or cells you wish to exclude.

   Pull down the Edit menu and choose exclude. The excluded values appear in blue italics.

2. Repeat the process to include the value again.

Tip: If you want to run some analyses both with and without the excluded values, copy the data table and exclude values from only one of the copies.

Using the clipboard

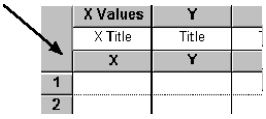
Prism uses the clipboard in a standard way. You need to understand four terms.

Term	Definition
Select	Indicate which region on the table you want to work with.
Copy	Place a copy of the region on the clipboard. Leave the original data alone.
Cut	Place a copy of the region on the clipboard. Erase the original data.
Paste	Paste the contents of the clipboard onto the data table at the indicated position.

Selecting

Use the clipboard to copy data from one location and paste it somewhere else. Before copying, you must select a region on the data table.

To Select	Mouse	Keyboard
A range of data.	Point to one corner of the block. Hold down the left mouse button and drag to the opposite corner.	Move to one corner of the block. Hold down the Shift key and move to the opposite corner (using arrow keys).

One or more columns.	Click on one of the column headers (“A”, “B”, etc.). Drag over the desired range of columns.	Hold Ctrl, and press the spacebar.
One or more rows.	Click on one of the row headers (“1”, “2”, etc.). Drag over the desired range of rows.	Hold Shift, and press the spacebar.
All data on the table.	Click here. 	Ctrl-A

Copy, cut, paste

There are four ways to access the cut, copy or paste commands:

- Press buttons on the toolbar. 
- Click the right mouse button and choose commands on the shortcut menu.
- Pull down the Edit menu and choose cut, copy or paste.
- Hold the control key, and press X, C, or V.

Notes:

- Before pasting, be sure to position the insertion point. You may also need to change the column format of the data table to match the new data.
- When pasting data, Prism simply places the data from the clipboard onto the data table. Prism does not attempt to figure out where to put X values, Y values, and error bars. It does not perform any calculations. It does not distinguish between columns that denote replicates, those that denote error values (SD) and those that denote distinct data sets.
- Prism can paste text into column titles, and into the X column if it is formatted for text.

Deleting data

The DEL Key

After selecting a range of data, press the DEL key to delete the selected range. Deleted data are not placed on the clipboard. Other numbers are not moved on the table to fill the gaps.

Even if you have selected an entire data set, pressing DEL does not completely erase the data set. All values are erased, but the data set itself still exists and it is still linked to analyses and graphs. If you enter new data, Prism will update the analyses and graphs. To completely delete a data set (and its links) pull down the Edit menu and choose Delete.

Note: The DEL key is the same as the Edit Clear command, but is not the same as the Edit Delete command.

Delete command

Select Delete from the Edit menu to delete a block of data completely, moving other data on the table to fill the gap.

*Note: Selecting the Edit Delete command is **not** the same as pressing the DEL key.*

If you have selected one or more entire rows or columns, Prism will delete them. Remaining numbers move up or to the left to fill the gap.

If you have selected a range of values, Prism presents three choices: Delete entire rows, delete entire data sets, or delete just the selected range (moving other numbers up to fill the gap).

Deleting an entire data table

To delete an entire data table, pull down the Sheet menu and choose Delete. Warning: When you delete the data table, you also delete those data from graphs. Results based on the deleted data become orphaned (see "Orphaned results" on page 150).

Importing data

Prism can import .GP files created by GraphPad InPlot (any version) and data formatted as plain ASCII text. Prism cannot import native spreadsheet files (i.e. *.WKS, *.XLS, *.WB1, etc.). Prism does not (yet) support object linking and embedding (OLE) or dynamic data exchange (DDE).

For related information, see “Exporting data” on page 143 or “Importing images” on page 112.

Importing text files

Prism imports text files with adjacent values separated by commas, spaces or tabs. Some programs refer to these files as ASCII files rather than text files. Each row in the file becomes one row on the data table.

All spreadsheet programs can create these files. From Excel (versions 4 or later) use the File Save As command and set the file type to Text or CSV (one uses tabs, the other commas to separate columns). Excel always exports an entire worksheet at once. From QuattroPro (version 5), first select the range of data. Then use the Tools Extract command and set the file type to Text.

To import data:

1. Change the column format if needed to match the new data.
2. Position the insertion point.
3. You may place the insertion point in any column and any row. That spot becomes the upper left corner of the imported block.
4. Select Import from the File menu.
5. Set the file type to Text and choose your file. Click OK.
6. Enter the first and last line of the file you want to import. The default values are the first and last line in the file.
7. Note that Prism automatically gave the data sheet the name of the data file.

Notes:

- It is not sufficient that the file have the extension .TXT. The file has to be a plain text file. If a file is not a text file, renaming it to have the extension .TXT won’t help.
- Prism can’t reformat text files. The data must be arranged as a table corresponding to the layout of a Prism data table.
- Often you'll find it easier to transfer data using the Windows clipboard. See "Using the clipboard" on page 41.
- When importing data, GraphPad Prism simply places the data from the clipboard onto the data table. Prism does not attempt to figure out where to put X values, Y values, and error bars. It does not perform any calculations. It does not distinguish between columns that denote replicates, those that denote error values (SD) and those that denote distinct data sets.

- Prism can import text into column titles, and into the X column if it is formatted for text.
- Hint: If the file contains more data than you want, import it onto a "scratch" data table (perhaps formatted as multiple replicates). Then create a new table for the selected data. Copy selected regions from the scratch table onto the clipboard, and paste onto the other data table.
- If you have trouble importing data, inspect the file using the Windows Notepad to make sure it contains only numbers clearly arranged into a table.

Importing data files (*.GP) created by GraphPad InPlot

Prism can import data from .GP files created by GraphPad InPlot. These files can contain tables of XY values, curves or bar graphs. Prism imports curves and XY values, but cannot import bar graphs (because the two programs format bar graph data so differently).

To import a .GP file:

1. Change the column format if needed to match the new data.
2. Before importing a curve, create a data table where the X values are defined as a sequence.
3. Position the insertion point at the first row of the X column.
4. Select Import from the File menu.
5. Drop down the "List files as type" control, and select InPlot files.
6. Choose a file.
7. Prism will warn you if the format of the data in the file does not match the column format of you selected.

Each *.GP file contains a single data set. If several files share the same X values, you may import them onto a single Prism data table. After importing the first *.GP file, move the insertion point to the first row of column B. Then import the second file. After you have imported the first *.GP file onto a table, Prism will import only the Y values of succeeding tables (ignoring the X values).

*Note: Prism cannot export data in *.GP format. However, InPlot can import text files created by Prism.*

Leave your data alone

The data table should contain your raw data, just as they were collected.

You don't need to change the format of the data table in order to change the format of your graph.

- You do not need to change the data table format to graph error bars. Choose SD, SEM or CI error bars from the graph.
- You do not need to erase one set of data to make room for another set. Place related data sets side-by-side on the data table. Create new data tables for data sets that don't share the same X values.
- You do not need to change the data to perform mathematical transforms. Use the Analyze button to transform the data. The results appear in a separate results table. The original data are not lost.

Entering multiple sets of data that don't share X values

Each data table has a single column for X and up to 52 sets Y values. What should you do if you have different X values for each set of Y values? There are two ways to solve this problem:

- Stagger the data entry. You don't have to start entering data in the first row. This example shows three data sets with three values each. The X values don't overlap.

	X	Y	Y	Y
1	1.0	1.0		
2	2.0	2.0		
3	3.0	3.0		
4	1.1		4.0	
5	2.1		5.0	
6	3.2		6.0	
7	1.2			7.0
8	2.2			8.0
9	3.2			9.0

- Enter each data set on a different table. You may create up to 100 tables in one project. The disadvantage with this method is that you lose some automation. Prism automatically analyzes all data on a table at once. If your related data are on different tables, you will have to repeat some commands.

The maximum number of values in a data set

The maximum number of data points depends on how you enter the data. If you enter a single Y value (no replicates, no SD or SEM), then you may enter up to 10,000 values per column. If you enter replicates or error bars, then the maximum number of values is reduced so that the total number of values is

10,000. For example, if you choose quadruplicates, you may only enter 2500 rows.

Embedding data tables in graphs or layouts

You can embed any portion of a data table onto a graph or page layout. Select the range of data, and copy to the clipboard. Then go to the graph or layout and paste. Double-click on the new table to bring up a dialog that lets you change its appearance. For more details, see "Embedding data and results tables" on page 113.

Introduction to graphing with Prism

Approach to making graphs

Prism provides all the tools you need to create customized graphs of scientific data. This chapter briefly presents an overview of how Prism makes graphs. The next few chapters present all the details.

Edit an automatically created graph or create a new graph?

You'll rarely need to create new graphs in Prism. Prism automatically creates graphs of all data tables and many analyses unless you tell it not to. In most cases, it will be easier to edit one of these than to create a new graph.

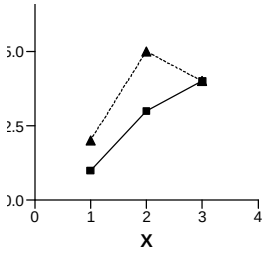
When you want to create a new graph, you have two choices:

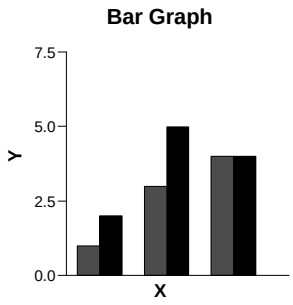
- Duplicate an existing graph, then edit one of the copies.
- Create a new graph.

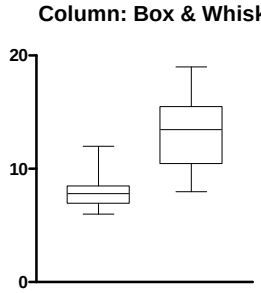
To duplicate or create, go to the Graphs section and press the New Graph button. See "Creating graphs" on page 64.

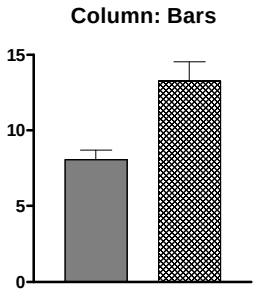
Choose the graph type.

Prism makes five kinds of graphs:









To learn more about different kinds of bar graphs, see "Choosing between three kinds of bar graphs" on page 64.

You choose the graph type on top of the Change graph dialog. When you create a new graph, Prism shows this dialog automatically. To change the graph type of an existing graph, press Change, and choose Graph type.

Select which data sets to graph.

The graphs that Prism makes automatically always include all data sets from one data table, perhaps with superimposed curve fits. You can easily remove data sets, and add additional data sets. There is no limit to the number of data tables or data sets that you can include on one graph.

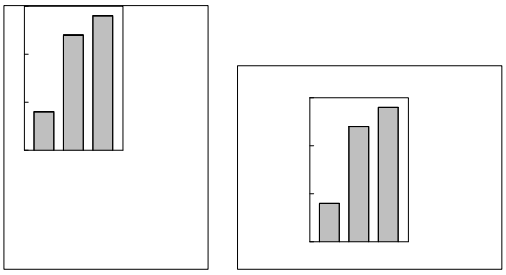
To change the data on a graph, press the Change button and select Data on graph. See "Selecting data to graph" on page 66.

Choose the page orientation.

If you plan to print the graph from the graph section, choose portrait or landscape orientation. If you only plan to print the graph from a page layout, the orientation in the graphs section is irrelevant.

With many programs you set page orientation in the Printer setup dialog. Prism is different because you can adjust the orientation of each sheet independently. Press Change, choose Page orientation, and make your choice at the top of the Change graph dialog.

Page orientation does not control the shape (aspect ratio) of the graph.



Portrait Landscape

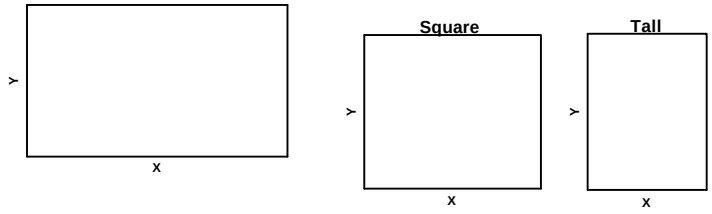
Adjust the graph shape and size.

The default graph size is about a quarter page. To change the default, see "Graph options" on page 134.

The easiest way to change the size of a graph is to place it on a page layout. Prism automatically expands or shrinks the graph so it fits on the layout. See "Page layouts" on page 101.

To adjust an individual graph's size and shape, do one of the following:

- Click once on an axis to select it. Then place the mouse over an axis end, hold down the mouse button and drag in the direction of that axis.
- Click the Change button and choose Graph size and shape. First choose the graph shape (wide, square, tall, or custom). Then enter the height or width of the graph. If you chose one of the standard shapes (wide, square or narrow), Prism will automatically adjust the other value. If you chose a custom shape, enter both height and width.



You can only change the shape of a graph from the Graphs section. When you place a graph on a layout, it grows or shrinks but does not change shape. You cannot change the shape of a graph by stretching it on a page layout.

Customize legends, titles, symbols, and axes.

The most important steps will be to customize the legends, titles, symbols, and axes. Prism offers plenty of flexibility. To see the possibilities, press the Change button to drop a menu.

Read more about	Page
Legends	68
Titles	69
Symbols and lines	71
Error bars	73
Graphing axes	85

Add text, images, drawings, and data or results tables.

Add text to annotate your graph. Import clip-art, photographs or equations. Draw lines, arrows, and boxes. Embed data or results tables.

Read more about	Page
Writing text	107
Drawing lines, arrows, ovals and boxes	111
Importing images	112
Embedding data and results tables	113
Aligning objects	119

Combine several graphs on a page layout.

The page layout lets you present multiple graphs on one page, along with text, drawings, images and tables.

When you create a new layout page, choose a standard arrangement of 1, 2, 3, 4, 6, 8, or 9 graphs. Initially the layout page will contain that many place holders rather than graphs. Double-click on a placeholder to assign a particular graph.

You can create additional placeholders, and can delete, move and resize placeholders. You may also write text, draw lines and boxes, import images, and embed data or results tables. See "Page layouts" on page 101.

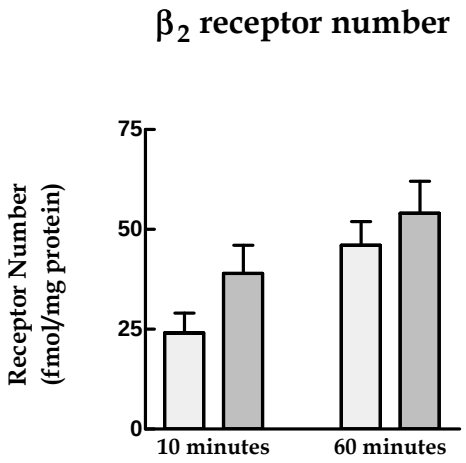
Print the graph or layout, or export to another program.

See "Printing" on page 137 and "Exporting graphs or layouts" on page 144.

Examples

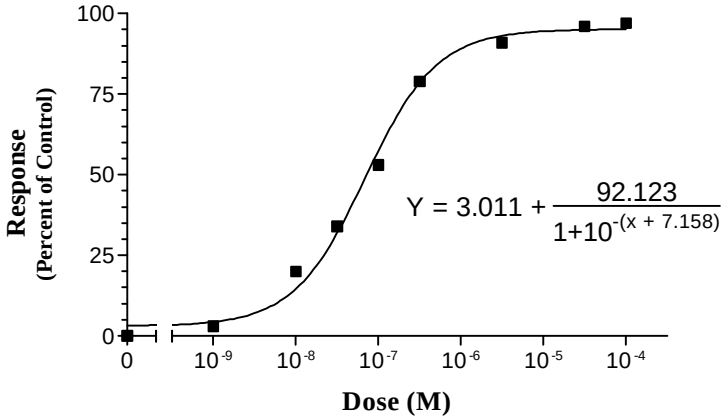
These examples show you some of Prism's versatility. Each example is accompanied by an explanation that tells you what to read for more details.

Example 1



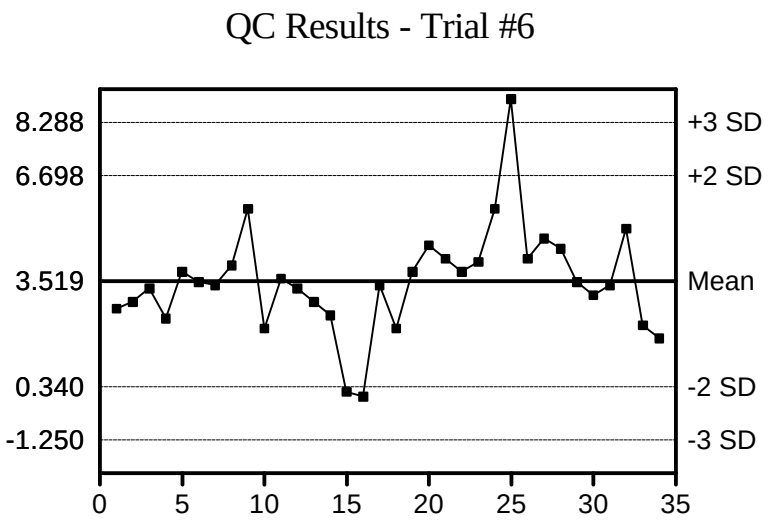
To	Do this
Graph the error bars.	You don't have to do anything special to plot error bars. Either enter data with multiple replicates (i.e. triplicate). Or enter data as mean, SEM (or SD), and N. Prism automatically plots SEM error bars. To change to SD error bars, double click on the bar, and click the appropriate option box.
Include Greek letters and subscript in the title.	Look on the tool bar for a button to enter Greek letters and another to generate subscripts. Also look at the Text menu for more options.
Move the title.	Move the cursor near the title until it turns into a two-headed arrow. Depress the mouse button and drag the title.
Make ticks point inward.	Double click on the axis to bring up the axis dialog. Under Tick options, set direction = right.

Example 2



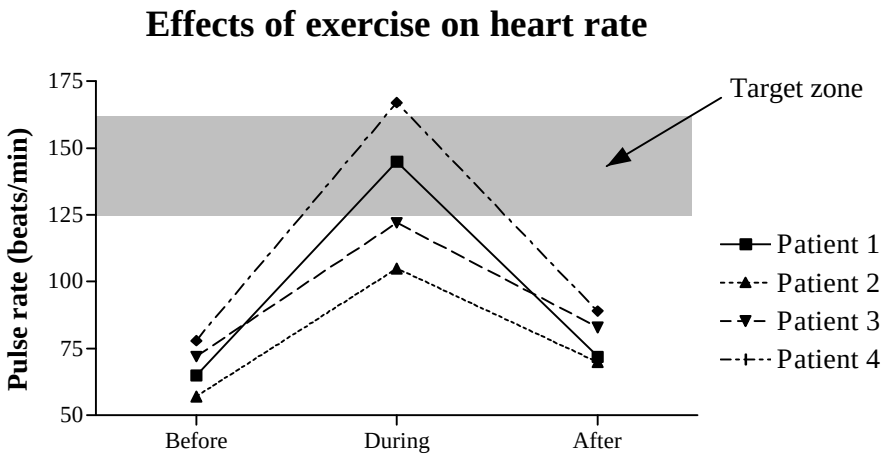
To	Do this
Create a logarithmic axis.	Enter the values as logarithms (e.g., enter a dose of 10-5 as X = -5). Double-click on the X axis to bring up the Axis dialog, and set the numbering format to powers of ten. See "Logarithmic axes" on page 93.
Make the first value be "0".	Enter the first X value as -10. Create a custom tick at: X=-10. Set the label to "0" with a tick. See "Custom ticks and labels" on page 96.
Create the gap in the X axis.	From the Axis dialog create a two-segment axis (insert a gap). Make the left segment be 10% of the length, ranging from -10 to -9.5, with an interval of 0.5 starting at -10. Set the right segment range from -9.5 to -3.5, with an interval of 1 starting at -9. See "Discontinuous axes" on page 90.
Add the equation.	Fit the data to a sigmoidal dose response curve, and note the results. Create the equation with three text objects. Draw the horizontal line separately. See "Writing text" on page 107 and "Drawing lines, arrows, ovals and boxes" on page 111.

Example 3



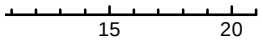
To	Do this
Create a second Y axis.	Double-click on an axis to bring up the Axis dialog. Select the option for two Y axes.
Calculate the SD lines.	Perform the column statistics analysis to determine the SD. Calculate two and three times the SD by hand. See "Column statistics" on page 181.
Draw the grid lines.	Double click on the left Y axis, and add four custom ticks with dotted lines. Check the option to plot custom ticks only (no regular ticks). Also add custom ticks to the right Y axis. For these, enter the text only with no ticks or lines. See "Custom ticks and labels" on page 96.

Example 4



To	Do this
Enter the data.	Enter before, during, and after values as X = 1, X = 2, and X = 3. Enter the pulse rate for each patient in a different column.
Create the X axis.	Double click on the X axis to bring up the Axis dialog. Set the X axis range to a minimum of -0.5 and a maximum of 3.5, with an interval setting of 1 starting at 1. Create custom ticks for "Before", "During", and "After". See "Custom ticks and labels" on page 96.
Add the target zone.	Draw a rectangle. Use the Properties menu to give the rectangle a solid gray fill and no border. Then select the rectangle, and use the Arrange menu to send it to be back, behind the data points.

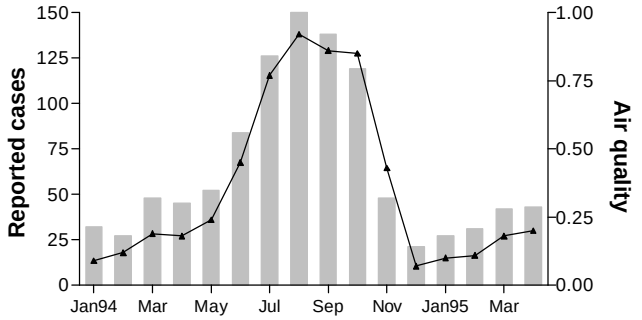
Example 5



bine (nM)

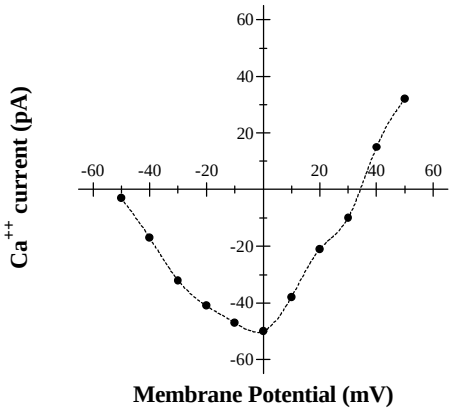
To	Do this
Fit the curve.	Use nonlinear regression with the one-site binding curve. Note the B_{max} and K_d values.
Add the grid lines at the Kd and Bmax.	Double-click on the Y axis to bring up the Axis dialog. Create a custom ticks at the B_{max} value with the text "Bmax" and a dotted line. Repeat to put the K_d grid line on the X axis.
Embed the table with the Bmax and Kd values.	Go to the nonlinear regression results. Select the four cells of the table, and copy to the clipboard. Go to the graph. Click the right mouse button, and select Paste. See "Embedding data and results tables" on page 113.
Graph error bars.	Enter the data as triplicate values. Prism graphs the error bars automatically. You don't have to do anything special.
Prevent the Y=1000 number from appearing on the axis.	Bring up the Axis dialog and place a custom tick (position Y = 1000, enter a space for the label, tick style = Tick).

Example 6



To	Do this
Enter the data.	Enter the X data on the data sheet as 1, 2, 3, … 16. Enter the number of reported cases in column A and the air quality index in column B.
Create the bars.	The bars are actually symbols plotted on an XY graph. Change the symbol shape to the fourth choice from the bottom (a solid bar). Set the color to gray and the size (thickness) to 6.
Create the right Y axis.	Press Change and select Axis range and ticks. Check the option to plot two Y axes. Then double click on a symbol to bring up the Symbols dialog. Check the option to plot that data set on the right Y axis.
Create the X axis.	From the Axis dialog, add custom tick marks at X = 1, 3, 5, etc. to replace the numbers with months.

Example 7



To	Do this
Create the curve.	Fit a spline curve.
Make the axes cross in the middle.	With the data shown here, Prism automatically crosses the axes at the origin. With other data, you might want to change the crossing point. Double-click on an axis to bring up the Axes dialog. Choose a custom origin, and enter the coordinates of the desired crossing point.
Position the axis numbering.	To avoid having numbers on top of data points, the X numbers were written on top of the axis rather than below. To do this, double-click on the X axis and set the numbering location to up.

Example 8

Unknowns calculated from standard curve	
X	Y

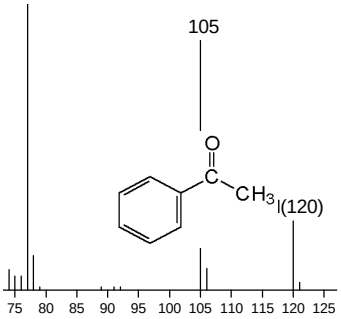
To	Do this
Fit the curve and determine the unknown values.	See "Reading unknowns from standard curves" on page 349.
Create the grid lines.	Double click on an axis to bring up the Axis dialog. Choose Frame with grid. See "Axis frame and offset axes" on page 86.
Graph the unknown values as lines extending down from the curve.	Press Change and select data on graph. The dialog lists the two data sets already on the graph (standard values and curve). Press Add, and select the results table with the unknown values. Double click on one of the unknown data points to pull up the Symbols and Lines dialog. Change the symbol shape to the last choice (a bar), and set the size to 0 or 1.
Embed the table with the unknown values.	Go to the results section and find the view that contains the XY coordinates of the unknown values. Select the table and copy it to the clipboard. Go back to the graph, press the right mouse button, and select Paste. Double click on the table to fine-tune its appearance. See "Embedding data and results tables" on page 113.

Example 9

P=

To	Do this
Create a column scatter graph.	Enter the control values in column A and the treated in column B. Leave the X column blank. Prism makes a column scatter graph automatically.
Offset the axes.	Double click on the Y axis to bring up the axis dialog. Select Offset X & Y axes.
Include the P value.	Use the t test analysis to compare the groups. Copy the P value to the clipboard. Paste it onto the graph as an embedded table (tables can be as small as one cell). Add the "P=" as a text object in front of the number. See "Embedding data and results tables" on page 113.

Example 10



To	Do this
Graph the spikes.	Create an XY graph. Double click on one of the data points to bring up the Symbols dialog. Change the symbol shape to one of the last 4 options (a bar) on the drop-down list. Set the size to 0. See "Symbols and lines" on page 71.
Add peak labels.	Use the text tool. See "Writing text" on page 107.
Add the chemical structure.	Create it in another program (we used ChemWindow from SoftShell International), copy to the clipboard, and paste. See "Importing images" on page 112.
Reduce the point size of the X axis numbering.	Click once on the X axis to select it. Press the  button on the tool bar to reduce the point size. Or pull down the Text menu and choose Font.

Creating Graphs

Automatic graphs

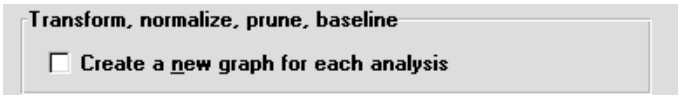
Prism automatically creates graphs of all data tables and many analyses. Prism chooses a graph type based on the format of the X column:

Format of X column	Automatic graph type
Text	Bar graph
Numbers	XY graph
No X column, or empty column.	Column scatter

Notes:

- Prism chooses the graph type based on the format of the X column. If you change the format of the X column (i.e. from numbers to text), Prism will automatically change the graph type (unless you set it manually).
- The New data table dialog includes an option box for creating a graph from that table. By default, this option box will always be checked. Uncheck it if you don't want an automatic graph. For example, you might choose not to create a graph if you plan to transform the data and only wish to see a graph of the transformed values.

- The Parameters dialog for transforming, normalizing, pruning, and subtracting baselines has a similar check box. Check the box to create a graph of the new data. You can decide whether this option is checked or unchecked by default in the Analysis options dialog. To reach this dialog, pull down the Edit menu and choose Options.



Creating graphs

There are two ways to create a graph:

- Duplicate an existing graph, then edit one of the copies. This is often the best approach.
- Create a new graph.

To duplicate a graph:

1. Press New Graph.
2. On the New graph dialog, choose to duplicate the graph you are looking at.
3. To change the appearance of the graph or the data sets included, press the Change button.

To create a new graph:

1. Press New Graph.
2. On the New graph dialog, choose to create a new graph.
3. On top of the Change graph dialog, choose a graph type (XY graph, bar graph, etc.).

   The bottom of the Change graph dialog shows the data sets included on the graph. Initially it is empty.
4. Press the Add button to bring up the Add data sets dialog.
5. At the top of the dialog, choose a data or results table. Select one or more data sets on the bottom of the dialog.

Choosing between three kinds of bar graphs

Prism makes three kinds of bar graphs: standard bar graphs, column bar graphs and histograms.

Standard bar graph

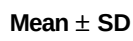
Each bar represents a single Y value or mean of replicate Ys. For example, if you enter three rows and two columns of data, the graph will have six bars. The graph could include error bars only if you had entered replicate values or had entered SD or SEM values. The bars representing each data set (column) have a distinct appearance and they may be separate, stacked, or interleaved (shown). See "Bar graphs" on page 75.

	X Labels	A	B
	X Labels	Control	Treated
	X	Y	Y
1	A	3.4	6.2
2	B	4.3	7.8
3	C	3.0	5.9

Column bar graph

Each bar represents the mean of an entire Y data set (column). Error bars are calculated from the range, SD, SEM or 95% CI of all the values (N = 5 in this example) in each column.

3.4	6.2
4.3	7.8
3.0	5.9
3.9	6.4
4.1	7.6



Histogram

A histogram is actually a type of XY graph. To turn an XY graph into a histogram, select one of the last four choices for shape on the Symbols dialog. The position of each bar is determined by the value in the X column (if you didn't enter numbers into the X column, you can't make a histogram). Each bar extends from the Y = 0 with a height determined by the value in the Y column.

	X	
	1.0	2.0
	2.0	4.0
	3.0	6.0
	4.0	8.0
	5.0	9.0
	6.0	7.0
	7.0	5.0
	8.0	4.0
	9.0	3.0

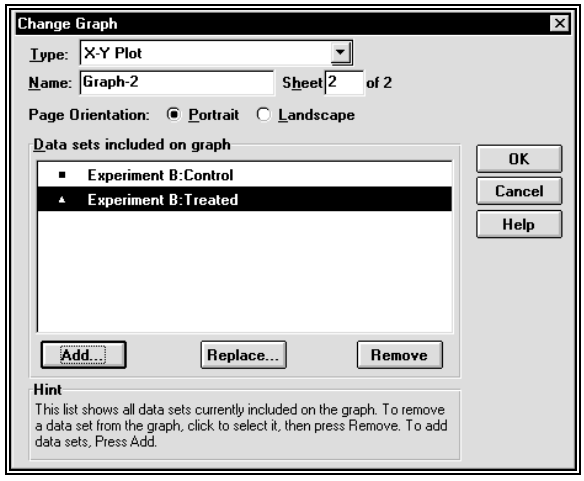
Histograms are often used to display frequency distributions. Prism can create a frequency distribution from a column of values. See "Frequency distributions" on page 163.

Selecting data to graph

When Prism creates a graph automatically, it includes all data sets from a single data table. You may add data sets from other tables, or remove selected data sets.

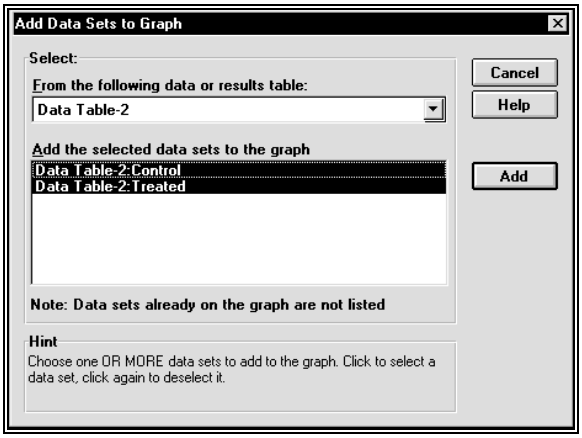
To add or remove data sets from a graph:

1. Make sure you are looking at the graph you want to change.
2. Press the Change button and choose "Data on graph".
3. The Change graph dialog shows you each data set on the graph identified with the symbol (or bar) used.



To	Do this
Remove a data set.	Select one or more data set names. Then press Remove.
Add one or more data sets from a single data or results table.	Press Add to bring up the Add data sets to graph dialog. Select the data or results table. Choose one or more data sets on that table. Press Add. Then press Close.
Add data sets from more than one data table.	Press Add to bring up the Add data sets to graph dialog. Choose a data or results table. Select one or more data sets. Press Add. Choose another data or results table. Select one or more data sets on that table. Press Add again. Press Close when done.
To replace one data set with another, without changing symbol color or size.	Click to select one data set. Press Replace. On the Replace data set dialog, choose a data or results table and choose a data set.

Use this dialog to add or replace data sets.



Notes:

- With many other programs, you define the range of data to graph by selecting a block on the data table. Prism is different. Selecting data on the table does not change what data is included on the graph.

- The curves generated by nonlinear regression are distinct data sets in a results table. To plot the original data points with a curve, you need to choose both. (Prism does this automatically in most cases).
- Here is a shortcut to bring up the Change graph dialog: Double-click on the background of the graph away from symbols and axes.

Legends

Prism creates an automatic legend on all XY and bar graphs.

When creating the legend, Prism matches the symbol shape, bar type, and color used in the graph. You can't change the shape, size or color of the symbol on the legend. If you change its appearance on the graph from the Symbols dialog, the legend will update automatically.

To	Do this
Move a legend.	Click once to select it. Drag it.
Move a group of legends.	Select all the legends, then drag to move them.
Edit the text.	Double-click on the text and edit it. The legend text will no longer be linked to the data set (column) title.
Change the font or size of text.	Select the entire legend, or a group of legends. Select Font from the Text menu.
Align legends.	Select all the legends by holding Shift and clicking on each. Pull down the Arrange menu and select Align objects.
Draw a border around a group of legends.	Use the drawing tools to draw a box (or a rounded box).
Delete a legend.	Click once to select it. Press DEL. The legend for each data set is a separate object.
Bring back a deleted legend.	Go to the Symbols dialog (or Bars or Columns) and select the option box "Show legend".
Use data set title as legend.	Go to the Symbols (or bars) dialog. Select "Use column title".

Notes:

- Initially, the text in the legend comes from the data set (column) title in the data table or is “Legend” if you didn’t enter a data set title. Although you can’t enter Greek letters, superscripts or subscripts in the data table, you can edit the legend directly on the graph.
- When you click on a legend to select it, it will show knobs on the corners. Resizing the legend by dragging the knobs has no effect. The symbol size is determined from the Symbols dialog. The text size is determined from the Font dialog.
- You can group all the legends together so they move as a block, and so you can change the font, size and color together. Select all the legends. Then pull down the Arrange menu and choose Group.
- Each legend is an independent object. To align and equally space the legends, see "Align objects dialog" on page 121.

Titles

Prism puts generic labels (“Title”) on top of the graph and on the X- and Y-axes. You may edit these standard titles and may add additional text anywhere on the graph.

To change a title:

1. Click in the title. If you are editing a Y axis title, it will be temporarily displayed horizontally.
2. Edit (or replace) the text. To enter more than one line, press ENTER between lines.
3. If you are editing a Y axis title, click anywhere on the graph to erase the temporary horizontal title and redraw the vertical title.

To delete a title:

1. Drag over the title to select it. If you are editing a Y axis vertical title, it will be temporarily displayed horizontally.
2. Press the DEL key.

To change the size or font of a title:

1. Select all or part of the title.
2. Pull down the Text menu and choose Font.

   or

   Click the + or - buttons to make the text bigger or smaller.

To left justify the title:

Usually the title is centered over the graph. Follow these steps to align the title with the left edge of the axis.

1. Put the insertion point in the title.
2. Pull down the Text menu.
3. Select Justify, and then Left.

To move a title:

1. Move the mouse near the title until the cursor turns into a two-headed arrow.

2. Hold down the left mouse button, and drag.

You can move titles closer or further from the axis, but cannot move them in the other direction. Titles are always centered under (or next to) an axis. If you want a title to appear somewhere else, erase it and then add the title as a free text object.

Notes:

- This section has explained about titles. Labels and legends work differently. See “Writing text” on page 107 and “Legends” on page 68.
- For information on Greek letters, international and math symbols, and changing font and size see "Writing text" on page 107.

Changing from one kind of graph to another

Prism makes five kinds of graphs (XY graphs, bar graphs, etc.). You may convert any graph type to any other type.

To change a graph from one type to another:

1. Make sure you are looking at the graph you want to change.
2. Press the Change button and choose “Graph type”.
3. Select a different graph type.

Notes:

- If you make a bar graph from a table with numerical X values, those numbers are used as group labels.
- If you make a XY graph from a table with text X values, Prism uses the row number as the X value (i.e. row 6 has X = 6).

Graphing symbols, lines, bars, and error bars

Symbols and lines

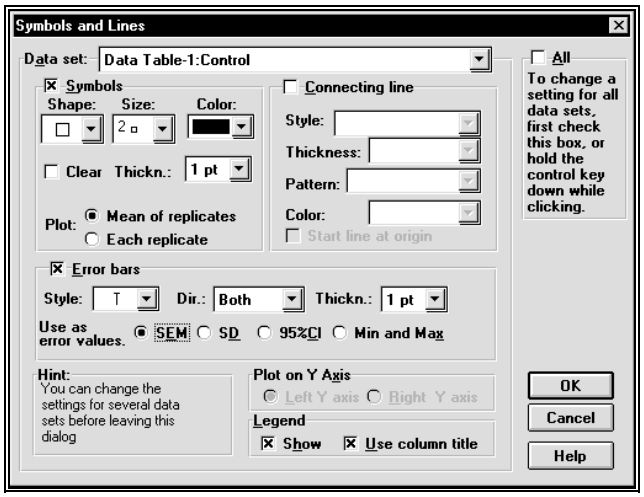
Use the Symbols dialog to change the appearance of symbols, connecting lines and error bars.

To bring up the Symbols dialog:

- Click the Change button, and select "Symbols & Lines" from the menu.

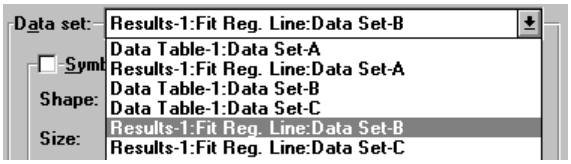
  or

- Double-click on any data point.



Choosing a data set

Choose symbols, lines and error bars one data set at a time. Select a data set from the drop down list, and choose the symbols, lines and error bars for that data set. Then select another data set and choose symbols for that data set. Repeat as often as necessary. Click OK to redraw the graph with all the changes.



Setting	Meaning
Symbols	To plot a symbol for each data point, select the "Symbol" check box and choose symbol type, size and color. If you pick an open symbol, also select a thickness. Check the clear option if you want point-to-point lines and other points to show through the open symbol. Otherwise the symbol will be filled with the background color.
Connecting line	To connect the symbols with lines, select "Connecting line" and choose the line's color, style (straight vs. staircase), thickness, and pattern (solid, dotted, dashed, etc.). Don't confuse the connecting line – which always goes from point to point – with a fit curve. If you have fit a curve through the data, the data and the curve are in separate data sets. For the data, choose symbols but no connecting line. For the curve, select a connecting line but no symbols.
Error bars	To plot an error bar with each data point, check the option box to create error bars and then select how they are to be calculated (SD, SEM, 95% CI, or range). Then choose how the error bar will be calculated (i.e. SD or SEM) and how it will appear (thickness, with or without caps). See "Error bars" on page 73.
Plot on which Y axis?	If your graph contains both left and right Y-axes, specify which axis to use for each data set. Initially Prism plots all data on the left Y axis. See "Creating two Y-axes" on page 97.
Legend	Select "Show legend" to create a legend for this data set. Select "Use column titles" to link the legend to the column title entered on the data table. Uncheck this box if you don't want the legend to change when you edit the column title. See

"Legends" on page 68.

You may also make global changes for all data sets. To do this, select "Change all data sets" on top of the dialog. While that check box is selected, any changes you make will apply to every data set on the graph. For example, you might want to increase the size of all symbols. Or you might want to change all error bars from SEM to SD.

Short Cut: Press the Ctrl key to temporarily make global changes. When you release the Ctrl key, you will revert to changing one data set at a time.

Symbols for survival curves, mass spectra and histograms

The fifth from the last choice is an upwards pointing tick. It is useful for plotting survival curves. See "Graphing survival curves" on page 228.

The last four choices are histogram bars. The bar will start at the X axis and go up (or down) to the Y position of the point. The thickness of the bar is set by the symbol size. If you pick the smallest size, the bar will actually be a spike suitable for graphing mass spectra. See an example on page 62. For a comparison of histograms with two other kinds of bar graphs, see "Choosing between three kinds of bar graphs" on page 64.

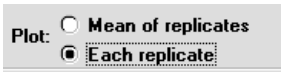
Error bars

Automatic error bar calculations

Prism plots error bars automatically from replicate data. The important word is automatic:

- You do not need to display the SD or SEM on the data table to see error bars on your graph.
- You do not need to choose an analysis to calculate the errors.

You don't have to plot error bars from replicate values. To plot each individual replicate, check "Each replicate" on the Symbols & lines dialog.



To include error bars on a graph:

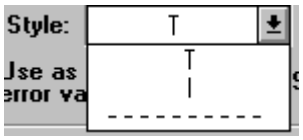
1. Check "Error bars" on the Symbols and Lines dialog (or the Bar Appearance dialog) This box will be gray if you have entered only a single value.

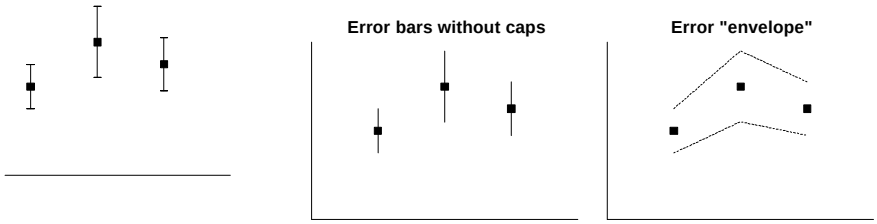
2. Choose between SD, SEM, 95% CI, or Min/Max error bars on the Symbols and Lines dialog (or the Bar Appearance dialog). Not all of these choices may be available. For example, if you entered your data as mean and SD only (no N), you can only plot SD error bars as it is not possible to calculate the SEM or CI without knowing N.

Note: Changing the settings on the Symbols and Lines dialog (or the Bar Appearance dialog) changes the appearance of the error bars for the data set selected at the top of the dialog. If you want to change the appearance of all error bars on the graph (say from SEM to SD), hold Ctrl while clicking. While holding Ctrl, the changes you make are global for the entire graph, rather than specific to a single data set.

Error bar appearance

Choose the error bar style on the Symbols and Lines dialog (or the Bar Appearance dialog). Choose plain error bars, capped error bars, or error envelopes. The last choice (shown in the dialog as - - - - -) does not show error bars, but instead creates an error envelope connecting the ends of the SD, SEM or 95% CI of adjacent points with a dotted line.





Also choose error bar thickness and whether error bars go up, down or both. This choice applies to the entire data set.

If you want some error bars to point up and others to point down, you need to enter your data on a data table for asymmetrical error values. You can then enter either the + error or the - error value for each data point individually.

Related information

Read about	Page
	169
Row means or totals	
Column statistics	181
Interpreting the SD, SEM and 95% CI	184

Bar graphs

Creating bar graphs

To create a bar graph automatically, create a new data table with the X column formatted for text. Before creating a bar graph, consider whether your data might be better plotted as a histogram or column bar graph. See "Choosing between three kinds of bar graphs" on page 64.

The Bars dialog

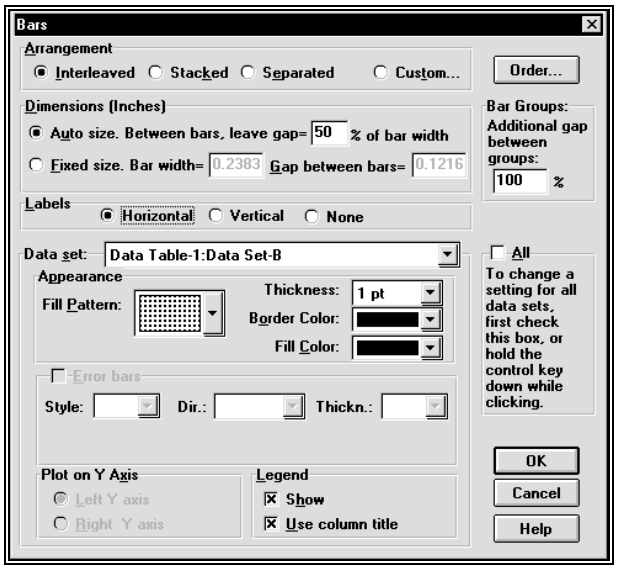
The Bars dialog lets you change the appearance and order of bars on a bar graph.

To bring up the Bars dialog:

- Double-click on any bar.

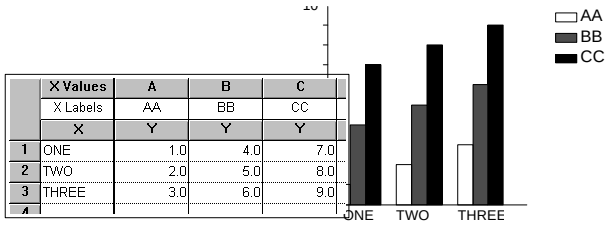
  or

- Press the Change button and select “Bar appearance”.

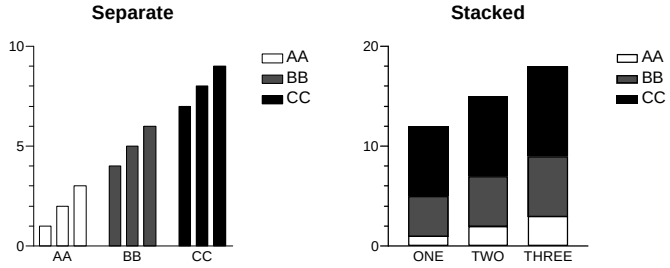


Arrangement

If your graph contains more than one data set, choose interleaved, stacked or separated.



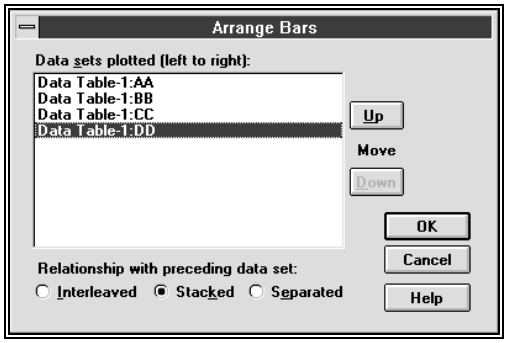
	X Values	A	B	C
	X Labels	AA	BB	CC
	X	Y	Y	Y
1	ONE	1.0	4.0	7.0
2	TWO	2.0	5.0	8.0
3	THREE	3.0	6.0	9.0
4				



With three or more data sets, create a custom bar graph to combine interleaved, separate and stacked bars.

To make a custom bar graph:

1. From the Bars dialog, select a "custom" arrangement. The Arrange Bars dialog immediately opens. (To return to this dialog, press the Order button on the Bars dialog.)



2. The data sets are presented in order from left to right. To change the order, click any data set and then the UP or DOWN buttons. Pressing UP moves the selected data set higher in the list, and thus moves it to the left on the graph.
3. Specify whether each data set is interleaved with, stacked on top of, or placed apart from the preceding data set.

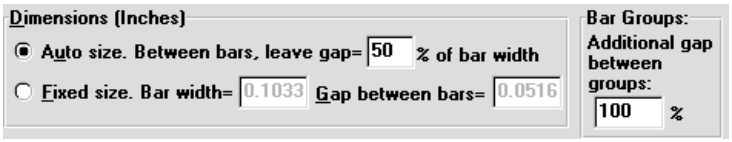
   In this example, data set B was stacked, data set C was separate, and data set D was stacked.

Two	2.0	5.0	8.0	11.0
Three	3.0	6.0	9.0	12.0

Dimensions

Let Prism choose bar widths automatically or enter the widths manually. If you choose manual bar widths, enter the width of each bar and the width of the gap between bars (in inches in the US and UK, in cm elsewhere). If you make the bars too large, all the bars may not fit on the graph.

If you choose automatic bar widths, Prism calculates the bar width based on the number of bars, the size of the gaps you choose, their arrangement, and the size of the graph. Enter the size of the gap between each pair of bars and between groups of interleaved or separate bars as a percentage of the width of a bar. In this example, the gap between adjacent bars will equal 50% of the width of one bar, and the gap between groups of bars will equal 150% of the bar width (the distance between adjacent bars plus the additional 100% gap between groups of bars).



Labels

Label the bars with the row titles from the data table. You may not edit the bar titles directly, but if you edit the row titles the graph will update. Choose whether you want to labels to be horizontal or vertical. Or check none to leave the labels off the graph.

Notes:

- If your data table has only one row, Prism will place the column titles, rather than row titles, under each bar.
- If you check the option box to make the labels vertical, the overall X axis title may be far below the titles. To move the title, see "Titles" on page 69.
- See “Writing text” on page 107 if you prefer to enter labels manually.

Data set

Choose the appearance of one data set at a time. Select a data set from the drop down list, and choose the fill pattern and color for that data set. Then select another data set and choose its fill pattern and color. Click OK to re-draw the graph with all the changes.

To make global changes for all data sets, select the check box on top of the dialog. While that box is checked, changes apply to every data set on the

graph. For example, you might want to plot all bars the same color or change all error bars from SEM to SD.

Short Cut: Press the Ctrl key to temporarily make global changes. When you release the Ctrl key, you will revert to adjusting one data set at a time.

Setting	Meaning
Appearance	Choose fill pattern and color, and the border thickness and color. The final appearance of the fill patterns is determined by the printer driver you use, and the appearance on screen is only approximate.
Error bars	To plot an error bar with each data point, check the option box to create error bars and then select how they are to be calculated (SD, SEM, 95% CI, or range). Then choose how the error bar will be calculated (i.e. SD or SEM) and how it will appear (thickness, with or without caps). See "Error bars" on page 73.
Plot on which Y axis?	If your graph contains both left and right Y-axes, specify which axis to use for each data set. Initially Prism plots all data on the left Y axis. See "Creating two Y-axes" on page 97.
Legend	Select “Show legend” to create a legend for this data set. The legend will be the bar used for this data set followed by text that you can edit. Select “Use column titles” to link the legend to the column title entered on the data table. If you edit the legend on the graph, Prism will automatically uncheck the option box. For more details, see “Legends” on page 68.

Transposing rows and columns of bar graphs

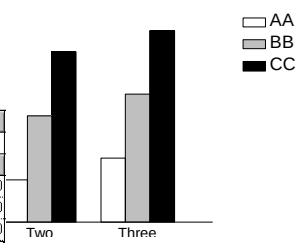
Each bar is part of two groups, defined by its row and column in the data table. Each column has a different color and fill pattern (which can be identified with the legend). Within each column, the values in different rows have the same color and fill pattern and are identified by their order and by labels on the X axis.

If the bar graph is hard to understand, try transposing the data table. After transposing, all values that were on the same row have identical color and fill patterns, and all values that were in the same column now have different appearances. See “Transposing rows and columns” on page 159.

This example shows an interleaved bar graph before and after transposing.

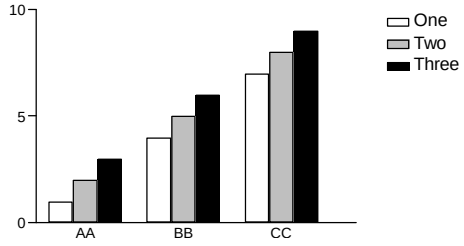
Original Data

	X Values	A	B	C
	X Labels	AA	BB	CC
	X	Y	Y	Y
1	One	1.0	4.0	7.0
2	Two	2.0	5.0	8.0
3	Three	3.0	6.0	9.0



After Transposing

	X Values	A	B	C
	X Labels	One	Two	Three
	X	Y	Y	Y
1	AA	1.000	2.000	3.000
2	BB	4.000	5.000	6.000
3	CC	7.000	8.000	9.000

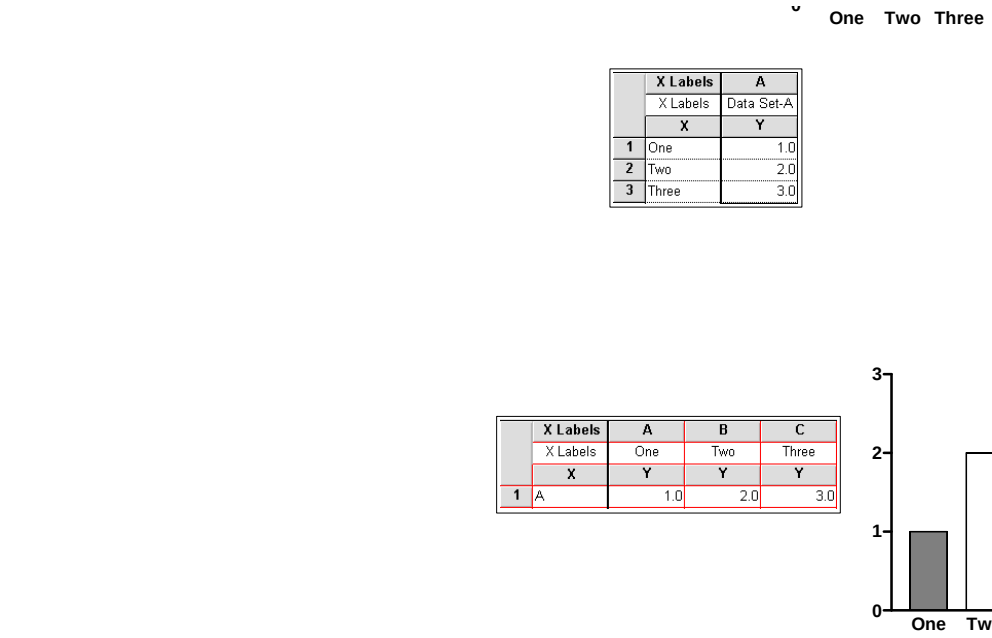


Bar graphs with one data set

If you have a single list of numbers, there are two ways to make a bar graph.

- Enter all the values in a single column. All bars have the same color and fill pattern.
- Enter all the data in one row. Each bar has a different appearance.

Use the transpose analysis to convert a single column to a single row.

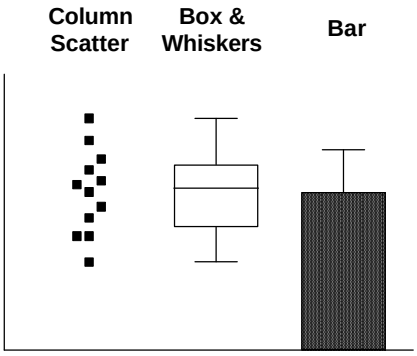


Box & whiskers, column scatter and column bar graphs

What are column graphs?

In a column graph, each data set (column on the data table) becomes one column of points (scatter plots), one set of box and whiskers, or one bar.

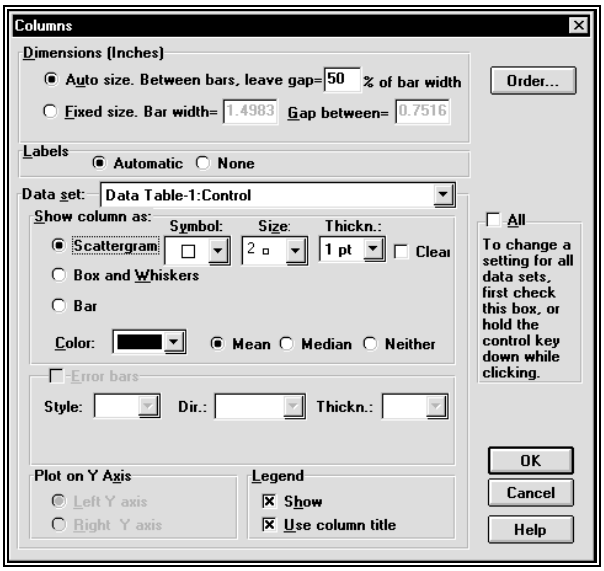
- Column scattergram. Each data point is shown as a symbol. Prism automatically spreads out symbols with similar Y values so they don't overlap. Choose the shape, size and color of the symbols. Optionally graph a horizontal line at the mean or median.
- Box and whiskers.The box extends from the 25th percentile to the 75th percentile, with a horizontal line at the median (50th percentile). Whiskers extend down to the smallest value and up to the largest. Select the color. Note that some programs plot whiskers differently. If you publish the graph, state that the whiskers show the range of the data.
- Bar. The bar extends from Y=0 to the mean of the values. Choose the color and fill pattern. To add an error bar, select "error bar" and select whether it is calculated from the SD, SEM or 95% CI of the column.



Notes:

- Prism automatically creates a column scatter graph when your data table has no X column or the X column is empty. To change it into a box & whiskers or column bar graph, press Change and choose Graph type.
- If you entered replicate values, Prism graphs only the mean values. If you entered SD or SEM values, these are ignored.
- After creating any of the three kinds of column graphs, you can individually change the appearance of each column. This lets you create a combination graph, such as the one shown above.

The columns dialog



Once you have created a graph, use the Columns dialog to change its appearance or to change the order of the data sets on the graph.

To bring up the Columns dialog:

- Double-click on the data area of any box and whiskers plot or column scatter graph.

  or

- Press the Change button and select "Column appearance".

Dimensions

Let Prism choose column widths automatically or enter the widths manually.

If you choose automatic column widths, enter the gap between each pair of columns as a percent of total column width. If you choose manual column widths, enter the width of each column and the width of the gap between columns (in inches in the US and UK, in cm elsewhere). If you make the columns too large, all the data sets may not fit on the graph.

Order

Initially the columns are plotted from left to right corresponding to their location on the data table.

To change the order of data sets on the graph:

1. Click the "Order" button.
2. Select any data set.
3. Click the Up or Down buttons to rearrange.
4. Moving the data set up in the list moves those data further to the left on the graph.
5. Repeat steps 2 and 3 as needed.

Labels

Select the "Automatic" option box to label the bars with data set (column) titles from the data table. If you edit the column titles in the data table, the graph will update. You may not edit the column titles directly on the graph.

Select "None" to leave the labels off the graph. To add your own labels, see "Writing text" on page 107.

Data set

Choose the appearance of one data set at a time. Select a data set from the drop down list, and choose the kind of column (box and whiskers, column scatter graph, or bar) and color for that data set. Then select another data set and choose its appearance. Repeat as often as necessary. Click OK to redraw the graph with all the changes.

You may also make global changes for all data sets, by selecting the "All" check box on top of the dialog. While that box is checked, changes apply to every data set on the graph. For example, you might want to plot all data sets as box and whiskers or make them all the same color.

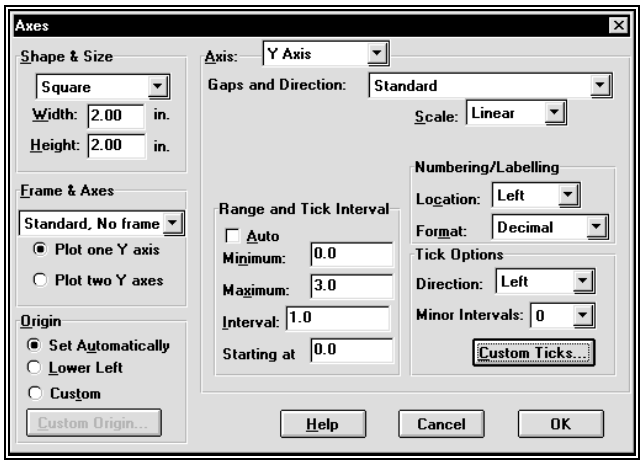
Short Cut: Press the Ctrl key to temporarily make global changes. When you release the Ctrl key, you will revert to adjusting one data set at a time.

Setting	Meaning
Show columns as:	For each column, choose between column scatter, box & whiskers, and column bar. This lets you create a combination column graph.
Plot on which Y axis?	If your graph contains both left and right Y-axes, specify which axis to use for each data set. Initially Prism plots all data on the left Y axis. See "Creating two Y-axes" on page 97.
Legend	Select "Show legend" to create a legend for this data set. The legend will be the symbol used for this data set followed by text that you can edit. Legends are rarely helpful with column graphs, as the column titles appear underneath each column. Prism does not show a legend unless you select "Show legend". For more details, see "Legends" on page 68.
Error bars	For column bar graphs, check the option box to create error bars and then select how they are to be calculated (SD, SEM, 95% CI, or range). The error bars are calculated from all the values in each column. You cannot add error bars to column scatter or box & whiskers graphs.

Graphing axes

The Axes dialog

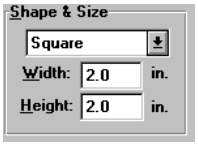
Adjust the size and shape of the graph, as well as the range and appearance of the axes, from the Axes dialog. The rest of this chapter explains how to use the dialog.



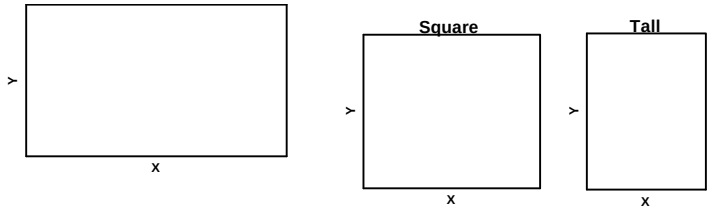
Graph shape and size

Setting shape and size by entering exact dimensions

Set the size and shape of the graph on the upper left corner of the Axes dialog.



We suggest that you choose one of the standard shapes (tall, square or wide) so that your graphs will match each other. If you enter a new width or height, Prism will adjust the other to maintain the aspect ratio (width/height). With wide and tall graphs, the long axis is 1.5 times the length of the short axis. For a nonstandard graph, select a "Custom" shape. Then enter both width and height.



Note: Prism uses the units (inches or centimeters) set in the International section of the Windows control panel.

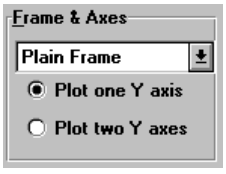
Setting shape and size by stretching an axis

The easiest way to adjust size and shape is to stretch the axes with the mouse.

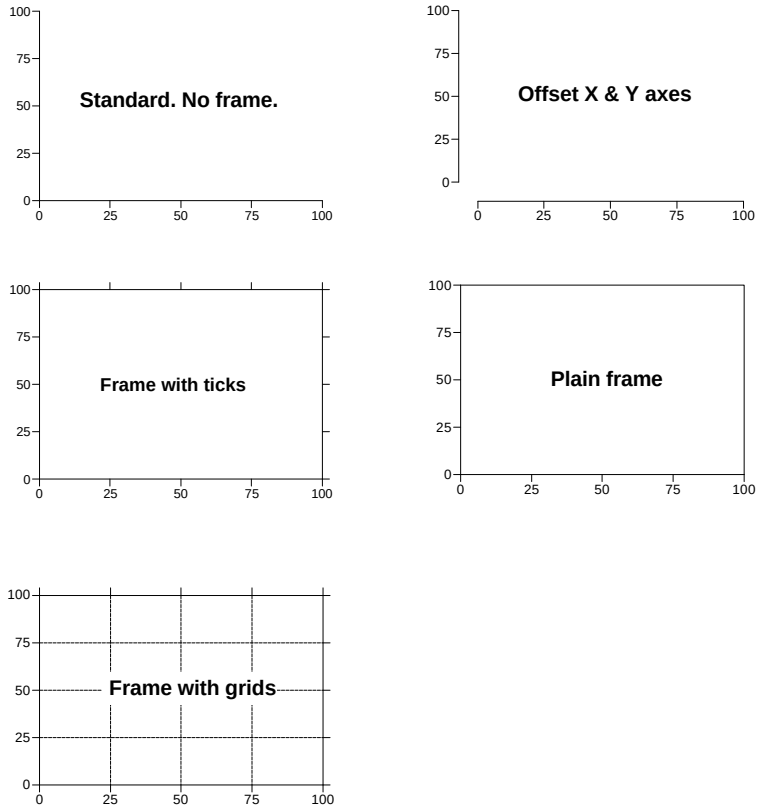
To change size and shape by stretching an axis:

1. Click once on either the X or Y axis to select it. Knobs appear on both ends.
2. Move the mouse over the right knob of the X axis or the top knob of the Y axis. Hold down the left mouse button, and drag the axis to make it longer or shorter.

Axis frame and offset axes



The drop-down menu in the Axes dialog offers five choices:



If you choose a frame with a grid or ticks, these will be placed at each major tick.

If you choose offset axes, you may adjust the distance between the axes.

To adjust offset axes:

1. Click on the X axis to select it. Knobs appear at both ends.
2. Hold the mouse over the left knob and drag to the right or left.
3. Click on the Y axis to select it. Knobs appear over both ends.
4. Hold the mouse over the bottom knob and drag it upwards or downwards.

Origin

The position where the X and Y axes intersect is called the origin. You have three choices, indicated at the lower left corner of the Axes dialog:

- Automatic. The X and Y axes intersect either at X = 0, Y = 0 or at the lower left corner of the graph, depending on the limits of the axes.
- Lower-left.
- Custom. Define the intersection point on a second dialog.

Range, numbering and ticks

Range and tick intervals

Prism determines the range and tick interval automatically and adjusts the range as needed if you edit the data. To override these settings, deselect "Auto" and set the range and interval manually on the Axes dialog.

Warning: Once you deselect Auto , Prism will not adjust the axes even if the range of the data changes considerably.

The axis will start at the minimum value and end at the maximum value you enter. The first major tick and number is at the "starting at" value. Additional major ticks and numbers are placed at the interval you specify.

If the axis has two or three segments (you placed a gap in the axis), you set the range of each segment independently. See "Reverse and discontinuous axes" on page 89.

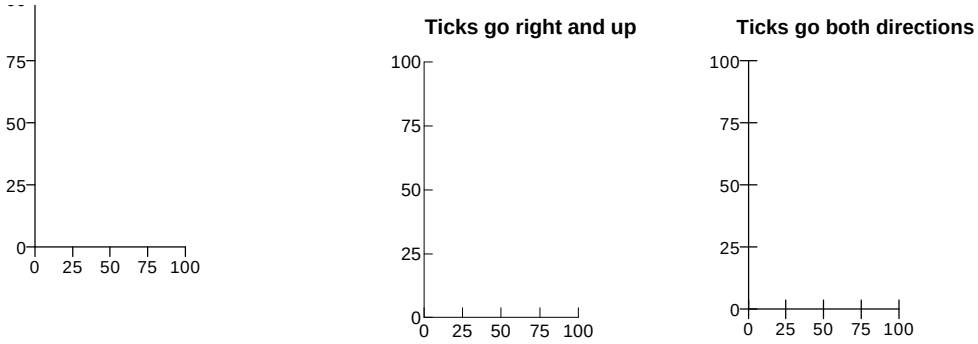
Numbering and labeling options

Use the Location drop-down list box to omit numbering from the axis or to set the numbering location (above or below the X axis, the right or left of the Y axis).

From the Format drop-down list box choose decimal 13 , Scientific $1.3x10^{1}$ or powers of ten 10^{13}.

Tick options

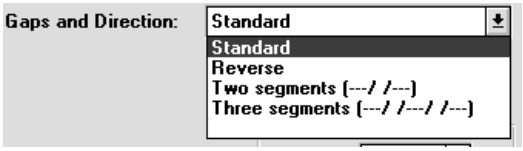
Choose whether the ticks are drawn above, below or across the X axis (to the right, left or across the Y axis). "Both" means that the tick extends across of the axis.



Enter the number of minor intervals (one greater than the number of minor ticks). If you enter "5", Prism will place four minor ticks between each pair of major ticks, to divide the space into five minor intervals.

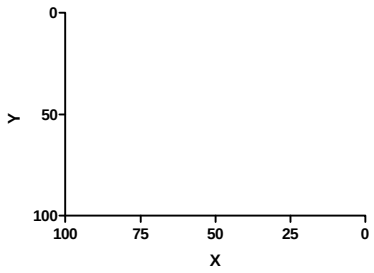
Reverse and discontinuous axes

Choose reverse and discontinuous axes from a drop-down list near the top of the Axes dialog, labeled "Gaps and Direction".



Reverse axes

Select a reverse axis to number the axis in descending rather than ascending order. If you choose a reverse axis, you cannot have a discontinuous axis. It is not possible to choose a reverse axis with bar graphs.



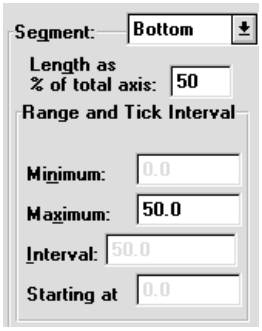
Discontinuous axes

When plotting scientific data, it is often useful to place a discontinuity in the axis.

To create a discontinuous axis:

1. Go to the Axes dialog, and change the "Gap & Direction" selection to "Two segments" (or three segments).

2. Separately set the range of each segment. Use the new pull-down list to choose between the left and right (or top and bottom) segments.



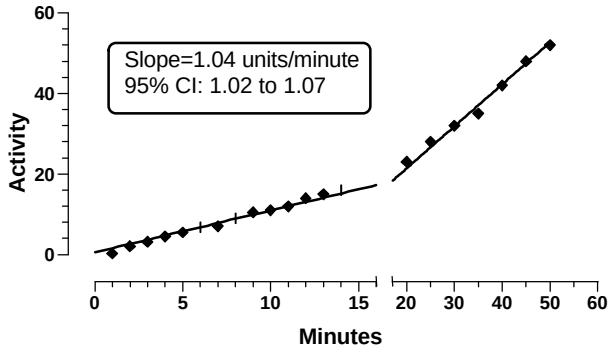
3. You can set the relative lengths of each segment in the Axes dialog as a percent of the total length of the axis.

To adjust the position and size of the gap visually:

1. Select the axis.
2. Drag the knob just to the left of (or below) the gap to change the position of the gap. If you drag to the left, the left segment will become smaller and the right segment will become larger. The size of the gap (in inches or cm) and the range of values shown on each segment won't be affected.

3. Drag the knob to the right of (or above) the gap to make the gap physically wider or narrower. The actual size of the gap between the two segments will grow or shrink, without affecting the relative lengths of the two segments.

Discontinuous axes: Example 1

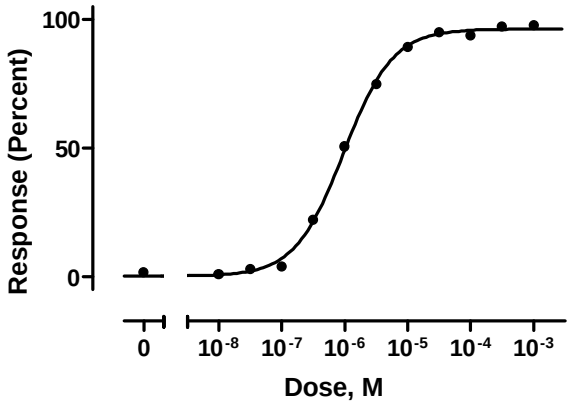


The X values range from 1 to 55. The data were analyzed by linear regression which created a single regression line. The Axes dialog has the following settings:

Setting	Value
Frame and axes	Offset X and Y
Gap and direction	Two segments
Left segment range	0 to 16
Left segment length	58% of total axis
Right segment range	17 to 60
Right segment length	41% of total axis

Because of the change in scale, the linear regression line appears to have two slopes. In fact the mathematical slope is identical everywhere, and the line was created by a single analysis of all the data.

Discontinuous axes: Example 2

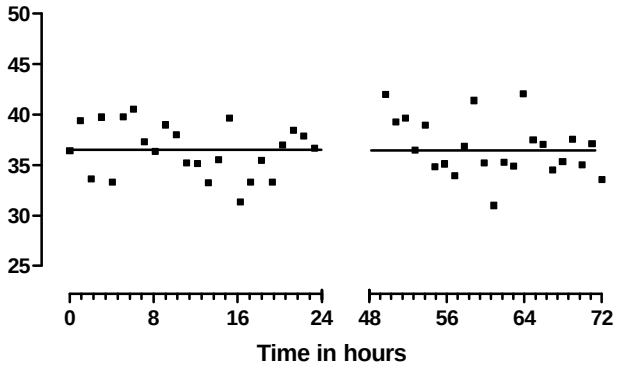


The data points at "0" concentration were actually entered with X=-9. The curves were created by nonlinear regression using a sigmoidal curve (variable slope).

The Axes dialog has the following settings:

Setting	Value
Frame and axes	Offset X & Y axes
Gap and direction	Two segments
Left segment range	-9.5 to -8.5
Left segment length	15% of total axis
Right segment range	-8.5 to -2.5
Right segment length	85% of total axis
Custom tick	"0" at X = -9 with a tick

Discontinuous axes: Example 3



The Axes dialog has the following settings:

Setting	Value
Size and shape	Custom shape. Height 2 inches. Width 4 inches
Frame and axes	Offset X and Y axes
Gap and direction	Two segments
Left segment range	0 to 24
Left segment length	50% of total axis
Right segment range	48 to 72
Right segment length	50% of total axis

Logarithmic axes

Many kinds of scientific data are best graphed with logarithmic axes. There are two reasons to create a log axis:

- Your data are the log of the values you really care about. You want to change the labeling on the axis to show the antilogs.
- Your data span a wide range of values. It is easier to visualize all the data when the axis is stretched into a log axis.

Creating log axes with data entered as logarithms

It is common to enter data as logarithms. For example, pharmacologists commonly enter concentrations as the logarithm of concentration (-8 to -3 in the example below), because it is necessary to fit data to sigmoidal curves.

The example below shows how to change the format of the axis numbering to show the original values. To do this, go to the axis dialog and set the numbering format to powers of ten. In this example, the minor ticks were set to log.

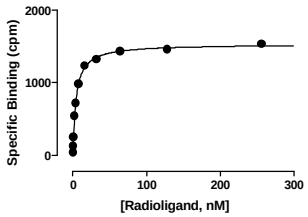
-6.0	18.0
-5.0	55.0
-4.0	92.0
-3.0	99.0

Creating log axes with data entered as actual values (not logs)

The other way to make a log axis is used when your data span a wide range of values and it is easier to visualize the data on a log scale. To do this, set the scale of the axis to Log10 on the top right of the Axis dialog.

This example shows the same data graphed with a linear X axis (left) and a log X axis (right). The curve was obtained by nonlinear regression using the one-site binding equation. On the linear axis, this appears as a hyperbola. On the log axis, it appears as a sigmoid curve. The two graphs show the same curve – the only difference is the X axis has been stretched into a log axis on the right.

	X Values	A
	X Title	Data Set-A
	X	Y
1	1.0	250.0
2	2.0	543.0
3	4.0	721.0
4	8.0	987.0
5	16.0	1234.0
6	32.0	1324.0
7	64.0	1432.0
8	128.0	1456.0

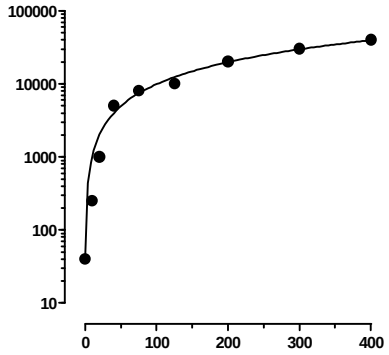


In the graph on the right, the X axis numbering was set to antilog. You had several choices:

Format	Result
Antilog	1, 10, 100, 1000
Powers of ten	10^0, 10^1, 10^2, 10^3.
Log	0, 1, 2, 3

Log axes and regression

When performing linear or nonlinear regression, Prism fits an equation to your data as they are displayed in the data table. Even if you initiate the analysis from a graph, the regression analysis completely ignores how your axis is displayed. Creating a log axis does not change your data. So a linear regression fit to data displayed on log axes will appear curved. These two graphs show the same data. In each case, the data was fit by linear regression. The graph on the right has a logarithmic Y axis, so the linear regression line appears curved.



Numbering of axes

Font and color of numbering

You may change the font, size, and color of the numbering.

To change the font used for all axes:

1. While looking at the graph, press Change.
2. Select "Font of Axis Numbers".
3. Choose font and point size on the Font dialog. Also choose whether you want the numbers to be bold.

To change the font or color used for one axis:

1. Click on an axis to select it.
2. Change the font or color using the text menu.

Notes:

- To change the size of the numbering, click on an axis then press the + or - button in the toolbar.
- To change the default font, see "Graph options" on page 134.
- To change the color of the axis itself, see "Thickness and color of axes" on page 98.

Position

Prism can place the numbers on the X axis either above or below the axis, and can place the numbering of the Y axis either to the left or right of the axis. Select the position in the "location" section on the Axes dialog. See "Range, numbering and ticks" on page 88.

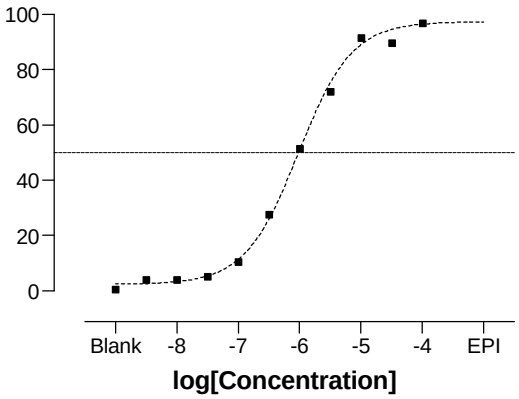
Custom ticks and labels

In addition to the regular numbering along the axis, you may create additional custom ticks and labels.

To create custom ticks and labels:

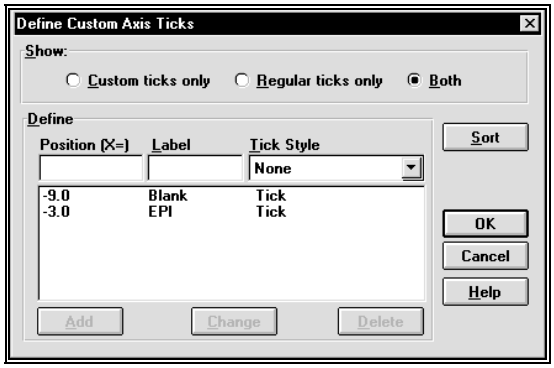
1. From the Axes dialog, click the Custom ticks button.
2. Choose whether Prism should graph custom ticks only, regular ticks only, or both.
3. Enter the coordinate of the first custom tick.
4. Enter the text that goes under (or next to) that tick.
5. Choose the tick style (tick, solid line, dotted line, tick with solid line, tick with dotted line or none).
6. Click the Add button.
7. Repeat steps 3-6 as often as necessary.

If you place the custom tick (or line) at the location of a regular tick and leave the text blank, the number under (or next to) the regular tick will still appear. To erase the regular numbering, enter a space as the text for the custom tick.



The example shows regular ticks as well as three custom ticks:

- The word "Blank" and a tick was placed at X = -9. Note that the custom tick replaces the regular tick ("-9") that would have appeared at that location.
- The word "EPI" and a tick was placed at X = -3.
- A dotted line with no text was placed at Y = 50.

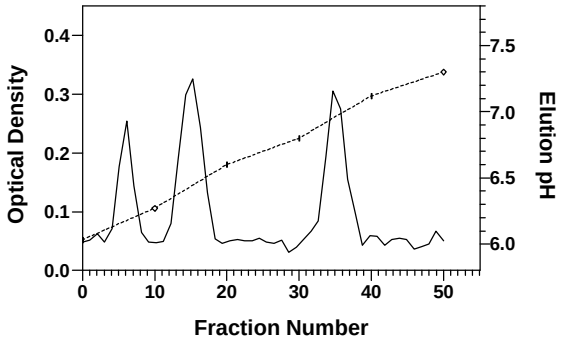


To change a custom tick:

1. From the Define custom axes ticks dialog, highlight the tick you want to change.
2. The coordinates and text appear above the list of ticks. Edit the information.
3. Press Change to edit the existing tick. If you changed the position, press ADD to keep both labels/ticks.

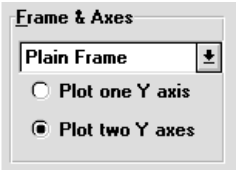
Creating two Y-axes

When you plot several kinds of data on one graph, it is often useful to plot some data using a Y axis on the left side of the graph and other data using a Y axis on the right side of the graph.



To create the second Y axis:

1. Press the Change button, then choose "Axes: Range and Ticks".
2. On the left of the Axes dialog, choose "Plot two Y-axes".



3. On top of the Axes dialog, select "Right Y axis".
4. Select the range and ticks.

Note: Graphs with two Y-axes look better with a frame.

To plot data using the right Y axis:

1. All data sets are initially graphed using the left Y axis. Depending on the scale of the axis, the data points may or may not be visible.
2. Double-click on a data point (symbol).

   or

   Press the Change button, then choose Symbols and Lines. At the top of the Symbols and Lines dialog, select the data set.
3. At the bottom of the Symbols and Lines dialog, select "Right Y axis".

4. Repeat for each data set to be plotted on the right Y axis.

Note: When you fit curves to data, Prism initially plots the curves on the left Y axis, even if the data are plotted on the right Y axis. To plot the curve on the right Y axis, go to the Symbols and Lines dialog, select the data set containing the curve, and choose to plot on the right Y axis.

Thickness and color of axes

To change the thickness or color of all axes (and the graph frame):

1. While looking at the graph, press Change.
2. Select "Color of axes and frame" or "Thickness of axes and frame".
3. Make a choice from the secondary menu.

Prism also lets you change the thickness or color of each axis, or each axis segment, independently.

To change the thickness or color of an axis or axis segment:

1. Select the axis you want to change. If the axis is divided into two or three segments, select one of them by clicking on it.
2. If you want to select additional axes (or segments) hold the Shift key while selecting them.
3. Pull down the Properties menu.
4. Select "Line & Border Color" or "Line & Border Thickness".
5. To change the color of the numbering, choose Color from the Text menu.

Notes:

- To choose the default thickness of axes, see "Graph options" on page 134.
- If your graph includes a frame, the thickness and color of each segment of the frame is determined by the thickness and color of the axis immediately opposite. It is not possible to set the frame color and thickness independently.

Page layouts

Introducing page layouts

If you only want to place a single graph on a page, you can do everything – create, polish, export, and print – from the Graphs section of your project. Use the Layout section to arrange two or more graphs on a page, along with data or results tables, drawings, and imported images.

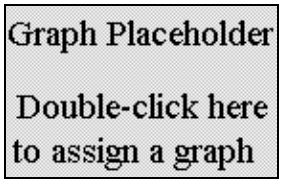
When you first click the Layout tab, you'll see an empty page with one or more empty placeholders.

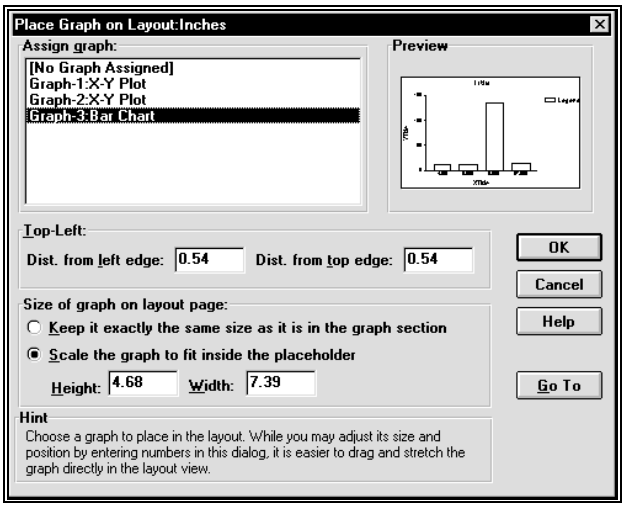
You can change the number and arrangement of place holders on the page, change the page from landscape to portrait, and replace each place holder with a graph or table.

To	Do this
Change the number of place holders on the page or the orientation of the page (landscape vs. portrait).	Press Change, then choose Number and arrangement of graphs.
Erase the place holders to start with an empty page.	Click on a placeholder to select it, then press DEL to delete.
Assign a graph to the place holder.	Double-click on the placeholder.

Assigning graphs to placeholders

Double-click on a placeholder to replace it with a graph.





Choose a graph on the top of the dialog. The preview on the top right of the dialog lets you check that you've chosen the correct graph.

Notes:

- The preview shows you what the graph looked like the last time you viewed it. If you've edited the data, the preview may be out of date. Prism will update the graph when it places the graph on the layout.
- The bottom part of the dialog sets the size and position of the graph. Since you'll usually do that visually, you can usually ignore the bottom part of this dialog.
- Prism usually scales the size of a graph to fit the place holder on the layout. If you have constructed a graph with a particular size that matters to you, place it on the layout at exactly that size, without scaling. To do this, select "Keep it exactly the same size as it is in the Graphs section."
- If you open a file created by Prism version 1 and go straight to the layout section, you won't see graph previews. Prism will create the previews automatically when you look at the graphs in the Graphs section.

Changing the number or arrangement of graphs

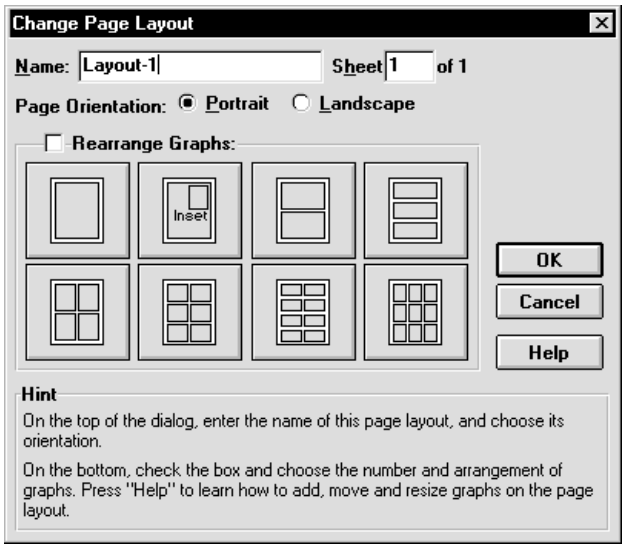
Use the Page Layout dialog to choose the orientation of the page, and the initial number and arrangement of graphs.

To access the Change Page Layout dialog:

- Press the Change button and choose "Change number and arrangement of graphs."

or

- Double-click on the background of the page layout (outside of all graphs and placeholders).



Page orientation

Choose between landscape or portrait orientation for this page layout. You may choose a different orientation for each layout in the project.

Unlike most programs, you do not set the orientation of graphs or page layouts in the Printer Setup dialog. By placing the control in the Change page layout dialog, Prism lets you choose a different orientation for each page.

Number and arrangement of graphs

If you are creating a new layout, press one of the buttons to start with 2, 3, 4, 6, 8 or 9 graphs. Later you may add, delete, rearrange, and resize graphs.

If you are editing an existing page layout, leave the "rearrange graphs" box unchecked to leave the arrangement of graphs alone (perhaps you are only changing the sheet name or orientation). If you press one of the other buttons, Prism will resize and rearrange the graphs to fit the layout on the button (any surplus graphs will be removed from the layout).

Adding and removing graphs manually

You don't have to keep the regular arrangement of page layouts provided by Prism. You can add, delete and move placeholders and graphs.

To add a placeholder to the page:

1. Click on the graph tool (just to the right of the arc tool). 
2. Point to the location where you want a corner of the graph to appear.
3. Drag to the opposite corner.
4. Double-click on the new placeholder to assign a graph.

To duplicate a graph or placeholder:

1. Select a graph or placeholder.
2. Pull down the Arrange menu and select Duplicate. The duplicate graph will overlap the original graph.
3. Move the new graph to a new position.
4. Double-click to assign a different graph to that position.

To make an inset:

1. Create both graphs in the Graphs section.
2. Place both the main graph and the inset on the layout page.
3. Reduce the size of the inset, and drag it to position.

To delete a graph or place holder:

1. Click once to select it.
2. Press DEL.

Adjusting the page layout

Borders and backgrounds

To change the background color for the whole page:

1. Pull down the Sheet menu.
2. Select Background color, and choose a color.

To change the background color for one graph only:

1. Select the rectangle (or rounded rectangle) tool and draw a box around the graph.
2. Pull down the Properties menu and select a fill pattern and fill color (and also the line/border color and thickness).

3. The rectangle will now block the graph, so the graph is invisible.
4. Pull down the Arrange menu and select Send to back.

To place a box (frame) around the graph:

1. Select a graph.
2. Pull down the Properties menu and select a line/border thickness and color.

Aligning graphs

To align the axes of two or more graphs:

1. Hold Shift, and click on each graph to select them.
2. Click the Change button and select Align, or select from the Arrange menu.
3. From the secondary menu, choose to align the X-axes or Y-axes.

Balance graph scaling

When copying a graph to the layout, Prism scales the graph object so it fits inside the placeholder. The scale factor depends on the relative sizes and shapes of the placeholder and graph object. Because the graph object can includes titles, legends, drawing, tables, and pictures, graphs that are the same size in the graphs section can scale differently so they are different sizes in the layout.

Prism can balance the graph scaling. It does this by equalizing the percent reduction (or increase) of the length of the axes. After scaling, two graphs that were made the same size in the Graphs section will be the same size in the layout. Since the graph object includes more than just the graph, the graph objects may not be equal size after balancing. Prism balances the length of the axes, not the size of the graph objects.

To balance graph scaling:

1. If you only want to balance selected graphs, hold Shift and click on each of them. (Skip this step if you want to balance all graphs on the layout).
2. Press the Change button and choose Balance Graph Scaling. (Or choose from the Arrange menu).
3. Choose whether to reduce the size of graphs that are too small, or to increase the size of graphs that are too large.

Note: The term scaling refers to the change in size when the graph is put on a page layout. It has nothing to do with the scale (range) of the axes.

Center objects on page

Before you print, center all objects on the page.

To center the page:

1. Pull down the Arrange menu.
2. Select Center page.

Adding text, drawings, pictures and tables

Writing text

Overview

You may write text on a graph, page layout or page of notes.

- Write on the graph to annotate that particular graph. See below for writing labels. Also see "Legends" on page 68.
- Write on the page layout to add text that applies to several graphs. You can even erase all graphs (or place holders) from the layout, and create a page layout of only text and drawings. This is useful for creating slides or transparencies.
- Write on a notes page to document where the data are stored, how the experiment was performed, any experimental problems, and your conclusions. See "Entering notes" on page 127.

Entering text

To enter new text onto a graph or layout:

1. Select the text tool by clicking on the button. 
2. Click on the graph or page to indicate where the text is to begin.
3. Start typing. If you want to enter several lines, press RETURN between lines.
4. When you are done entering text, click on the pointer button. 

   or

   Click elsewhere on the sheet.

To enter a paragraph with automatic word wrapping:

1. Enter the first line or two using the instructions in the previous paragraph.
2. Click on the pointer button  to select the entire text object. You'll see knobs on the corners.
3. Drag the knobs to make the text object the correct width. Don't worry about the length, as it will grow automatically as you enter text.
4. Double-click inside the text object to bring back the insertion point.
5. Type your text. Don't push Return at the end of the line. Prism will automatically wrap the text as you type.

To enter a Greek letter:

1. While entering text, press the Greek button. 
2. Select a character from the menu provided.

To create bold, italics, underlining, subscripts, and superscripts:

1. Select the letters that you want to be bold, italics, etc. (Drag the mouse over those letters.)

   or

   Place the insertion point before you start typing.
2. Click on the bold, italics, underline, subscript or superscript buttons.

   If you previously had selected text, it will be bold, italicized, etc.

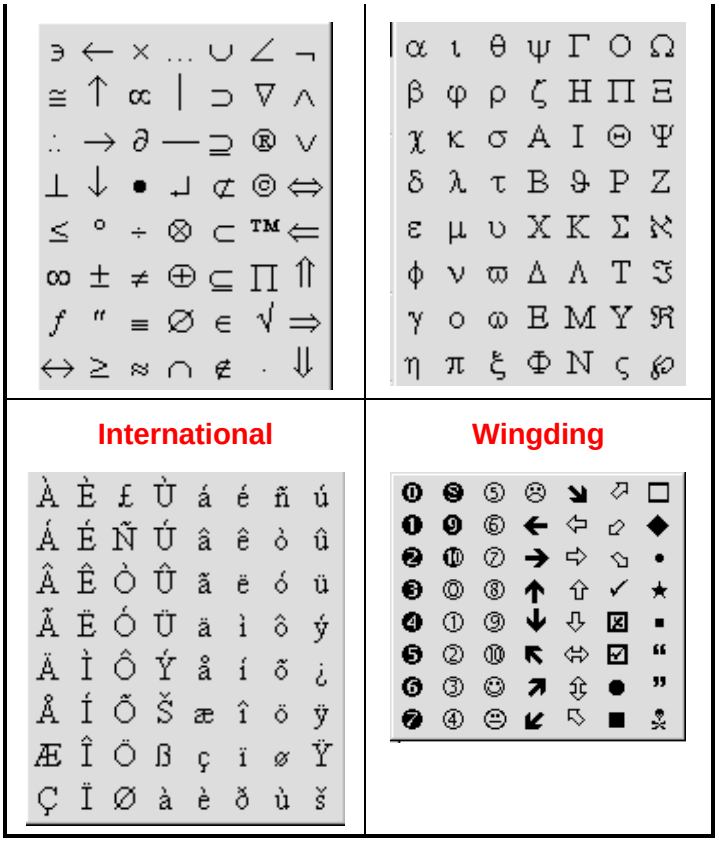
   If you didn't select text, the text you type from now on will be bold, italicized, etc. Press the button again to return to normal text.

Shortcut: Hold Ctrl and press + or - for superscript or subscript. Use the gray + and - keys next to the keypad, not the keys on top of the keyboard. Hold Ctrl and press U, I or B for underline, italics or bold.

To enter a international character (i.e. ä or é) or math symbol (i.e. ← or ±) or Wingding (i.e. or):

1. While entering text, pull down the text menu.
2. From the menu, select Insert International, Insert Math, or Insert Wingding.
3. Select a character from the menu provided.

Math	Greek



Editing text

To edit text:

1. Double-click to place the insertion point in the text.
2. Add, delete or edit.

To change the point size or font:

1. Select the characters you want to change.

   or

   Select the entire text object.

2. Press the + or - buttons to increase or decrease the point size.

   or

   Pull down the Text menu and select Font.

Note: Stretching the text object to change its size will not change the point size of the lettering.

Working with blocks of text

When you enter text, the entire block is considered to be one object. You can place a border around it.

To select the text object:

1. Point to an edge of the block.
2. Click the left mouse button.

Note the difference between selecting characters (indicated as inverted text), and selecting the entire block of text (indicated by six or eight knobs).

To place a border around a block of text:

1. Select the text object.
2. Pull down the Text menu .
3. Select a border thickness.
4. To adjust the border color, select border color from the Properties menu.
5. If you want to justify the text within the box, pull down the Text menu and choose Justify.

To give the background a color:

1. Select the text object.
2. Pull down the Text menu.
3. Select a background color.
4. To change the text color, select Color from the Text menu.

To move a block of text:

1. Select the text object.
2. Point to any area inside the text block.
3. Press and hold the left mouse button. The mouse cursor changes its appearance. Drag the text to its new location.
4. Point and click elsewhere to deselect the text you moved.

Note that graph titles are not text objects. They are always centered on an axis and can only be moved closer to or further away from the axis. See "Titles" on page 69.

To change the size and shape of the text object:

1. Select the text object.
2. Hold the mouse over one of the knobs. Hold down the left mouse button and drag.
3. Prism will wrap the text to match the new shape of the object.

Note that changing the size and shape of the text object does not change the point size of the lettering. To change the lettering size, you need to change the point size of the font.

To justify a block of text:

1. If necessary, stretch the object to an appropriate shape and size.
2. Select the text object.
3. Pull down the Text menu and choose Justify.
4. Choose left, right, center or full justification. Prism justifies with respect to the object boundaries.

Shortcut: Bring up the text menu by clicking the right mouse button.

Drawing lines, arrows, ovals and boxes

You may add drawings to graphs and pages.

To draw a line, arrow, rectangle, oval or arc:

1. Click on a drawing tool. 
2. Note that the mouse cursor changes to a pencil. 
3. Position the mouse pointer at one end of the line or arrow, or at one corner of the box or oval.
4. Hold down the left button and drag to the other end, or to the diagonally opposite corner.

Notes:

- To draw a perfect square or circle (rather than rectangle or oval), hold the Ctrl key while dragging.

- After you complete the drawing, the mouse cursor reverts to the selection arrow. If you want to use the same tool repeatedly, hold the Shift key while selecting the tool in step 1 above. After drawing all the objects you want, click the selection arrow tool, just to the left of the drawing tools.
- When drawing an arrow, start at the tail. After drawing the arrow, adjust the arrowhead size and direction from the Properties menu.
- Arcs are always one quarter of an ellipse, starting vertically from the first mouse click and ending horizontally at the second.

To	Do this
Cancel drawing.	Click on the pointer button to the left of the line tool. 
Select a drawing.	Point to its edge. Click the left mouse button.
Move a drawing.	Select it. Point to an edge. Hold down the left mouse button and drag.
Stretch a drawing.	Select it. Point to a handle. Hold down the left mouse button and drag.
Delete a drawing.	Select it. Press DEL.
Change a drawing's color or appearance.	Select it. Pull down the Properties menu. Choose Line/Border color, Line/Border pattern, Fill color or Fill pattern.

Importing images

You can place images created in other programs onto graphs or page layouts. One approach is to paste from the Windows clipboard. The other approach is to import a file. Prism does not (yet) support object linking and embedding (OLE).

Also see "Importing data" on page 43.

To paste a picture created in another program:

1. Select a graph or picture in any Windows program and copy it to the clipboard.
2. Switch to Prism. Go to the appropriate graph or page layout.
3. Point the mouse to the place where you want the upper left corner to go.
4. Click the right mouse button, and select Paste from the shortcut menu.

To import a picture created in another program:

1. Save the picture in .TIF, .PCX, .BMP, .GIF or .WMF format. If you have a choice, choose WMF (Windows metafile).
2. Go to the appropriate graph or page layout in Prism.
3. Pull down the File menu. Choose Import.
4. Select the file type and the file.
5. Prism places the object at the top of the page. Grab a handle or edge to move or resize it.

When you transfer pictures from one program to another, you should know the difference between vector and bit map formats. Vector formats store the instructions required to create the picture. No information is lost. Bit map formats store the picture as a number of individual pixels. The resolution is defined by the file, and the image may look grainy when printed.

Prism imports Windows Metafiles (WMF), which is a vector format, and .BMP, .PCX, GIF and .TIF formats, which are bit maps. Avoid bit map files when possible, because the files tend to be huge and the picture quality tends to be poor. If you have to use a bit map file, avoid TIF, as the other bit map formats tend to work better.

Embedding data and results tables

You can place portions of a data or results table on a graph or layout. Like graphs and analyses, these tables remain linked to the original data. If you edit or replace the original data, Prism will update the tables.

Step 1. Select and copy a portion of a table.

1. Go to a data table or to a results table.
2. Select a portion of the table. If you want to paste a single value just click on that cell – you don't have to do anything special to select it.
3. Copy the selection to the clipboard. To do this, press the Copy button, pull down the Edit menu and choose Copy, press Ctrl-C, or click the right mouse button and choose Copy.

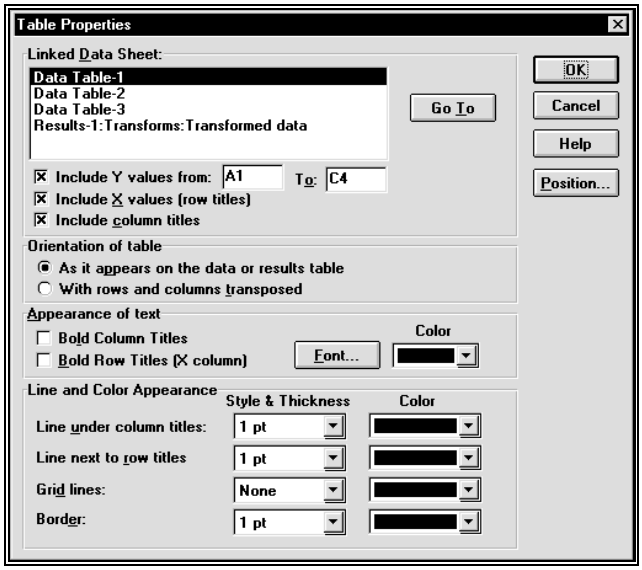
Step 2. Paste the table

1. Go to a graph or layout.
2. Point with the mouse to the location where you want the upper left corner of the table to go.
3. Click the right mouse button.
4. Choose Paste from the shortcut menu.
5. Move and resize the table.

You can also press the Embed Table button. The resulting dialog lets you choose whether to transpose the table when you paste. However, you have no control over the initial position of the table.

Step 3. Fine-tune the appearance of the table

1. Double-click on the table to bring up the Table properties dialog.
2. Change the dialog as necessary.



Linked data or results table

You first designate the portion of the table you want to include when you copy it to the clipboard. If you change your mind, you can change the selection on top of the dialog.

- Which table? Choose a data or results table.
- Include Y values? Enter the upper left and lower right coordinate on the table. Examples:

From	To	Comment
A1	C4	From the first row of the first column (A) to the fourth row of the column C.
A1	E:	All of columns A through E. Since there is no final row number, the entire columns are included.

A1	ZZ:	The entire table.
E3	E3	Only the value in the third row of column E.

- Include X values? For results tables, the X values may more accurately be called row titles. If you include X values, but not Y, the range of rows comes from the range of Y values you enter. If you chose a range from A1 to C5, but checked only X values, you'd get the X values from rows 1 to 5.
- Include the column titles?

Orientation of the table

Embed the table in the same orientation as it appears in the data table or transpose rows and columns. If you transpose, the first row you selected becomes the first column in the embedded table and so on. Transposing only works when the table is formatted for single Y values. Transposing will not work if your table is formatted for duplicate or triplicate Y values, or for Y with SD or SEM.

Results table with selection

	X Labels	A	B
		Control	Treated
	X	Y	Y
1	Equation 1		
2	Variables		
3	SPAN	1369	1311
4	K	0.07701	0.1273
5	PLATEAU	434.8	250.9
6	HalfLife	9.000	5.441

As it appears in the results table

	Control	Treated
K	0.07701	0.1273

Rows and columns transposed

	K
Control	0.07701
Treated	0.1273

Appearance of text

Check whether you want the column titles or X column to be bold. Beyond that, all numbers and text in the table have the same appearance. You can't adjust the font of some cells separately from the rest.

Color and line style

Setting	Meaning
Line under column title.	If you transpose the table, then this is really the line next to the first column.

Line next to X column.	If you transpose the table, then this is really the line under the first row.
Grid lines.	Lines, if any, between Y columns and rows.
Border.	Lines, if any, that surround the entire table.

All options turned on

A273
50.9
.441

All options turned off

	Control	Treated
Variables		
SPAN	1369	1311
K	0.07701	0.1273
PLATEAU	434.8	250.9
HalfLife	9.000	5.441

Notes:

- The size of the table is determined by the number of rows and columns and by the font. Stretching the table will alter the point size. If you change the point size, its size will change.
- The "table" can be a single value. The label below is actually two objects side-by-side (a text object next to a single-cell table).

- You cannot edit the numerical format inside the table. However, you can adjust the number format in the data table and results table. For data tables (and results tables like Transforms), select one or more columns and choose Number format from the Edit menu. For other results, choose the number format on the Parameters dialog. From the results sheet, press Change and then choose Analysis parameters.
- The Tools menu includes the Embed table command. This command is provided so that people who don't read manuals or help screens will figure out how to embed tables. There is no advantage to using that command to copy. There is one minor advantage to using that command (or button) to paste: you can choose to transpose the table when it is first created. If you use the Paste command, the table is always pasted just as it looks on the data or results table. Transposing requires an extra step.

- Embedding a table works only within Prism. Prism does not yet support Object Linking and Embedding, which is used to enable communication between programs.
- You can only embed a table created within Prism. If you want to embed a spreadsheet table, you'll need to copy it to a Prism data table first.
- You can create a table containing only text. To do this, create a new data table with both X and Y columns formatted to contain text. Then copy that table to the clipboard, and paste to embed it on a graph or layout.

Selecting objects

What is an object?

When you work in the Graphs section of your project, the graph itself as well as text, legends, drawings, and imported images are all referred to as "objects". When you are working in the Layout section of your project, each graph (with all its objects) becomes a single object. Images imported onto a page layout are also objects. Images imported onto a graph are objects in the graph section, but are part of the graph object in the page layout section.

Selecting objects

To position one or more objects, you first must select them:

To select one object:

- Point to a border and click the left mouse button. Prism displays knobs on the selected object. You must click on the edge of the object, not inside it.

To select several objects:

- Hold down the Shift key while selecting the objects one at a time.

  or

- Move the mouse above and to the left of all the objects. Hold down the left mouse button and drag to a spot below and to the right of all objects. Let go of the mouse button. All objects entirely contained within the imaginary rectangle are selected.

Moving objects

To	Do this

Move an object with the mouse.	Point to an edge. Press the left mouse button and drag.
Move an object vertically only.	Hold the Shift key while moving.
Move an object horizontally only.	Hold the Ctrl key while moving.
Position an object by entering the distance of the upper left corner from the top and left edge of the page.	Select Position object from the Arrange menu.
Nudge an object a tiny distance.	Press the arrow keys. Each press of an arrow key moves the object one pixel.
Lock an object so you can't accidentally move it.	Select Lock from the Arrange menu.
Align two or more objects.	Select Align objects from the Arrange menu.
Group two or more objects so they move together.	Select Group from the Arrange menu.
Force objects to snap to grid. This makes it easier to align objects, but harder to fine-tune the positions.	Select Snap to grid from the Arrange menu.

Overlapping objects

When you move an object so it overlaps with another, you only see and select the one in front. Normally, new objects are placed in front of existing objects.

To bring an object to the front or send it to the back:

1. Select one or more objects.
2. Pull down the Arrange menu.
3. Select "Bring to Front" or "Send to Back".

To draw a colored box behind graphs and other objects:

1. Draw a box (or rounded box) around a graph or a group of objects.
2. You will see the objects inside the rectangle. The rectangle is selected (has knobs on its corners).
3. Pull down the Properties menu, and choose a fill pattern and/or color. Also choose a border thickness and color.
4. The filled rectangle lies on top of the other objects, which are now invisible. The rectangle is still selected.
5. Pull down the Arrange menu. Choose "Send to Back".

6. The objects appear on top of the filled rectangle.

Aligning objects

Zoom (enlarge) the view

For precise alignment, you'll want to display the objects larger.

To change the zoom, do one of the following:

- Press the Zoom Large or Zoom Small buttons. 
- Pull down the View menu and choose one of the five Zoom levels.
- Hold down both the Ctrl and Shift keys and click the left mouse button to toggle between small and full size.
- Hold Ctrl and press M to zoom larger or Y to zoom smaller. To remember these keys, note that M and Y are the first and last letters in "magnify".

Prism initially displays page layouts at the smallest zoom, and displays graphs at 100% or 75% size, depending on your choice in the Graph Options dialog. See Graph options on page 134. Since Prism doesn t know how big your monitor is, 100% actual size may not be exactly *the size of the printed page, but it should be close. On some monitors, 75% zoom is closer to actual size than is 100% zoom .*

View the ruler, grid, and coordinates

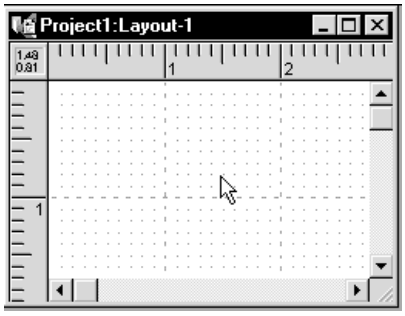
From the View menu, you can show or hide the rulers, grid and coordinates.

- **Show rulers**. The rulers appear on top and to the left of the graph and page layout screens. Show the rulers to align objects precisely. Hide the rulers to leave more space for the graph. Prism never shows rulers at small or 50% zoom, because they would be too small to be useful.
- **Show grid.** Superimpose a grid on the page layout or graph. The grid is not included in printouts and is shown to help you position objects.
- **Show coordinates.** This works differently on graphs and page layouts.

  On both graphs and page layouts, the position of the mouse is shown in the corner where the two rulers meet. The position is shown as inches or centimeters (depending on your Windows control panel) from the left and top edge of the page. These coordinates are visible only when rulers are showing.

When you show coordinates of a graph, Prism also shows the X and Y coordinates (in data units) at the top of the screen, next to the help button. You do not have to view rulers to see these XY coordinates.

This window shows rulers, coordinates and grid.

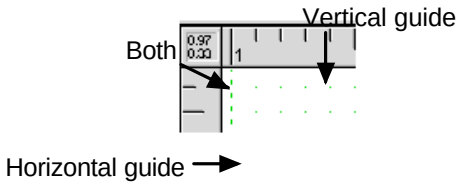


Alignment guides

Alignment guides are temporary lines you can drop from the ruler to check on the position and alignment of objects.

To use alignment guides:

1. If the rulers are not in view, pull down the View menu and select Show Rulers. (This works only if the zoom magnification is 75% or greater.)
2. Click the mouse in either ruler, or in the upper left corner where the two rulers intersect.



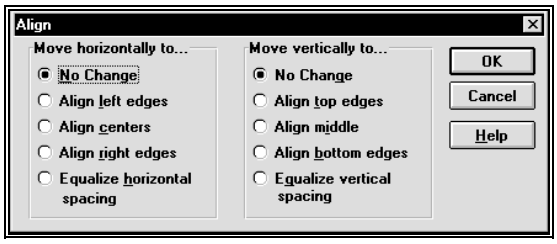
3. Hold down the mouse button.
4. Thin lines drop down from the top ruler, across from the left ruler, or both. Drag the mouse around the screen to move the guides.

5. The alignment guides go away when you let go of the mouse button.

Align objects dialog

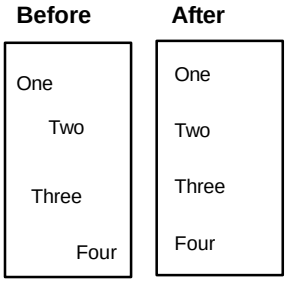
To align two or more objects:

1. Select the objects. See "Selecting objects" on page 117.
2. Pull down the Arrange menu and choose Align. Then choose Align objects.
3. Complete the dialog.



On the left side of the dialog, specify how the objects should be moved horizontally – to align the left sides, centers or right sides. Or check Space equally to equalize the horizontal distances between objects (this choice makes sense only if you have selected three or more objects). On the right side of the dialog, specify how the objects should be moved vertically – to align the tops, middles or bottoms of the objects. Or select Space equally to equalize the vertical distance between objects (again, this only makes sense if you have selected three or more objects).

This example shows the effect of moving horizontally to align the left edges and moving vertically to equalize spacing.



Properties — color, fill patterns and line thickness

Use the Properties menu to change the appearance of the selected object(s) (see "Selecting objects" on page 117). The menu lets you change the following properties:

Shortcut. Click the right mouse button and select Properties from the shortcut menu.

Property	Explanation
Fill color.	The color of the interior of boxes and circles.
Fill pattern.	The pattern of the interior of boxes and circles.
Line & Border thickness.	The thickness of lines used to create drawings and axes.
Line & Border color.	The color of lines used to create drawings and axes. Also the color of axes. When working with page layouts, this creates the border (if any) around graphs and imported images.
Line/thickness pattern.	The appearance (solid, dotted, dashed) of lines used to create drawings and axes. When working with page layouts, this creates the border (if any) around graphs and imported images.
Arrowhead direction.	Add or remove arrowheads from either end of any line or arrow.
Arrowhead size.	Choose the size of the arrowheads.

Copying, pasting and duplicating objects

You can use the clipboard to copy and paste objects from one graph or layout to another.

To	Do this
Select one object.	Click on its border.
Select several objects by clicking.	Hold the shift key, and click on each object's border in turn.
Select several objects by region.	Point to an area above and to the left of all objects. Hold the left mouse button and drag to an area to the right and below all the objects. This selects all objects entirely contained within the region.
Copy.	Select Copy from the Edit menu. Or press Ctrl C. Or click the right mouse, and select Copy from the shortcut menu.
Paste.	Point the mouse to the place where you want to paste. Click the right mouse button, and select

	Paste from the shortcut menu. Alternatively, select Paste from the Edit menu, press Ctrl V, or click the paste button. With these methods, the object is pasted in the upper left of the screen.
Duplicate.	Select Duplicate from the Arrange menu.

Managing your projects

Going to a linked sheet

If your project contains many sheets, Prism provides a shortcut to move between related (linked) sheets. For example, you can go from a data sheet to a graph that contains those data. Or you can go from a graph to results plotted on that graph, or to a results table embedded on that graph.

To go to a linked sheet:

1. Press the "Go To" button at the right of the tool bar. 

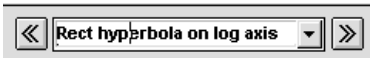
2. Choose a sheet from the menu.

   The Go To menu precedes each sheet name with [D], [R], [G], or [L] to designate data tables, results, graphs and layouts.

Renaming and reordering sheets

Renaming a sheet

Every sheet has a name. Prism initially assigns generic names such as "Table 1", "Results 3", or "Graph 2". You'll find it easier to manage large projects if you give the sheets descriptive names. To rename a sheet, click in the sheet name in the toolbar and edit.



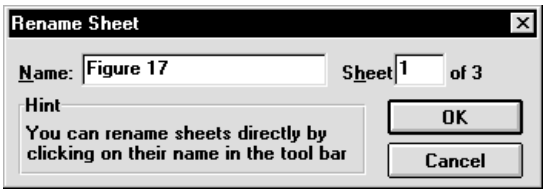
Don t confuse the name of each sheet with the file name for the entire project.

Reordering sheets

The sheets ordinarily appear in the drop-down menu in the order they were created. You may find it easier to manage a large project if you reorder the sheets.

To move a sheet to a different order in the section:

1. Make sure that the sheet you want is visible.
2. Press the Change button, and select Sheet Name.
3. Change the name on the dialog.
4. Enter a new value for position (i.e. sheet 1 of 3). The sheet that used to be at that position and all later sheets will be renumbered and moved down one position.



Deleting a sheet

As you work with a large project, you may accumulate unneeded graphs and analyses.

To delete a sheet:

1. Make sure that the sheet you want to delete is visible.
2. Pull down the Sheet menu and select Delete sheet.

Note: If you frequently find yourself deleting automatically created graphs of analyses, change the Analysis options to create fewer automatic graphs. See *Analysis options on page 135*.

Because sheets are linked, note the following.

- When you delete a data table, you also remove those data from all graphs, but do not delete the graph sheets. You'll probably want to separately delete the graph sheets as well.
- When you delete a data table, all results based on those data become orphaned. You can still view and graph orphaned results, but you can't change the calculation parameters.

- When you delete a results sheet, you also remove those results from all graphs. If you want to see a graph of results, you must not delete the results sheet.
- When you delete a graph sheet, you also delete that graph from all page layouts but do not delete the page layout sheets themselves.

Entering notes

Prism provides a notes editor so you can explain how the experiment was done and note any irregularities in the data. You may also want to document where the raw data are stored and how you interpreted the results.

Click on the Notes tab to enter notes. You may create up to 100 notes sheets, and all are stored in the project file along with data, results and graphs. Notes are not linked to a particular data table or graph.

To	Do this
Change the font, size and color of text.	Use the text tools or the Text menu.
Enter international or Greek characters.	Select Insert international or Insert Greek from the Text menu.
Change the margins or line spacing.	Select Notes format from the Text menu.
Export as a text file.	Select Export from the File menu.
Import a text file.	Select Import from the File menu.

Note: Use the notes section to document your data and conclusions. If you want to include text on graphs or page layouts, write the text directly in the graph or page layout section. See Writing text on page 107.

Merging and splitting projects

As you work with projects, you may want to merge two projects or divide one project into two.

Merging two projects

While you are working on one project, you may append a second project.

To merge a project:

1. Pull down the File menu and select Merge.
2. Choose a file.

3. All sheets from that file will be appended to your current project. An asterisk (*) is placed in front of the name of each appended sheet.

Prism cannot merge if the combined size of any section will exceed 100 sheets.

Splitting a project

After a project file grows, you may want to split it into two files. There are no special commands to split a project into two files, so you need to go through several steps.

To split a project into two files:

1. Save the project file.
2. Use the Sheet Delete command to delete the sheets you don't want in the first of the new files.
3. Pull down the File menu and choose Save As. Enter a name for the first new file.
4. Open the original file again.
5. Use the Sheet Delete command to delete the sheets you don't want in the second new file.
6. Pull down the File menu and choose Save As. Enter a name for the second new file.
7. Delete the original file if you no longer want it.

Moving a graph from one project to another

To copy a graph from one project to a layout in another:

1. Look at the graph and make sure nothing is selected.
2. Press the Copy button (or pull down Edit menu and choose Copy).
3. If the other project is already open, use the Window menu to go there. If it is not open, use the File menu to open it.
4. Go to a layout sheet.
5. Point the mouse to the location where you want the graph to go. Click the right mouse button and choose Paste from the shortcut menu.
6. The graph you paste onto the layout will NOT be linked to any data. It is just a picture that cannot be edited. It will not be updated if you edit the original data in the other file. It is not possible to change the symbol size or color. It is not possible to change the axis range.

Opening several projects at once

You may open up to four projects at one time. Each will appear in its own window.

To	Do this
Start a new project.	Select New Project from the File menu.
Open a project file.	Select Open from the File menu. Or click the Open button.
Work on another project visible on screen.	Click on its window.
Work on a project whose window is hidden behind another window.	Pull down the Windows menu. Select the Window you want from the list at the bottom of the menu.
Close a project (window).	Select Close from the File menu. If you haven't already saved the project, Prism will prompt you to do so.
View all open projects at once.	Pull down the Windows menu. Select Tile Windows or Cascade Windows.
Copy data from one project to another.	Select a range of data in one window, and select Copy. Then move to the other window and select Paste.
Copy a graph from one project to another.	See "Moving a graph from one project to another" on page 128.

Analyzing repeated experiments

Prism makes it very easy to analyze and graph data from repeated experiments. Because a Prism project file maintains the links between data, analyses and graphs, you can analyze and graph a repeat experiment very quickly and easily.

There are four approaches to analyze repeated experiments.

Approach 1: Replace the old data with the new

If you no longer want your original data, and no longer want analyses or graphs based on those data, then you can use this easy approach. Simply replace the original data with the new data. Click on the Graphs tab to see the new graph.

The problem with this approach is that the old data are erased.

Approach 2: Add the repeated experiment to the same project file

Since a project may contain many data tables, results, and graphs, you may include several experiments in one file.

To add a repeat experiment to the same project file:

1. Go to the data table containing the old data.
2. Click on the New table button. On the New Table dialog, choose to duplicate the table.
3. Delete the old Y values, and enter the new. Edit the X values, if necessary. Name this data table.
4. Go to the first results sheet from the old data.
5. Click on the New analysis button, and choose to duplicate the results you are looking at.
6. On the Analyze data dialog, select the new data table.
7. Repeat steps 4-6 if needed for other analyses.
8. Go to the graph of the old data. Click on the New graph button. Choose to duplicate the graph.
9. Press the Change button, and choose "Data on graph". Remove all data sets from the old table and add data sets from the new data table. Also remove any analysis tables of the previous experiment, and add analysis tables of the new experiment.

Approach 3: Using a template file for each experiment

If you frequently perform the same kind of experiment, create a template file. Once the file is created, you can analyze and graph each experiment very quickly. For an explanation of the subtle difference between ordinary and template files, see "Template files" on page 142.

To create a template file:

1. Enter data for a representative experiment, and go through all the steps to analyze and graph the data. Add notes to explain what the template does.
2. If those particular data are worth saving, save the project.
3. Erase the portions of the data that will change with each experiment. Often, you will erase the Y values but leave the X values and the column titles.

4. Switch to the sheet that you want to see when you open the template. Most often this will be a data table. If so, position the insertion point at the spot where you want it to be when you open the template.

Hint: If the template will be used by others, create a notes page that explains how to use it. Go to that notes sheet before saving the template. When people open the template, they see those instructions first.

5. Pull down the File menu and choose Save As.
6. Change the File type to Prism template and enter a name for the template file.

To use a template file:

1. Open the template file.
2. Enter or import data.
3. Click on the Results, Graphs or Layout tabs. Prism will automatically analyze and graph the data.
4. If you want to save the data, use the File Save command. Since the file you opened is a template, Prism will prompt you for a new name. This way you can save your data without disturbing the template.

Approach 4: Combining several experiments in one file

Approach 3 saves the data, results and graphs for each experiment in a separate file. Approach 4 combines several experiments into one larger file.

Before you run the first experiment, create a template file that analyzes and graphs a single experiment. Follow the steps in approach 3, and save the template.

Note: As you follow the steps below, distinguish between the template file you just created and the project file that will accumulate your data, results and graphs from several experiments.

Each time you run the experiment, follow these steps:

1. If this is your first experiment, start a new project by selecting New on the File menu. This project will hold all your experiments.

   or

   If this is a repeated experiment, open the project file (saved in step #7 below).
2. Pull down the File menu and choose Import.
3. Set the file type to Prism file. Select the experiment template.

4. Prism appends the template to your project.
5. Enter data onto the new data table and look at the automatically prepared results and graphs.
6. Rename the new data table, results and graph sheets to identify this experiment.
7. Press the Save button to save the project.
8. If this is your first experiment, Prism will prompt for a file name. Use a name different than the template name. Save a project file (*.PZM) not a template file (*.PZT).
9. Repeat steps 1-6 to add another experiment to the project.

Using a startup template

You can create a startup template file. Prism opens the startup template file whenever you start a new project – either when starting the program or when you choose New project from the File menu.

A startup template can be as simple or complex as you want. The template may contain a single data table with column labels. If you always analyze the same kind of data, the startup template could contain a data table with X values, and analyses and graphs.

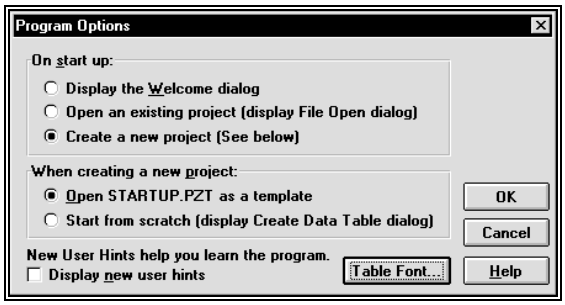
To create a startup template

1. Using the usual tools, create a project containing just those sheets you want to have available whenever you create a new project. The template may contain only a single data table or it may contain many tables with graphs, results and notes.
2. Go to the sheet you want to appear first when a new project is created.
3. If the sheet is a data table, position the insertion point at the spot you want it to appear first when a new project is created. (Probably in a column title or in the first row of the X column).
4. Pull down the File menu and choose Save As.
5. Select the file type Prism template.

6. Enter the file name STARTUP. Prism adds the extension .PZT automatically.
7. Click OK.

To use a startup template:

1. Pull down the Edit menu. Choose Options, then Program options.
2. Set the options as shown here.



Sharing Prism with several users

Prism normally stores preferences (such as the list of user defined equations, choices of the Options dialog, and default data directory) in the Windows directory. That way the same defaults will be used for all users. If several people use the same computer, each may want to maintain their own preferences and equations.

To use difference preferences for different users:

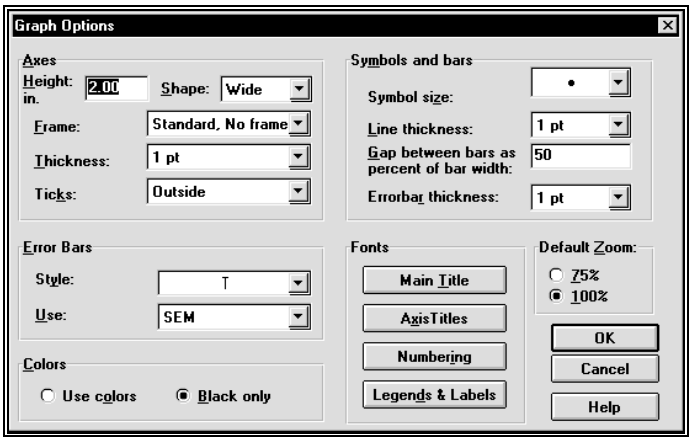
1. Create a different icon for each user. Consult your Windows manual to learn how to do this.
2. Enter the full path name of the Prism program as the first part of the command line.
3. Immediately after the program name, type "-O" (letter o, not zero) and then the full name of the file you want to use for options. For example, the command line might be:

   C:\PRISM\PRISM.EXE -OC:\PRISM\JSPRISM.CNF

Options

To set options (preferences) pull down the Edit menu and choose Options. You can set Program options, Graph options and Analysis options.

Graph options



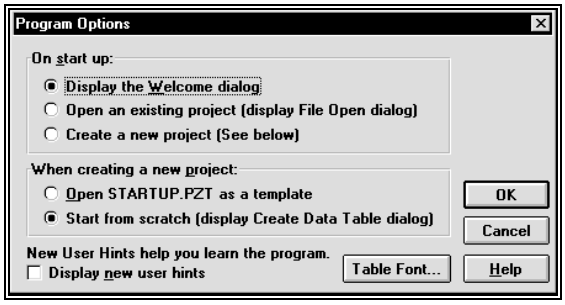
Define the appearance of graphs that Prism creates automatically. These choices affect the initial appearance of the graph. For example, you can choose the shape and size of the graph, whether it includes a frame, and the fonts used for titles and legends.

These choices only affect the initial appearance of the graphs. You can customize each graph after it is made.

Also choose 100% or 75% as the default zoom for the graphs section of your project. If you choose 75%, you can make the default graph size larger yet still see the entire graph on the screen.

Note: These choices only affect graphs you make in the future. Graphs you have already made are not affected by this dialog.

Program options

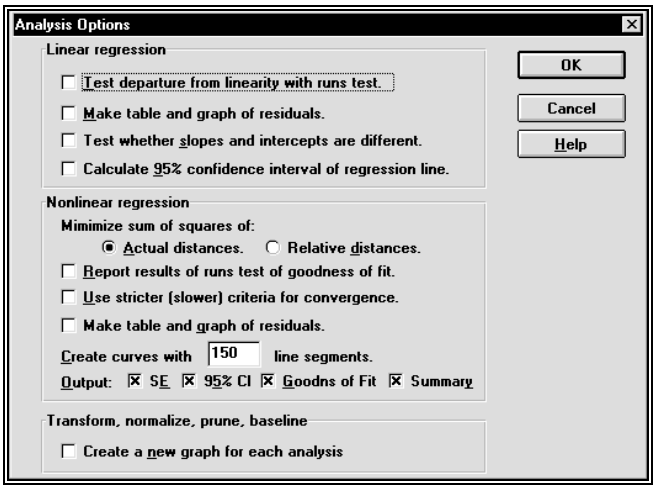


When Prism begins, it normally shows the Welcome dialog. From there, you can choose to start a new project or open an existing project.

To save time, you can opt to skip the Welcome dialog and start instead either with the File Open dialog to open an existing file, or with the Create Data Table dialog to create a new project.

Save even more time by using a startup template. See "Using a startup template" on page 132. Using a startup template will make it quicker to get started when you usually want to do the same thing.

Analysis options



The choices for linear and nonlinear regression are self-explanatory. These choices become the default choices when you perform linear and nonlinear regression in the future. You can always override the defaults.

These choices only affect analyses you do in the future. They do not affect analyses already performed.

Select "Create a new graph for each analysis" to create an automatic graph every time you transform, normalize, prune, or subtract baselines. This makes it easy to see every step in data analysis, but tends to create more graphs than you need.

Freezing results and graphs

Normally data, results and graphs are linked. If you edit the data, Prism automatically recalculates the results and redraws the graphs.

You can freeze a set of results or a graph. Frozen results and graphs are not updated when the data are edited or erased. Freeze results and graphs when you are sure they are correct, and don't want to risk inadvertent changes.

To freeze a graph or set of results, choose Freeze from the Sheet menu. If you pull down the Sheet menu again, you'll see a check mark in front of the Freeze command. Select the command again to unfreeze.

Prism places the word "[Frozen]" in front of the sheet name of frozen graphs and results. You cannot adjust the parameters of frozen results. You cannot change the appearance of frozen graphs.

It is not possible to freeze data tables, page layouts or notes.

Printing and making slides

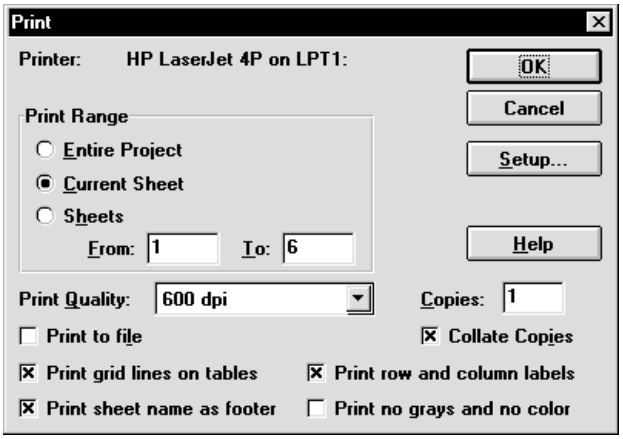
Printing

To print:

- Pull down the File menu and select Print.

  or

- Press the Print button. 



The Print dialog shows the currently selected printer. Press Setup to select a different printer.

Choose the print range:

- The entire project.
- Current sheet. If you are on a results sheet with several views, print only the view you are looking at.
- Range of sheets in the section you are currently looking at. If you are in the results section, all views of selected sheets will be printed.

Finally, set the following print options.

Print Option	Effect
Print quality.	Change the number of dots per inch (DPI). The larger the number, the more attractive the graph (but the slower it prints). Not available with all printers.
Copies.	Print multiple copies of each sheet.
Print to file.	Prism will prompt for you to enter a file name. Not available with all printers.
Print grid lines on tables.	Check this box to include the horizontal and vertical lines on data and results tables. Omit it for a cleaner look.
Print row and column titles.	When printing data tables, print the row numbers on the left of every page, and the column letters on top of every page.
Print sheet name as footer.	Print the file name, sheet name, date and time at the bottom of every page.
Print no grays and no colors.	Convert all colors to black. Don't print colors as shades of gray.

Notes:

- The Printer Setup dialog lets you choose between portrait and landscape orientation. This section applies only to data and results tables. Your choice here does not affect graphs or layouts. Choose the orientation of each graph and page layout by pressing Change from that sheet.
- If you have problems printing, see "Printing problems" on page 364.
- Printing data and results can waste paper. An alternative is to embed portions of your data and results tables on a page layout. This lets you create one page that combines graphs, data and results, or a layout page with only embedded tables. See "Embedding data and results tables" on page 113.
- You cannot change the font or point size used for printing data and results. If you embed the table onto a graph or layout, then you can change the font and point size.

Making slides

There are two approaches to making slides from Prism graphs or layouts.

Prism prints to any Windows printer driver. If you have a slide maker connected to your computer, you can install the slide maker as a printer. Then choose that "printer" and output slides directly from Prism.

Another approach is to copy graphs or page layouts to a presentation graphics program (we recommend Microsoft PowerPoint), and then output the slides from that program. Compared to Prism, a presentation graphics program provides three advantages:

- More options for background color.
- Better tools for making text slides (bullet lists, numbered lists, etc.).
- A more widely-known file format. If you bring your photo lab a PowerPoint file, they'll probably be able to make slides without difficulty.

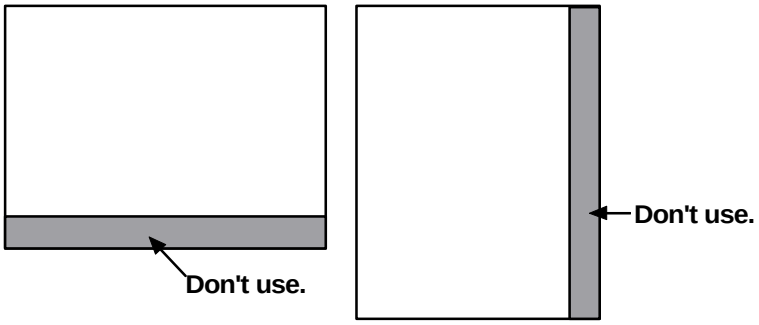
Follow these steps:

1. From a graph or layout sheet, make sure that nothing is selected. Press the Copy button or select Copy from the Edit menu.
2. Go to the presentation graphics program and paste. Since Prism places the graphs and layouts on the clipboard in Windows metafile (wmf) format, you may need to install the import filter for this format.
3. Since Prism does not include information about the background color when it copies to the clipboard, select a background color in the presentation graphics program.
4. The entire graph or page layout will be a single object in the presentation graphics program. With some programs, you can select that object and ungroup it. Once you have done this, you can move titles, adjust fonts, etc.

Notes:

- Programs differ in their ability to accept wmf files. We've found that Microsoft PowerPoint imports Prism graphs and page layouts perfectly every time (use the Insert Picture command). Not all programs read complicated wmf files correctly, so you may end up with a scrambled image. You may find that your program imports Prism graphs fine, but scrambles page layouts. Or you may find that it imports the graph or layout perfectly, but scrambles them when you try to ungroup. If you encounter problems, contact the manufacturer of the other program to see if they have a more recent version of their wmf import filter.

- An alternative approach is to export the graph from Prism to a disk file, and then import the picture into the other program. When doing this, use the wmf format when possible. Avoid the bit map formats, as these have the same resolution you see on the computer screen and result in grainy slides.
- The aspect ratio (width/height) of a slide (1.50) differs from the aspect ratio of a printed page. If you use standard US paper (8½x 11 inches) and wish to make a graph or layout fit on a slide, don't place any objects within about an inch of the bottom edge of a landscape page or the right edge of a portrait page . In other words, pretend as though the short dimension of the page is 7½ inches, rather than 8½ inches. If you use A4 (European) paper, the discrepancy between the aspect ratios is much smaller. You shouldn't place any objects within one cm. of the right edge of a portrait page or the bottom edge of a landscape page.



Saving and opening files

Opening files

You may open up to four files at one time, and each will be displayed in its own window.

To open a project file:

1. Pull down the File menu and select Open.

   or

   Click the Open button. 

2. Choose a different disk or directory if necessary.
3. Choose a file. Prism displays both project files (.PZM) and template files (.PZT). You don't need to distinguish between version 1 and version 2 files.

After the project opens, it will appear just as it was when saved. Prism remembers which sheet wass visible and the location of the insertion point.

Saving files

Save vs. Save As

The first time you save a file, use the Save dialog to choose a disk, directory and file name. After that, the Save command will save the project with that name and will not prompt for a file name. Use the Save As command to save a project with a different name or to a different drive or directory.

Save

Prism files contain the entire project — data, results, graphs, page layouts and notes. Because the files contain far more than data, we call them "project files" rather than "data files".

To save a file:

1. Pull down the File menu and select Save.

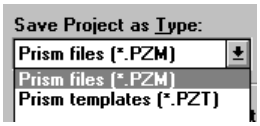
   or

   Click the Save button. 

Note: After you have saved a project once, the Save command will no longer prompt for disk, directory or file name. Use the Save As command to change file name or location.

2. Select a disk and directory.
3. Enter a file name.
4. Select whether you want to save the file in large or small format. Large files are complete. Small files omit results and preformatted graphs, so they occupy only about one half as much space on your disk. After you open a small file, Prism will reanalyze the data and recreate the graphs, so nothing is irreversibly lost. It does this when you look at results, graphs or layouts that require the omitted information. Opening the file won't take longer, but you'll wait later when you go to view results or graphs. Decide whether you'd rather save disk space or time.

5. Choose whether you want to save a project file or a template.



Note: When Prism saves the project, it remembers which sheet is showing and the location of the insertion point. When you open the file, it will appear just as it did when you saved it.

Template files

When you save a file, you can choose the standard Prism format (*.PZM) or a Prism template (*.PZT).The difference is subtle, as the two kinds of files contain exactly the same information.

Prism opens the two formats differently. After you open a template (.PZT) file, Prism leaves the project unnamed. When you save the project, therefore, Prism prompts you for a new file name. This protects you from accidentally

overwriting the template file. Choose a template file when you plan to use the file with several different sets of data in the future.

Compatibility with Prism version 1

Version 2 saves file in a format that is completely different than that used by version 1.

- Version 2 can open files created by version 1. You don't have to do anything special.
- Version 1 cannot open version 2 files, and can't even recognize them as Prism files.
- Version 2 cannot save files in version 1 format.

Exporting data

Prism can export data files. The term "export" refers to writing files to disk. In many cases, you'll find it easier to transfer information via the Windows clipboard. See "Using the clipboard" on page 41.

Prism can export data formatted as plain ASCII text with adjacent values separated by commas or tabs. These files have the extensions *.CSV or *.TXT.

Also see "Importing text files" on page 44 and "Exporting graphs" on page 144.

To export data:

1. Make sure you are looking at the data table you want to export.
2. Choose Export from the File menu.
3. At the bottom of the Export dialog, choose Tab-delimited text or Comma-delimited text. In most cases, your choice won't matter as spreadsheet programs can import either.
4. Choose the disk and directory, if necessary. Enter a file name. Press OK.
5. On the next dialog, specify whether the exported file should contain column titles.

Prism always exports the entire data table. It is not possible to export a specified range. However, you can export a selected range using this trick.

To export a selected range of data:

1. Select the desired range.
2. Press the Copy button. Or select the Edit Copy command.

3. Go to the Notes section. If you have already entered notes, start a new notes sheet.
4. Press the Paste button. Or select the Edit Paste command.
5. The selected range of data are now in the notes section.
6. Choose Export from the File menu.
7. Enter a file name.

Exporting graphs or layouts

You can move Prism graphs to other programs either via the Windows clipboard or via an exported file. Prism does not (yet) support object linking and embedding (OLE).

Also see "Importing images" on page 112 and “Exporting data” on page 143.

To copy a graph or page layout to the clipboard:

1. Go to the graph or layout. You don't need to select anything.
2. Select Copy from the Edit menu (or click the Copy button).
3. Go to another program and paste.

To export a graph or page layout as a file:

1. Go to the graph or layout you want to export.
2. Select Export from the File menu.
3. Select a file format. Prism can export graphs as a Windows Metafile (*.WMF) or as a bit map (*.TIF, *.GIF, *.PCX or *.BMP). Windows metafiles produce the highest quality images. Choose metafile if you can. TIF, PCX, GIF and BMP are bit map formats that have exactly the resolution you see on your computer screen. Avoid bit maps if possible. If you must use a bit map, make your graphs as large as possible so you get the best possible resolution. Choose BMP or PCX format, if possible, as TIF files are less standard, and Prism takes a long time to create a GIF file.

Analyzing data with Prism

Introduction to data analysis with Prism

Prism makes it very easy to analyze and manipulate your data. Prism uses the term "analyze" to include both statistical analyses (i.e. nonlinear regression) and data manipulation (i.e. mathematical transforms).

Garbage in, garbage out

Computers are wonderful tools for analyzing data. But like any tool, data analysis programs can be misused. If you enter incorrect data or pick an inappropriate analysis, the results won't be helpful. Heed the first rule of computers: Garbage in, garbage out.

This manual cannot replace a statistics text. You should learn about the analyses — especially nonlinear regression, survival curves, and ANOVA — from a statistics text before using Prism.

Note: GraphPad provides free technical support when you encounter problems with the program. We do not provide free statistics tutorials or consulting (although these may sometimes be arranged for a fee).

Automated analyses

Prism has a unique approach that automates data analysis. Prism uses these principles:

- Since most experiments compare several treatment groups, you can analyze all related data sets with one series of commands. It is no harder to analyze fifty related data sets than to analyze one.
- If you change any data, Prism updates the analyses automatically. The program remembers the links between data, analyses and graphs.

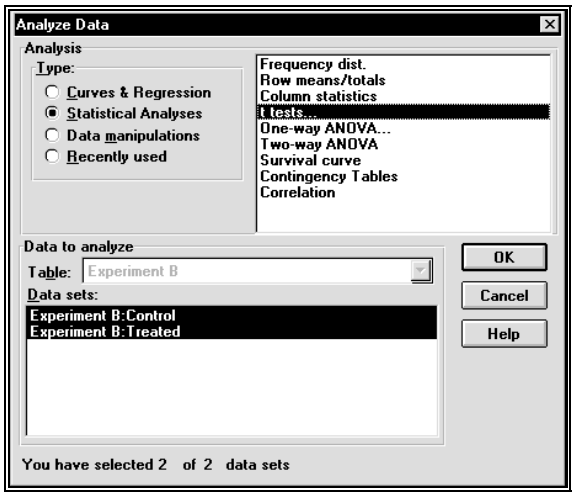
Analyzing data

There are three ways to analyze data with Prism:

- You can analyze a data table.
- You can fit a curve to a graph.
- You can further analyze a results table to chain analyses. For example, normalize data (the first analysis) and then fit a curve to the normalized values (the second analysis).

To analyze a data table:

1. Go to the data table you want to analyze.
2. Press the Analyze button.



3. The top of the Analyze dialog lists the available analyses. First choose a category: Curves and regression, Statistical analyses, or Data manipulations. Then choose an analysis from the list. A fourth category, Recently used, lists the seven analyses you have used most recently.
4. The bottom of the Analyze dialog lists all the data sets in the table you are analyzing. Initially all the data sets are selected. To analyze only some of these data sets, click on the name of the others to deselect them.
5. Press OK to bring up the Parameters dialog.
6. Choose the details of the analysis on the Parameters dialog.

To fit a curve directly on a graph:

1. Go to the graph you want to fit.

2. Press the Analyze button.
3. On top of the Analyze dialog choose the kind of curve fit (nonlinear regression, linear regression, or spline/lowess). You can also generate a theoretical curve or smooth a curve. To do other analyses, you must start from a data or results table.
4. On the bottom of the Analyze dialog, select the data sets you want to include.
5. Choose the details of the analysis on the Parameters dialog.
6. Prism will fit the curve and superimpose it on your graph. Click on the Results tab to see numerical results.

To chain analyses:

1. Look at the results of the first analysis. Note that some analyses produce several output views. Make sure you are looking at a results table you want to analyze further.
2. Click the New Analysis button.
3. On the New Analysis dialog, choose "Analyze the data you are looking at".
4. Complete the Analyze dialog.
5. Complete the Parameters dialog.

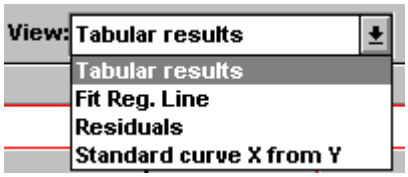
You may chain as many analyses as you want. When you edit or replace the original data, Prism will update the entire chain in order.

Viewing results

Depending on which analysis you choose, your results will be shown on one or more pages, which Prism calls "output views". For example, linear regression can produce one view that tabulates results, another that contains the XY coordinates of the regression line, another with residuals, and another with standard curve calculations.

To view results of an analysis:

1. Click on the Results tab to move to the Results section of your project.
2. Select the appropriate results sheet.
3. Drop down the list box to select a view. The list box is in the bottom row of the toolbar.



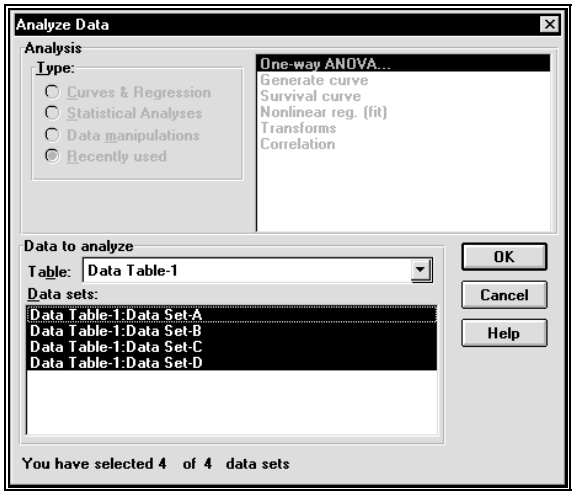
Shortcut: After dropping the list box, press the UP or DOWN arrow keys to browse through all the output views. As you change the selection in the list box, you also change the view displayed on screen.

Changing an analysis

After reviewing the results of an analysis, you may realize that you made invalid choices. If this happens, don't repeat the entire analysis. Fix the problem instead.

To change which data table and data sets to analyze:

1. Go to the results and press Change.
2. Select Data Analyzed.
3. Prism displays the Change Analysis dialog. The top portion of the dialog is dim, because you can not change one analysis to another.



4. Choose the data or results table you want to analyze. Each analysis can only use a single table as input.
5. Select one or more data sets to analyze. Click on a dataset name to select it. Click again to deselect it.

To change the parameters of an analysis:

1. While looking at the results sheet, press Change Parameters. 

2. The Parameters dialog shows your previous choices. Each analysis has a different Parameters dialog. Change settings as needed (i.e. pick a different equation for nonlinear regression). Press OK to recalculate the analysis.

Notes:

- You can also change parameters directly from a graph. Press the Change button and choose Analysis parameters.
- The word "parameters" is used in two contexts. Here, we refer to the Parameters dialog where you make all the choices for the analysis. In other contexts, the word refers to variables in an equation used for nonlinear regression.

To delete an analysis sheet:

1. Go to the results sheet you want to delete.
2. Pull down the Sheet menu.
3. Select Delete.

Frozen and orphaned results

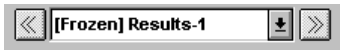
Frozen results

Data are linked to results. If you change the data, Prism automatically recalculates the analyses and updates the results.

Freeze results so they won't be updated when you edit the data. For example, you might want to freeze the results before excluding some outliers from the data. You can then repeat the analysis, and have both copies of the results in the project.

To freeze a results sheet:

1. Make sure that the sheet you want to freeze is visible.
2. From the Sheet menu, select Freeze.
3. When a sheet is frozen, the word "frozen" appears in front of the sheet name.



To unfreeze the sheet, choose Freeze from the Sheet menu again. Prism immediately recalculates the analysis.

Orphaned results

If you erase all values from a data set using the DEL key, Prism still maintains the links between the data and the results. Since there are no data, the results will be blank. When you later add data, Prism will update the results automatically.

Since Prism maintains the links between empty data tables and results, you can save a template file with no data. Add data, and Prism automatically performs the analyses. See “Template files” on page 142.

If you delete an entire data set using the Edit Delete command or an entire data sheet using the Sheet Delete command, then any results based on those data become orphaned. You can view and graph orphaned results, but you cannot adjust the parameters. When you view orphaned results, the word “orphaned” appears in front of the sheet name.

Embedding results tables in graphs or layouts

You can embed any portion of a results table onto a graph or page layout. Select the range of the table, and copy it to the clipboard. Then go to the graph or layout and paste. Double-click on the new table to bring up a dialog that lets you change its appearance. For more details, see "Embedding data and results tables" on page 113.

Preprocessing data

General

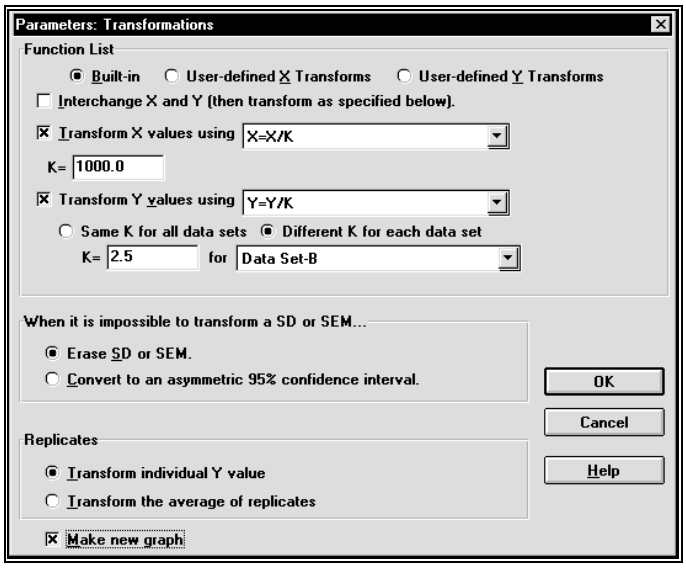
Before analyzing and graphing your data, you may need to preprocess the data by transforming, normalizing, pruning, subtracting baseline (nonspecific) values, or transposing rows and columns. These are all choices on the Analyze dialog.

These calculations do not change the original data. The results appear on a table in the results section of your project. You may analyze these processed data further.

Select the check box "Create new graph" on the Parameters dialog to create a new graph of the processed data. The default selection (whether or not to create a new graph) is set in the Analysis options dialog (see "Analysis options" on page 135).

Note: You can sort your data by X using the Edit Sort command. This command changes your data table without creating a results table.

Transforming data



Interchange X and Y values

If you check this box, Prism interchanges X and Y values.

Notes:

- If you entered more than one data set, Prism interchanges X and Y of the first data set, and will ignore all others on the sheet. There is only one X column.
- If you entered replicate Y values or mean with SD or SEM, Prism puts the mean Y value into the X column and ignores the other information.
- After interchanging, Prism performs the X and Y transforms you selected (if any). It applies the Y transform to the data that were originally in the X column, and the X transform to the data originally in the Y column.

Built-in transforms

You may transform X values, Y values or both by checking one or both option boxes.

Choose from one of these functions for transforming Y values (analogous functions are available for X):

Function	Comments

Y = Y * K	Enter K in the box provided.
Y = Y + K	"
Y = Y - K	"
Y = Y / K	"
Y = Y squared	
Y = Y ^ K	Enter K in the box provided.
Y = log(Y)	Log base 10
Y = -1*log(Y)	
Y = ln(Y)	Natural logarithm (base e)
Y = 10 ^ Y	Ten to the Y^{th} power.
Y = exp(Y)	e^{Y}
Y = 1/Y	
Y = SQRT(Y)	Square root.
Y = LOGIT(y)	ln(Y/1-Y)
Y = sin(Y)	Y is in radians.
Y = cos(Y)	"
Y = tan(Y)	"
Y = arcsin(Y)	Result is in radians.
Y = ABS(Y)	Absolute value. If Y is negative, multiply by -1.
Y = Y + Random	Gaussian. Mean = 0. SD = K (you enter).
Y = X / Y	(Not available when transforming X.)
Y = Y / X	"
Y = Y - X	"
Y = Y + X	"
Y = Y * X	"
Y = X - Y	"

Many of the functions include the variable "K", and you enter a value for K on the dialog. When transforming Y values, you can enter one value of K for all data sets or a separate value of K for each data set.

Prism always transforms all data sets selected for the analysis.To analyze only selected data sets on a data table, see "Changing an analysis" on page 148.

User-defined functions

At the top of the Parameters dialog for transforms, switch between built-in transforms and user-defined transformations of X or Y.

To enter a new user-defined function:

1. Choose User-defined X or Y transform at the top of the Transform Parameters dialog. You must specify whether you are transforming X or Y values.
2. Click Add.
3. Enter a name for the transform.
4. Enter the function. Here is an example:

```
Y=A*exp(Y)+B
```

Note: If you are transforming X values, you may not use Y in the function (because there might be several Ys for each X).

For general information, see "Choosing or entering equations" on page 287. For a list of functions you may use, see "Functions you can use in user-defined equations" on page 297.

To select a user-defined transform:

1. At the top of the Parameters dialog, select an option box for X or Y transforms.
2. Select one of the transforms in the list.
3. Enter values for the variables used in the transform.

Transferring transforms with data files

Prism maintains a list of user-defined transformations. Whenever you transform data, you can choose from transformations you used before.

What happens when you want to transfer a file to another computer? There are no explicit commands to import or export transforms. Prism handles the situation automatically by including the transform in the project file. When you open the file, Prism uses these rules:

1. Prism first reads the name of the transform from the file.
2. If a transform with exactly the same name already exists in the equation list on that computer, the equation is read from the list even if it is different than the one saved on the file. Prism will not use the transform saved with the file, but will instead use the transform with the same

name already stored on the computer. This allows you to use a revised function with stored data.

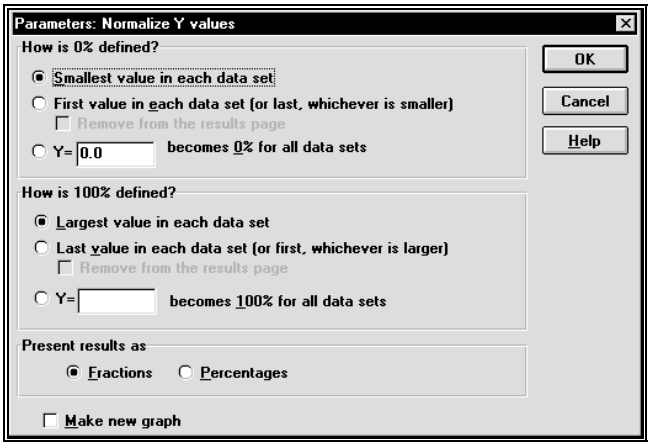
3. If the list does not contain an transform with the same name, then Prism reads the transform from the file and adds it to the list stored on the computer.

Replicates and error bars

If you entered replicate Y values, transform each replicate or the means of the replicates.

If you entered data as mean, SD (or SEM) and N, Prism tries to transform the error bar as well as the mean. When a transform is intrinsically asymmetrical (i.e. logarithms), it is mathematically impossible to transform a SD and end up with a SD. You have two choices. You may either transform the mean only and erase the error bars. Or you may convert the error bars to 95% confidence intervals, and then transform both ends of the confidence interval. The resulting 95% CI will be asymmetrical.

Normalizing data



Normalize each data set to aid comparisons. You have to make three decisions:

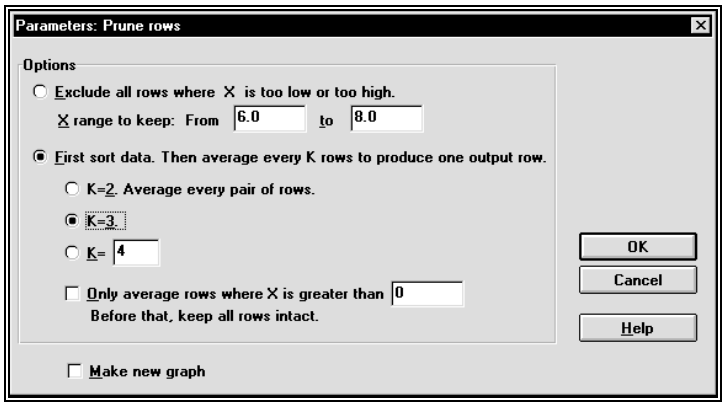
1. What is the definition of zero? You may normalize to the smallest value in each data set, to the value in the first row in each data set, or to a value you enter.
2. What is the definition of one hundred? You may normalize to the largest value in each data set, to the value in the last row in each data set, or to a value you enter.

3. Do you want to express the results as fractions or percents?

Notes:

- If you have entered replicate values, zero and one hundred percent are defined by the mean of the replicates.
- The X values are copied to the results table. They are not normalized.
- Each SD or SEM is normalized appropriately.
- If you normalize to the smallest and largest value in the data set, you can remove those values (which would be 0.000 and 1.000) from the results.

Pruning rows



This analysis reduces the size of large data sets to speed curve fitting and graphing. Use it to preprocess large data sets imported from an instrument.

There are two ways to prune data:

- Exclude all rows where X is too low or too high. Keep only data where X is between limits you enter. The output table is sorted by X value.
- Average every K rows to produce one output row (you enter K).

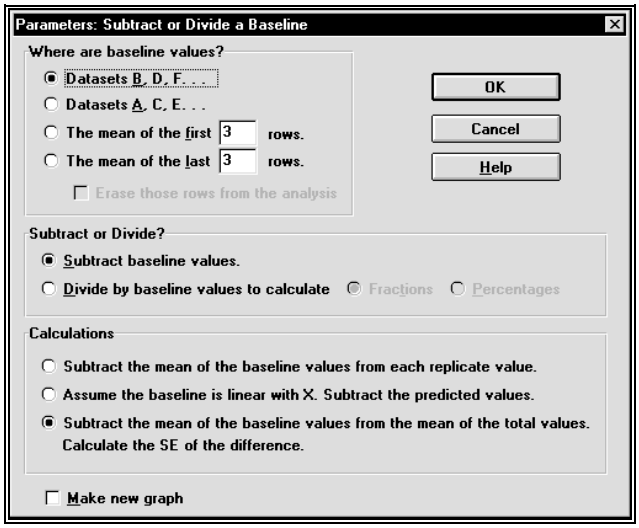
  First Prism sorts the table by X (if not already sorted). Then it averages every K rows to produce one output row. For example, if K = 3, the first X value in the results table is the average of the X values of the first three rows. The first Y value of each data set is the average of the Y values in the first three rows. The second row in the results table is the average of rows 4 to 6, and so on.

  Select the check box to only average rows after a threshold X value. Before that threshold, keep all data intact. After that threshold, reduce the number of rows by a factor of K. This is useful if your experiment

reaches a plateau value, and you only want to prune the values near the plateau.

Note: After pruning, the project contains both the original data and the pruned data. Therefore pruning increases the size of the project file. If you don't want the original data any more, you should go to that data table and use the Sheet Delete command to remove it.

Subtracting (or dividing by) baseline values



Subtracting or dividing by a baseline or nonspecific value is often the first step in analyzing data.

Where are the baseline values?

- Baseline values are in columns B, D, F, etc. Each baseline value is paired with the value just to the left of it (in column A, C, E, etc.).
- Baseline values are in columns A, C, E, etc. Each baseline value is paired with the value just to the right of it (in columns B, D, F, etc.).
- Baseline is the average of the Y values of the first N rows (you enter N).
- Baseline is the average of the Y values of the last N rows (you enter N).

Use one of the first two choices when you have measured a baseline value at every value of X. For each row, Prism will subtract the baseline value from the other value. The results table will have half as many data sets (Y columns) as the original table, but the same number of rows.

Use one of the last two choices when you have collected data over time, and the first or last few points define the baseline. The results table has the same number of data sets as the original table. If you check the option box to erase the values that define the baseline, the results table will have fewer rows than the data table.

Subtract or divide?

Choose to subtract the baseline values from the total values, or to divide the total values by the baseline values. If you divide, express the results as a fraction of baseline or as a percent of baseline.

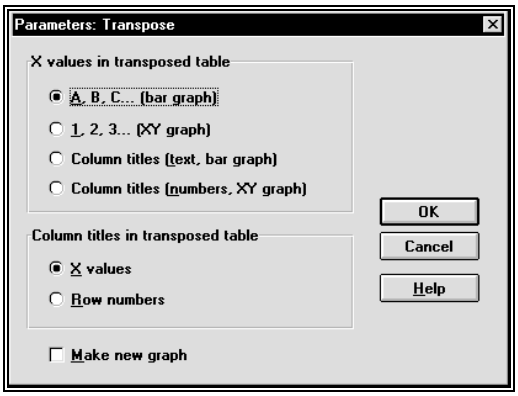
The X values of the original data table are copied to the results table without change.

Calculations

If the baseline or nonspecific values are in separate data sets (columns), there are three ways to perform the calculations:

- For each row, calculate the mean of the baseline values and subtract (divide) this value from the total values. If you entered replicate (i.e. triplicate) values, Prism subtracts the mean of the baseline from each replicate, so the results are still in replicate format.
- Assume that the baseline is linear with X. Prism performs linear regression with the background values (it does not report the regression results). It then subtracts (or divides) the Y value predicted from that line. If you know that the nonspecific or background values must be linear with the X value (for example nonspecific binding as a function of ligand concentration) this method is more accurate. It is particularly useful when you have not collected baseline or nonspecific measurements at every value of X, as Prism will fill in the missing nonspecific values from linear regression.
- For each row, calculate the mean of the total values and the mean of the nonspecific values. Prism reports the difference and its standard error. This option is not available if you choose to divide (rather than subtract).

Transposing rows and columns



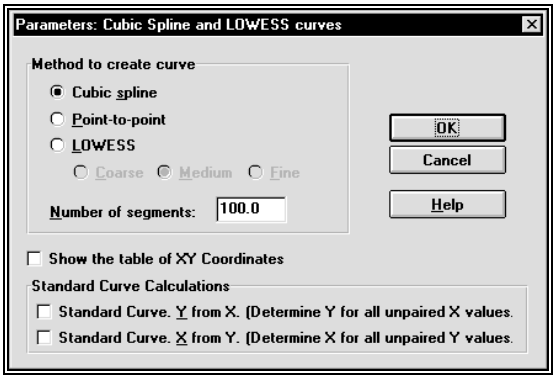
Prism transposes the Y values using these rules:

- The first row becomes the first data set (column), the second row becomes the second data set (column), etc.
- The first data set (column) becomes the first row, the second data set (column) becomes the second row, etc.
- You may not transpose a data table with more than 52 rows, because Prism cannot create a table with more than 52 columns.
- The column and row titles are determined by your choices in the dialog.

Transpose rows and columns to change the appearance of bar graphs. For an example, see "Transposing rows and columns of bar graphs" on page 79.

Miscellaneous data analyses

Creating a LOWESS or cubic spline curve



Create a curve that goes through, or near, your data points using one of three methods.

- A cubic spline curve goes through every point. If you entered replicate values, then the cubic spline goes through the mean of each set of replicates. The spline will bend and twist to hit every point, no matter how scattered the data.
- A point to point curve is a series of line segments that join every point. Don't create a point-to-point curve just so you can connect points with a line on the graph. You can do that by checking a box on the Symbols & Lines dialog from the Graphs section of your project. Only select the point-to-point analysis if you want to use the point-to-point line as a standard curve or to calculate area under the curve.
- A LOWESS (locally weighted) curve is a relatively smooth curve that follows the trend of the data. It is calculated from an algorithm adapted from Graphical Methods for Data Analysis, John Chambers et. al., Wadsworth and Brooks, 1983. A LOWESS curve is helpful when the data progresses monotonically, but is not helpful with peaks or valleys.

LOWESS curves are not recommended unless you have well over twenty data points.

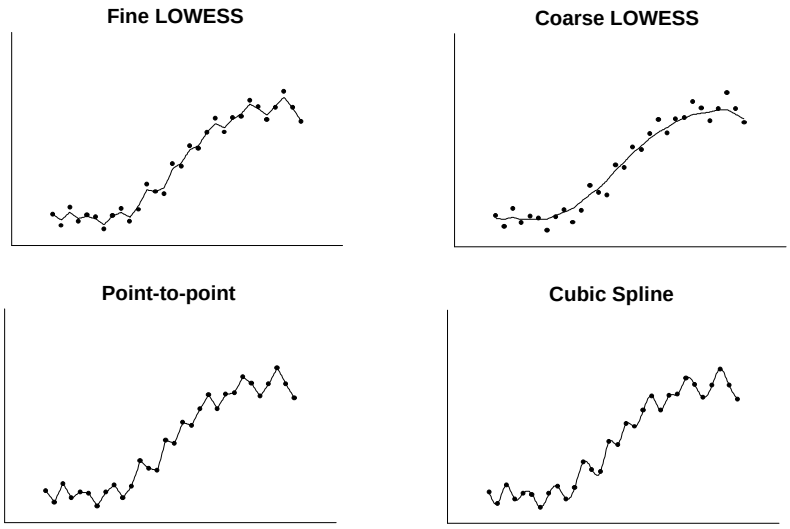
Choose between fine, medium and coarse LOWESS curves. The fine curve reveals the fine structure of the data, but tends to wiggle a lot. The coarse curve shows only the coarse trend, but obscures the detail.

Options when creating spline, point-to-point or LOWESS curves

You have several options:

- The curve is, in fact, a series of line segments. Enter the number of segments you want.
- If you want to see the table of XY coordinates, check the option box.
- You may use the LOWESS, point-to-point, or spline curve as a standard curve from which to read unknown values. See "Reading unknowns from standard curves" on page 349.

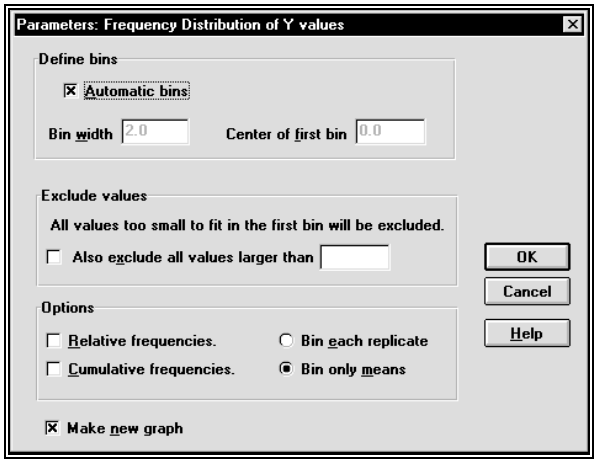
Examples of cubic spline, point-to-point and LOWESS curves



Frequency distributions

A frequency distribution shows the distribution of Y values in each data set. The range of Y values is divided into "bins" and Prism determines how many values fall into each bin. The size of each bin is set automatically or by you.

Note: A frequency distribution completely ignores any values you entered in the X column.



Make these choices:

- Clear "Automatic bins" to enter bin width manually. Histograms look best when the bin width is a round number and there are 10-20 bins. Also enter the center of the first bin. All data too small to fit in the first bin are excluded from the analysis.
- If you entered replicate values, decide whether you wish to bin each replicate or only the means.
- Enter an upper limit if you want to exclude larger values from the analysis.
- Select "Relative frequencies" to express the results as fractions rather than counts. Assume that you have 50 data points, and 15 fall into the third bin. If the results are expressed as relative frequencies, the results for the third bin will be 0.30 (15/50) rather than 15.
- Select "Cumulative binning" to bin cumulatively rather than individually. In this case each bin contains the number of data points that fall within that bin or any smaller bin, and the last bin contains the total number of values.

Interpreting and graphing frequency distributions

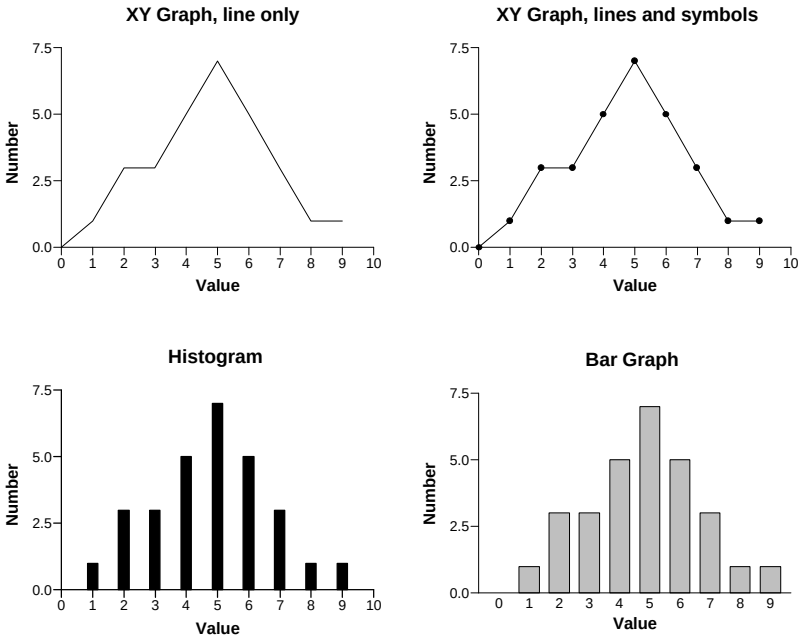
Prism creates two results views:

- The frequency distribution is shown on a table. The X values are the centers of each bin. The Y values in the table are the number of observations that fall into each bin.
- The second results view contains the column statistics. For more information, see "Column statistics" on page 181.

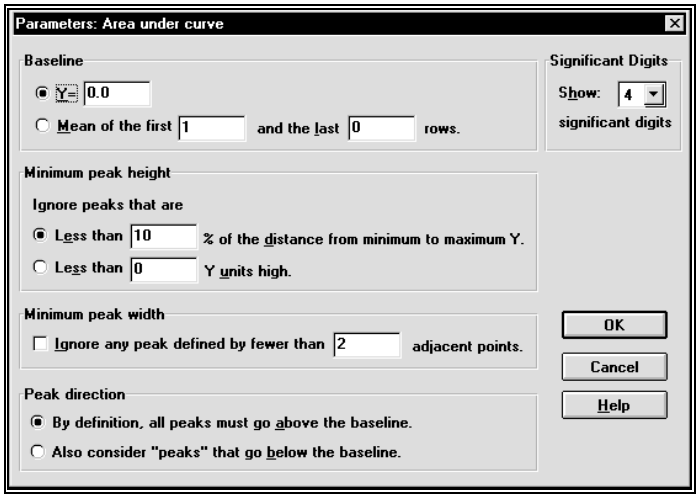
The figure below shows four ways to plot the frequency distribution. In all cases, the X axis shows the bin centers and the Y axis plots the number of data points in that bin.

- XY graph. The upper left panel shows the graph that Prism creates automatically with connecting lines but no symbols. The upper right panel shows data points as well as connecting lines.
- Histogram (XY graph with bars). The lower left graph is a histogram. To make this graph, double-click on one of the symbols to bring up the Symbol dialog. Then choose one of the last four choices for shape.
- Bar graph. The lower-right example is a bar graph. To make this graph, press Change and select Graph Type. Choose Bar graph.

  When you make a bar graph, you have more control over the fill pattern than you do with a histogram. You can also control the size of the gap between bars (or choose no gap). However you have less control over the X axis. Each bar is labeled with the bin center. If you have lots of bins, the labeling will be cluttered.



Area under the curve



Parameters for area under curve

- Define the baseline with a particular Y value (usually Y=0) or calculate the baseline as the mean of the first and last few values (you define how many). The baseline must be horizontal (parallel to the X axis).
- In order to avoid finding many "peaks" in noisy data, define the minimum height that you consider worth finding. Enter either the height measured in units of the Y axis, or as a percentage of the way between the minimum and maximum Y values.
- Optionally define the minimum width of a peak worth considering. Give the width as the number of adjacent rows in the data table. Do not use X units.
- Choose whether to consider negative peaks.

Interpreting area under the curve

Peaks are defined using the options you specified in the dialog. The area of a peak is determined by the trapezoid rule.

For each peak, Prism shows:

- The X and Y coordinates of the highest point.
- The X coordinates of the beginning and end of the peak.
- The area under the peak. The units are the units of the X axis times the units of the Y axis.
- The area under the peak as a fraction of the area under all peaks.

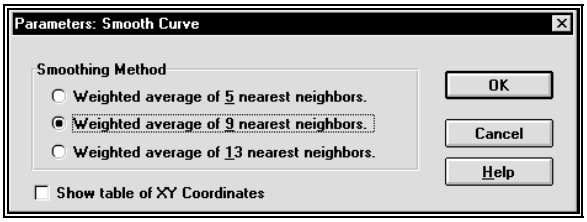
Note the limitations of this analysis:

- The baseline must be horizontal.
- There is no smoothing or curve fitting.
- Prism cannot separate overlapping peaks. It will not be able to distinguish two peaks unless the signal descends all the way to the baseline between peaks. It is not able to find a peak within a shoulder of another peak.
- If the signal starts (or ends) above the baseline, the first (or last) peak will be incomplete. Prism reports the area under the tails it sees.

Finding the area under the entire curve

If all values are above the baseline, then this analysis finds one peak which is the area under the entire curve. This is useful in analysis of pharmacokinetic data.

Smooth curve



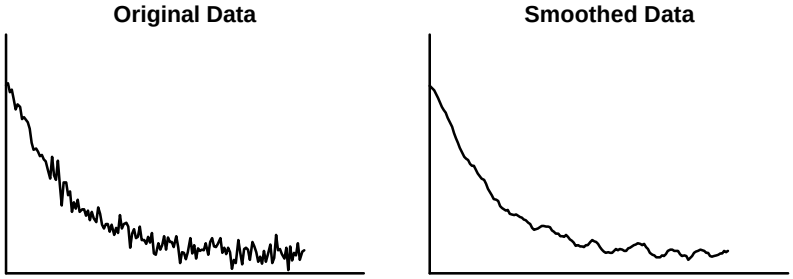
If you import a curve from an instrument, smooth the data to improve the appearance of a graph.

Prism can only smooth data sets where the X values are equally spaced. The X values in the table may be formatted either as individual numbers or as a sequence (you define the first value and the interval, and Prism fills in the rest).

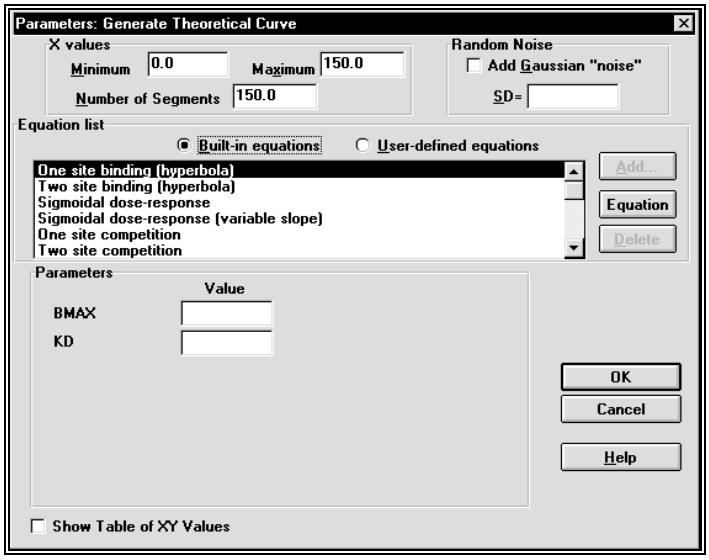
Each point in the curve is replaced by the weighted average of its nearest five, nine or thirteen neighbors by the method of Savitsky and Golay (Analytical Chemistry, 36:1627-1639, 1964) using a cubic equation.

If you want to see the table of XY coordinates, check the option box.

Warning: You lose information when you smooth a curve. It does not help to smooth data prior to running linear or nonlinear regression.



Generate a theoretical curve



Generate a curve from a built-in or user-defined equation, perhaps with random error. There are three reasons to do this:

- To test analysis methods. Create a curve with random error, and then analyze the simulated data.
- To teach theory.
- To get a feel for what the parameters in an equation mean. Generating curves is the best way to learn about equations. See how the curve changes when you alter a variable.

To generate a curve:

1. Press the Analyze button from a data table, or press New Analysis from a results sheet.
2. Select Generate Theoretical Curve.
3. Enter the range of X values.
4. Enter the number of line segments you want to calculate. Prism generates a curve as a series of XY coordinates. You need to decide how many to generate. The default value of 150 is sufficient for most purposes. Increasing the number of segments may make the curve appear smoother.
5. Check the option box if you want to add random error to the curve to simulate experimental error. Prism generates random errors that follow a Gaussian (bell-shaped) distribution with a SD you enter.

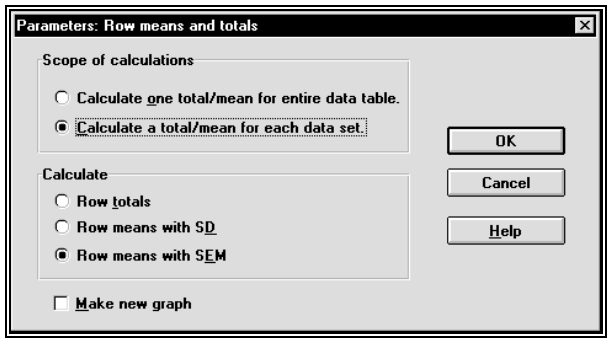
6. Choose an equation. Pick a built-in equation or enter your own. For more information, see "Choosing or entering equations" on page 287.
7. Enter values for all the variables in the equation.
8. Check the option box if you want to see a table of XY values.

A note about random numbers

Prism can add random values to each of the calculated Y values to simulate experimental error. Prism generates random numbers using routines from Numerical Recipes in C, (W. H. Press et al, Cambridge Press, 1992). It generates uniformly distributed random numbers using RAN3 and converts these to a Gaussian distribution using GASDEV. Prism uses the time of day when calculating the first random number, so you will get a different series of random numbers every time you run the program.

The only way to generate truly random numbers is through a random physical process such as tossing dice or measuring intervals between radioactive decays. Prism, like all computer programs, generates "random" numbers from defined calculations. Since the sequence of numbers is reproducible, mathematicians say that the numbers are "pseudo-random". It turns out that the difference between truly random and pseudo-random numbers rarely creates a problem. Computer generated random numbers are random enough to simulate data and test analytical methods.

Row means or totals



This analysis lets you consolidate several data sets on one table. It also lets you see the mean and SD (or SEM) of replicate values. You don't need to run this analysis just to plot SD or SEM error bars Prism does that automatically but use this analysis when you want to see a table of the SD or SEM values.

First, decide on the scope of the analysis. If you have entered more than one data set on the table, you have two choices:

- Calculate one row total/mean for the entire table. Consolidate all data from all data sets by row. With this choice, the results table will have a single data set containing the totals or means.
- Calculate a row total/mean for each data set. Consolidate all replicates from each data set by row. With this choice, the results table will have the same number of data sets as the input table. Instead of showing replicate values, the table will have a single value (for totals) or means with SD or SEM per data set.

Then decide on the calculation. You have three choices:

- Row totals.
- Row means with SD.
- Row means with SEM.

Here are some examples:

- You entered quadruplicates (four replicates) for each X value and want to see the mean and SD at each X value. Note that you don't need to perform this analysis in order to graph the mean and SD — Prism can graph error bars from replicate values directly from the Symbols and lines dialog. Only select this analysis if you want to see the mean and SD as numbers.
- You entered quadruplicate values for each X value for three different experiments. Now you want to consolidate the data, and generate one mean and SD from all three experiments. Prism will first average the replicates for each experiment, and then take the average and SD of the three means.
- You entered quarterly sales figures into the four columns. Now you want to make a table of total sales.

Introduction to statistical comparisons

Why do we need statistical calculations?

When analyzing data, your goal is simple: You wish to make the strongest possible conclusion from limited amounts of data. To do this, you need to overcome two problems:

- Important differences can be obscured by biological variability and experimental imprecision. This makes it hard to distinguish real differences from random variability.
- The human brain excels at finding patterns, even from random data. Our natural inclination (especially with our own data) is to conclude that differences are real, and to minimize the contribution of random variability. Statistical rigor prevents you from making this mistake.

Statistical analyses are most useful when you are looking for differences that are small compared to experimental imprecision and biological variability. If you only care about large differences, you may follow these aphorisms:

> *If you need statistics to analyze your experiment, then you've done the wrong experiment.*
>
> *If your data speak for themselves, don't interrupt!*

But in many fields, scientists care about small differences and are faced with large amounts of variability. Statistical methods are necessary.

This chapter reviews the basic principles of statistics. Later chapters help you make the appropriate choices and interpret the results from specific tests.

Population vs. samples

The basic idea of statistics is simple: you want to extrapolate from the data you have collected to make general conclusions. Statistical analyses are based on a simple model. There is a large population of data out there, and you have randomly sampled parts of it. You analyze your sample to make inferences about the population.

Situation	Sample	Population
Quality control.	The items you tested.	The entire batch of items produced.
Political polls.	The voters you polled.	All voters.
Clinical studies.	Subset of patients who attended Tuesday morning clinic in August.	All similar patients.
Laboratory research.	The data you actually collected.	All the data you could have collected if you had repeated the experiment many times the same way.

The logic of statistics assumes that your sample is randomly selected from the population, and that you only want to extrapolate to that population. This works perfectly for quality control. When you apply this logic to scientific data, you encounter two problems:

- You don't really have a random sample. It is rare for a scientist to randomly select subjects from a population. More often you just did an experiment a few times and want to extrapolate to the more general situation. It is sufficient that your data be representative of the population, and that the population be hypothetical.
- You want to make conclusions that extrapolate beyond the population. The statistical inferences only apply to the population your samples were obtained from. Let's say you perform an experiment in the lab three times. All the experiments used the same cell preparation, the same buffers, and the same equipment. Statistical inferences let you make conclusions about what would happen if you repeated the experiment many more times with that same cell preparation, those same buffers, and the same equipment. You probably want to extrapolate further to what would happen if someone else repeated the experiment

with a different source of cells, freshly made buffer and different instruments. Statistics can't help with this further extrapolation. You can use scientific judgment and common sense to make inferences that go beyond statistics. Statistical logic is only part of data interpretation.

Assumption of independence

It is not enough that your data are sampled from a population. Statistical tests are also based on the assumption that each subject (or each experimental unit) was sampled independently of the rest. The assumptions of independence is easiest to understand by studying counterexamples.

- You are measuring blood pressure in animals. You have five animals in each group, and measure the blood pressure three times in each animal. You do not have 15 independent measurements, because the triplicate measurements in one animal are likely to be closer to each other than to measurements from the other animals. You should average the three measurements in each animal. Now you have five mean values that are independent of each other.
- You have done a laboratory experiment three times, each time in triplicate. You do not have nine independent values. If you average the triplicates, you do have three independent mean values.
- You are doing a clinical study, and recruit ten patients from an inner-city hospital and ten more patients from a suburban clinic. You have not independently sampled 20 subjects from one population. The data from the ten inner-city patients may be closer to each other than to the data from the suburban patients. You have sampled from two populations, and need to account for this in your analysis. This is a complicated situation, and you should probably contact a statistician.

Confidence intervals

Statistical calculations produce two kinds of results that help you make inferences about the populations from the samples. Confidence intervals are explained here, and P values are explained in the next section.

95% confidence interval of a mean

Although the calculation is exact, the mean you calculate from a sample is only an estimate of the population mean. How good is the estimate? It depends on how large your sample is and how much the values differ from one another. Statistical calculations combine sample size and variability to generate a confidence interval for the population mean. You can calculate intervals for any desired degree of confidence, but 95% confidence intervals are used

most commonly. If you assume that your sample is randomly selected from some population, you can be 95% sure that the confidence interval includes the population mean. More precisely, if you generate many 95% CI from many data sets, you expect the CI to include the true population mean in 95% of the cases and not to include the true mean value in the other 5%. Since you don't know the population mean, you'll never know when this happens.

Confidence interval for the difference between means

When comparing groups, calculate the 95% confidence interval for the difference between the population means. Again interpretation is straightforward. If you accept the assumptions, there is a 95% chance that the interval you calculate includes the true difference between population means.

Other situations

Methods exist to compute a 95% confidence interval for any calculated statistic, for example the relative risk or the best-fit value in nonlinear regression. The interpretation is the same in all cases. If you accept the assumptions of the test, you can be 95% sure that the interval contains the true population value.

P values

What is a P value?

You've measured values in two samples, and the means are different. How sure are you that the population means are different as well? There are two possibilities:

- The populations have different means.
- The populations have the same mean, and the difference you observed is a coincidence of random sampling.

The P value answers this question:

If the populations really did have the same mean overall, what is the probability of observing such a large difference (or larger) between sample means in an experiment of this size?

The P value is a probability, with a value ranging from zero to one.

Common misinterpretation of a P value

Many people misunderstand what question a P value answers.

If the P value is 0.03, that means that there is a 3% chance of observing a difference as large as you observed even if the two population means are identical. It is tempting to conclude, therefore, that there is a 97% chance that the difference you observed reflects a real difference between populations and a 3% chance that the difference is due to chance. Wrong. What you can say is that random sampling from identical populations would lead to a difference smaller than you observed in 97% of experiments and larger than you observed in 3% of experiments.

You have to choose. Would you rather believe in a 3% coincidence? Or that the population means are really different?

Statistical hypothesis testing

The P value is a fraction. In many situations, the best thing to do is report that number to summarize the results of a comparison. If you do this, you can totally avoid the term "statistically significant", which is often misinterpreted.

In other situations, you'll want to make a decision based on a single comparison. In these situations, follow the steps of statistical hypothesis testing.

1. Set a threshold P value before you do the experiment. Ideally, you should set this value based on the relative consequences of missing a true difference or falsely finding a difference. In fact, the threshold value (called α) is traditionally almost always set to 0.05.
2. Define the null hypothesis. If you are comparing two means, the null hypothesis is that the two populations have the same mean.
3. Do the appropriate statistical test to compute the P value.
4. Compare the P value to the preset threshold value.
    - If the P value is less than the threshold, state that you "reject the null hypothesis" and that the difference is "statistically significant".
    - If the P value is greater than the threshold, state that you "do not reject the null hypothesis" and that the difference is "not statistically significant".

Statistical significance

The term *significant* is seductive, and it is easy to misinterpret it. A result is said to be *statistically significant* when the result would be surprising if the populations were really identical.

It is easy to read far too much into the word *significant* because the statistical use of the word has a meaning entirely distinct from its usual meaning. Just because a difference is *statistically significant* does not mean that it is important or interesting. And a result that is not *statistically significant* (in the first experiment) may turn out to be very important.

If a result is statistically significant, there are two possible explanations:

- The populations are identical, so there really is no difference. You happened to randomly obtain larger values in one group and smaller values in the other, and the difference was large enough to generate a P value less than the threshold you set. Finding a statistically significant result when the populations are identical is called making a Type I error.
- The populations really are different, so your conclusion is correct.

There are also two explanations for a result that is not statistically significant:

- The populations are identical, so there really is no difference. Any difference you observed in the experiment was a coincidence. Your conclusion of no significant difference is correct.
- The populations really are different, but you missed the difference due to some combination of small sample size, high variability and bad luck. The difference in your experiment was not large enough to be statistically significant. Finding results that are not statistically significant when the populations are different is called making a Type II error.

"Extremely significant" results

Intuitively, you may think that P = 0.0001 is more statistically significant than P = 0.04. Using strict definitions, this is not correct. Once you have set a threshold P value for statistical significance, every result is either statistically significant or is not statistically significant. Some statisticians feel very strongly about this. Many scientists are not so rigid, and refer to results as being "very significant" or "extremely significant" when the P value is tiny.

Prism summarizes the P value using the symbols shown in the first column of this table. You may refer to the result using the wording of the third column.

Summary	P value	Wording
ns	> 0.05	Not significant
*	0.01 to 0.05	Significant
**	0.001 to 0.01	Very significant
***	< 0.001	Extremely significant

You can copy the summary symbol to the clipboard, and then paste it onto a graph. If you edit or replace the data, Prism will automatically update the analysis, and update the P value summary as well. See "Embedding results tables in graphs or layouts" on page 150.

*The definitions of *, **, and *** are not entirely standard. If you report the results in this way, you should define the symbols in the figure legend.*

One- vs. two-tail P values

When comparing two groups, you must distinguish between one- and two-tail P values.

Start with the null hypothesis that the two populations really are the same and that the observed discrepancy between sample means is due to chance.

- The two-tail P value answers this question: Assuming the null hypothesis, what is the chance that randomly selected samples would have means as far apart as observed in this experiment with either group having the larger mean?
- To interpret a one-tail P value, you must predict which group will have the larger mean before collecting any data. The one-tail P value answers this question: Assuming the null hypothesis, what is the chance that randomly selected samples would have means as far apart as observed in this experiment with the specified group having the larger mean?

A one-tail P value is appropriate only when previous data, physical limitations or common sense tell you that a difference, if any, can only go in one direction. The issue is not whether you expect a difference to exist – that is what you are trying to find out with the experiment. The issue is whether you should interpret increases and decreases the same.

You should only choose a one-tail P value when you believe the following:

- Before collecting any data, you can predict which group will have the larger mean (if the means are in fact different).
- If the other group ends up with the larger mean, then you should be willing to attribute that difference to chance, no matter how large the difference.

It is usually best to use a two-tail P value for these reasons:

- The relationship between P values and confidence intervals is more clear with two-tail P values.
- Some tests compare three or more groups, which makes the concept of tails inappropriate (more precisely, the P values have many tails). A two-tail P value is more consistent with the P values reported by these tests.
- Choosing a one-tail P value can pose a dilemma. What would you do if you chose a one-tail P value, but observed a large difference in the op-

posite direction to the experimental hypothesis? To be rigorous, you should conclude that the difference is due to chance, and that the difference is not statistically significant. But most people would be tempted to switch to a two-tail P value or to reverse the direction of the experimental hypothesis. You avoid this situation by always using two-tail P values.

A bit about Bayes

The P value does not answer the question you really want the answer to:

If I observed data like these many times, in what fraction of the experiments would the results reflect real differences between the populations?

To answer this question requires understanding a bit about Bayesian statistics. You can't answer the question unless you can define the probability that there is a real difference between populations, based on information you knew before collecting these data. Bayesian statistics combine this prior probability with the experimental evidence to calculate the post probability that answers the question posed above. Prism does not do these calculations, but the companion program, GraphPad StatMate, does. See the StatMate manual for details.

When comparing three or more groups, what's wrong with performing multiple t tests?

When comparing three or more groups, your first thought might be to compare the groups one pair at a time. For example, to compare four groups, you might perform six tests (A vs. B, A vs. C, A vs., D, B vs. C, B vs. D, C vs. D). You should not do this.

Making multiple comparisons increases the chance of finding a statistically significant difference by chance. Even if all populations have the same mean, there is a 5% chance that each t test would result in a P value less than 0.05 and thus be deemed statistically significant just by chance. But you'd be performing six t tests. Even if the null hypotheses were all true, there is far more than a 5% chance that at least one of these P value would be less than 0.05.

When you have more than two groups, you should analyze them all at once. The statistical method that does this is called one-way analysis of variance.

Statistical comparisons performed by Prism

This table lists all the statistical comparisons offered by Prism. The next few chapters give the details.

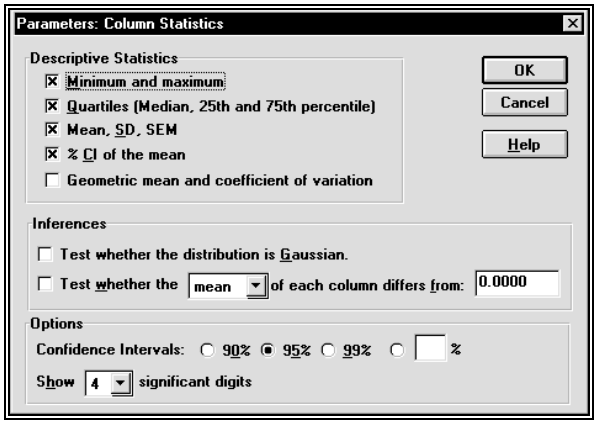
Comparison	Statistical test	Prism analysis

One column with hypothetical value	One-sample t test Wilcoxon test	Column Statistics
Compare two columns	Unpaired t tests Paired t tests Mann-Whitney test Wilcoxon test	t tests
Three or more columns	One-way ANOVA Repeated measures ANOVA Kruskal Wallis test Friedman test	One-way ANOVA
Simultaneously investigate row and column effects	Two-way ANOVA	Two-way ANOVA
Two or more survival curves	Log-rank test	Survival curves
Two proportions	Fisher's exact test	Contingency table
Contingency table	Chi-square test	Contingency table

Column statistics and one-sample t test

Column statistics

The column statistics analysis calculates statistics that describe the distribution of values in a column. It also tests whether the mean or median differs significantly from a hypothetical value and whether the distribution of values differs significantly from a Gaussian distribution.



Variable	Meaning
Range	Largest and smallest value.
Quartiles	The 25th percentile, median, and 75th percentile. One quarter of the values are less than or equal to the 25th percentile. Three quarters of the values are less than or equal to the 75th percentile. The median is the 50th percentile.

Mean	Average.
SD	Standard deviation. A measure of variability.
SEM	Standard error of the mean. A measure of how close this sample mean is likely to be to the true population mean.
95% CI	Assuming that your sample is representative of a larger Gaussian population, you can be 95% sure that the interval contains the true population mean. You can choose other levels of confidence (rather than 95%).
Geometric mean	The antilog of the mean of the logarithm of the values. It is less affected by outliers than the mean.
CV	Coefficient of variation. The standard deviation divided by the mean (expressed as a percent). Because it is a unitless ratio, you can compare the CV of variables expressed in different units.

Notes:

- If you entered replicate values, Prism first calculates the mean of the replicates, and then calculates the column statistics of the means. If you enter ten rows of data in triplicate, the column statistics are calculated from the ten row means, not from the 30 individual values.
- If you entered mean data with SD or SEM, Prism calculates column statistics for the means and ignores the SD or SEM values.
- If you entered X values, they are ignored.
- The SD, SEM and CI are explained later in this chapter.

Testing whether the distribution is Gaussian

Many statistical tests depend on the assumption that the values in your sample were obtained from a population that follows a Gaussian bell-shaped curve.

Prism tests for deviations from Gaussian distribution using the Kolmogorov-Smirnov test. The P value comes from Dallal and Wilkinson's approximation to Lilliefors' method (Am. Statistician, 40:294-296, 1986). The plain Kolmogorov-Smirnov test cannot be used because we don't know the mean and SD of the population. Instead, we estimate these values from the sample being analyzed.

Prism reports the Kolmogorov-Smirnov distance, abbreviated KS distance. If the sample followed a Gaussian distribution exactly, this would equal zero. Larger values of the KS distance correspond to larger deviations from an ideal Gaussian distribution.

Prism also reports a P value which answers this question: If the population were really Gaussian, what is the probability that a randomly selected sample of this size would deviate as much from a Gaussian distribution (or more so) than observed here. More precisely, the P value answers this question: If the population were really Gaussian, what is the chance that a randomly selected sample of this size would have a KS distance as large, or larger, than observed?

Your conclusion must depend on the P value and the sample size.

- If the P value is small, the data failed the normality test. You can conclude that the population is unlikely to be Gaussian (regardless of sample size).
- If the P value is large and the sample size is large, the data passed the normality test. You can conclude that the population is likely to be Gaussian, or nearly so. How large is large? There is no firm answer, but one rule-of-thumb is that the normality tests are only useful when your sample size is 100 or more.
- If the P value is large, and the sample size is small, you will be tempted to conclude that the population is Gaussian. Don't do that. A large P value just means that the data are not inconsistent with a Gaussian population. That doesn't exclude the possibility of a nongaussian population. Small sample sizes simply don't provide enough data to discriminate between Gaussian and nongaussian distributions. You can't conclude much about the distribution of a population if your sample contains less than a dozen values.

Notes:

- If a data set has less than four values, it is impossible to test normality.
- If the P value is greater than 0.10, Prism simply reports "P>0.10" and does not try to calculate what the P value is.
- This test has limited usefulness in helping you decide whether you want to use standard (t tests, ANOVA) or nonparametric tests. The normality test only works well with large samples. But if you have a large sample, it doesn't matter too much whether you choose a parametric or nonparametric test (see "Why sample size matters when choosing between parametric tests and nonparametric tests" on page 193).

The one-sample t test and Wilcoxon test

The one-sample t test and Wilcoxon test determine whether the values in each column differ significantly from a hypothetical value. You need to:

- Enter the hypothetical value, often 0, 1, or 100.
- Choose the test. You can compare the mean with the hypothetical value with a one-sample t test or compare the median with the hypothetical value using the Wilcoxon signed rank test.

Both tests calculate a P value that answers this question: If the true mean (median) equals the hypothetical value, what is the chance of randomly selecting N data points and finding a mean (median) as far (or further) from the hypothetical value as observed here.

The one-sample t test assumes that the data are sampled from a Gaussian population. Both tests assume that the sample is randomly selected from, or at least representative of, the overall population, and that each value was obtained independently.

Interpreting the SD, SEM and 95% CI

Standard deviation

The standard deviation (SD) quantifies variability. Prism calculates the "sample SD" (which uses a denominator of N-1), not the "population SD" with a denominator of N. If the data follow a bell-shaped Gaussian distribution, then 68% of the values lie within one SD of the mean (on either side) and 95% of the values lie within two SD of the mean. The SD is expressed in the same units as your data.

Standard Error of the Mean

The standard error of the mean (SEM) is a measure of how far your sample mean is likely to be from the true population mean. The SEM is calculated as the SD divided by the square root of sample size. With large samples, therefore, the SEM is always small. By itself, the SEM is difficult to interpret. It is easier to interpret the 95% confidence interval, which is calculated from the SEM.

Confidence interval

The mean you calculate from your sample of data points depends on which values you happened to sample. Therefore, the mean you calculate is unlikely to equal the overall population mean exactly. The size of the likely discrepancy depends on the variability of the values (expressed as the SD) and the sample size. Combine those together to calculate a 95% confidence interval (95% CI), which is a range of values. You can be 95% sure that this interval contains the true population mean. More precisely, if you generate many 95% CI from many data sets, you expect the CI to include the true population mean in 95% of the cases and not to include the true mean value in the other 5%. Since you don't know the population mean, you'll never know when this happens.

Comparing groups (t tests, ANOVA, etc.)

Entering data for column comparisons

The t test analysis compares two columns. The one-way ANOVA analysis compares three or more columns. There are three ways to enter data for a t test or one-way ANOVA.

- Enter raw data by column (data set). The data table should be formatted to hold single Y values. Each data set must have more than one row.

	A	B
	Control	Treated
	Y	Y
1	34.0	45.0
2	43.0	47.0
3	39.0	52.0

- Enter averaged data. Format the data table for mean, SD (or SEM) and N. With this format, you can't pick nonparametric or paired tests which require raw data.

A			B		
Control			Treated		
Y	SEM	N	Y	SEM	N
38.667	2.603	3	48.000	2.082	3

- Enter raw data on one row. Format the data table for replicate Y values. Enter all the data in one row. With this format, you can't pick non-parametric or paired tests, and you must enter only a single row of data. We don't recommend using this format.

	A			B		
	Control			Treated		
	Y1	Y2	Y3	Y1	Y2	Y3
1	34	43	39	45	47	52

Notes:

- If your data has more than one row and more than one replicate, Prism averages the replicates and compares the means.
- If data are missing, just leave a blank spot in the table. With the exception of repeated measures ANOVA, Prism can perform all the tests even if values are missing. Repeated-measures ANOVA cannot be calculated when values are missing.
- Many statistics programs expect you to enter data in an indexed format, as shown below. One column contains all the data, and the other column designates the group. Prism cannot analyze data entered in index format.

Group	Value
1	34
1	43
1	39
2	45
2	47
2	52

Advantages of transforming the data first

Before comparing columns, consider whether you should first transform the values. The t test and ANOVA depend on the assumption that your data are sampled from a population that follows a Gaussian distribution. If your data do not follow a Gaussian (normal) distribution, you may be able to transform the values to create a Gaussian distribution. This table shows some common normalizing transformations:

Type of data and distribution	Normalizing Transform
Count (C comes from Poisson distribution)	$\sqrt{C}$
Proportion (P comes from binomial distribution)	$\text{Arcsine}\sqrt{\text{P}}$
Measurement (M comes from lognormal	Log(M)

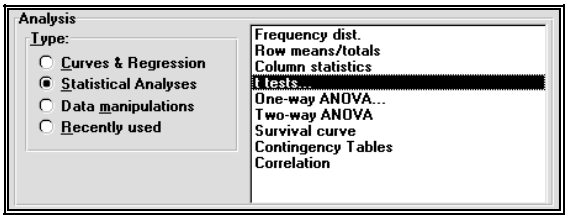
distribution)	
Time or duration (D)	$\frac{1}{D}$

If you know the distribution of your population, transforming the values to create a Gaussian distribution is a good thing to do. A normalizing transform will add power to your analyses. If you cannot transform your data to create a Gaussian distribution, choose a nonparametric test. See "Parametric or non-parametric test?" on page 192.

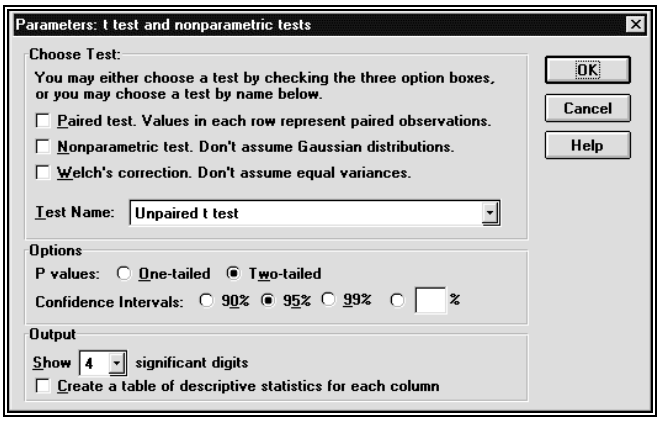
Choosing the analysis

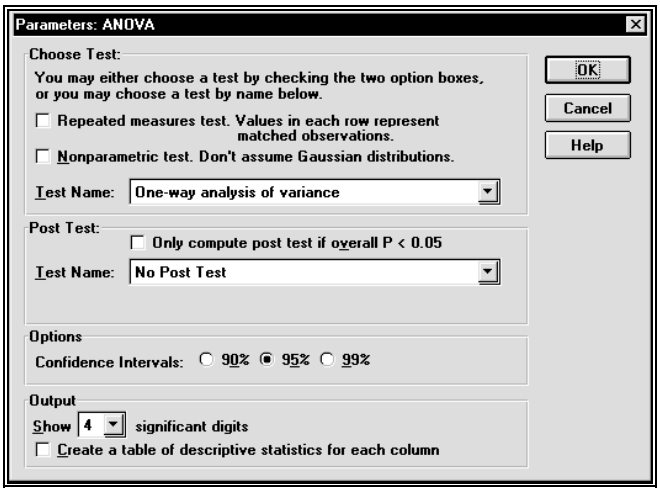
To compare two or more columns:

1. Start from the data or results table you wish to analyze.
2. Press the Analyze button.
3. From the Analyze dialog, select the statistical analyses section.



4. Choose t tests (two groups) or One-way ANOVA (three or more groups).
5. If you don't wish to analyze all columns in the table, select the columns you wish to compare at the bottom of the Analyze dialog.
6. Press OK to bring up the Parameters dialog. The dialog will be different depending on whether you are comparing two columns (t tests) or three or more (ANOVA).





7. Select the name of the test you want from the drop down list.

   or

   Check any or all of the option boxes, and Prism will choose the test for you.

If you are comparing two groups, Prism uses this scheme.

Test	Paired	Non-parametric	Unequal variances
Unpaired t test	No	No	No
Welch's t test	No	No	Yes

Paired t test	Yes	No	N/A
Mann-Whitney test	No	Yes	N/A
Wilcoxon test	Yes	Yes	N/A

If you are comparing three or more groups, Prism uses this scheme:

Test	Paired	Nonparametric
Ordinary one-way ANOVA	No	No
Repeated measures one-way ANOVA	Yes	No
Kruskal-Wallis test	No	Yes
Friedman test	Yes	Yes

The rest of this chapter discusses the decisions you need to make to select the right test.

Paired and repeated measures tests

When choosing a test, you need to decide whether to use a paired test (called a repeated measures test when you compare three or more groups).

You should choose a paired or repeated measures test when the experiment used paired or matched subjects. Here are some examples:

Two groups	Three or more groups
You measure a variable in each subject before and after an intervention.	You measure a variable in each subject before, during and after an intervention.
You recruit subjects as pairs, matched for variables such as age, ethnic group and disease severity. One of the pair gets one treatment, the other gets an alternative treatment.	You recruit subjects as matched sets. Each subject in the set has the same age, diagnosis and other relevant variables. One of the set gets treatment A, another gets treatment B, another gets treatment C, etc.
You run a laboratory experiment several times, each time with a control and treated preparation handled in parallel.	You run a laboratory experiment several times, each time with a control and several treated preparations handled in parallel.
You measure a variable in twins, or	You measure a variable in triplets, or

child/parent pairs.	grandparent/parent/child groups.

More generally, you should select a paired or repeated measures test whenever you expect a value in one group to be closer to a *particular* value in the other group than to a *randomly selected* value in the other group.

Ideally, the decision about repeated measures analyses should be made before the data are collected. Certainly the matching should not be based on the variable you are comparing. If you are comparing blood pressures in two groups, it is OK to match based on age or zip code, but it is not OK to match based on blood pressure.

Parametric or nonparametric test?

The Gaussian (normal, bell-shaped) distribution plays an important role in statistical analyses. Many common tests, including the t test, are based on the assumption that your data are sampled from a population that follows a Gaussian bell-shaped distribution.

If you cannot make that assumption, alternative tests, known as nonparametric tests, can be used instead. The nonparametric tests make fewer assumptions about the distribution of the data, but are less powerful (especially with small samples).

Situations where it is easy to choose a parametric or nonparametric test

You should definitely choose a parametric test (t test or ANOVA) when:

- You are sure that your data are sampled from a population that follows a Gaussian distribution, at least approximately.

You should definitely choose a nonparametric test in these circumstances:

- The outcome variable is a rank or score (i.e. Apgar score). Clearly the population cannot be Gaussian in these cases.
- A few values are "off scale", too high or too low to measure. Even if the population is Gaussian, it is impossible to analyze these data with a t test or ANOVA. Using a nonparametric test with these data is easy. Assign values too low to measure an arbitrary low value, and values too high to measure an arbitrary high value. Since the nonparametric tests only consider the relative ranks of the values, it won't matter that you didn't know a few values exactly.
- You are sure that the population is far from Gaussian. Before choosing a nonparametric test, consider transforming the data (i.e. logarithms, reciprocals). Sometimes a simple transform will convert nongaussian

data to a Gaussian distribution. See "Advantages of transforming the data first" on page 188.

Situations where it is difficult to choose between parametric and nonparametric tests

Unfortunately, the rules of the previous section are rarely helpful. Usually, it is difficult to know whether or not to use a nonparametric test. When deciding, consider these points:

- If you have many (hundreds) of values, a glance at a frequency distribution histogram will tell you whether the data approximate a Gaussian distribution.
- Think about previous experiments that measured the same variable. The assumption is about the distribution of the overall population of values, not just the sample you have obtained in this particular experiment.
- Consider the source of the scatter. When variability is due to the sum of numerous independent sources, with no one source dominating, you expect a Gaussian distribution.
- Look at the results of normality testing (see "Testing whether the distribution is Gaussian " on page 183). Normality tests are of limited usefulness if you have less than a hundred or so values. If the P value of the normality test is low, you have strong evidence that your population is not Gaussian. If the P value is high, don't conclude that the population is Gaussian unless your sample is large. Normality tests have little power to detect nongaussian populations with small samples.

Why sample size matters when choosing between parametric tests and nonparametric tests

If your sample is large, the decision makes little difference.

- ANOVA and t tests are robust to violations of the Gaussian assumption with large data sets. If the population is not Gaussian and you choose a parametric test, the P value will be nearly correct if your sample is large enough. The difficulty is that it is hard to say exactly how large is large enough, as it depends on the shape of the nongaussian distribution. Unless the distribution is really weird, you are probably safe using a t test or ANOVA if you have more than a dozen or so values in each group.
- Nonparametric tests are quite powerful with large samples, even if the population is Gaussian. With large samples drawn from Gaussian popu-

lations, the nonparametric tests need only about 5% more subjects to have the same power as the parametric tests. This rule is true only if the sample size is very large.

- With large samples, it is easy to tell whether the Gaussian assumption makes sense. Either look at a frequency distribution or test for normality (see "Testing whether the distribution is Gaussian " on page 183).

If your samples are small, it makes a big difference which test you choose.

- The P values from ANOVA or a t test are misleading if the samples are small and the data are sampled from nongaussian populations.
- Nonparametric tests lack statistical power with small samples. If your data really do come from a Gaussian distribution, the P value from the nonparametric test will be too high. With very small samples, the nonparametric test always produce a P value greater than 0.05. For example, the P value from a Mann-Whitney test (unpaired nonparametric test that compares two groups) is always greater than 0.05 (two tail) with seven or fewer values (total in both groups).
- With small samples, it is impossible to tell whether the Gaussian assumption is valid. Small samples simply don't contain enough information to let you make inferences about the shape of the distribution in the entire population. Formal normality tests don't help – you simply don't have enough information.

This table summarizes the difference between small and large samples.

	Small samples	Large samples
Normality test	Not very useful. Little power to discriminate between Gaussian and nongaussian populations.	Useful.
Parametric tests	Not robust. If the population is not Gaussian, the P value may be misleading	Robust. P value will be nearly correct even if population is not Gaussian.
Nonparametric test	Not powerful.	Powerful. Almost as powerful as the parametric tests with Gaussian data.

With small data sets, it is impossible to tell whether the population is Gaussian but in matters a lot, because the parametric tests are not robust and the nonparametric tests are not powerful. The best way to choose between parametric and nonparametric tests is to look at data from other experiments (your own or in the literature) measuring the same variable.

Assume equal variances?

The unpaired t test assumes that the two populations have the same variance (and thus the same standard deviation). A modification of the t test (developed by Welch) can be used when you are unwilling to make that assumption. Check the box for Welch's correction if you want this test.

This choice is only available for the unpaired t test. With Welch's t test, the degrees of freedom are calculated from a complicated equation and the number is not obviously related to sample size.

One-way ANOVA also assumes that the populations have equal variances. Modifications of ANOVA that don't make this assumption have been developed, but are rarely used and are not currently available in Prism.

Welch's t test is used rarely. Don't select it without good reason.

One- or two-tail P value?

If you are comparing two groups, you need to decide whether you want Prism to calculate a one-tail or two-tail P value. To understand the difference, see "One- vs. two-tail P values" on page 177.

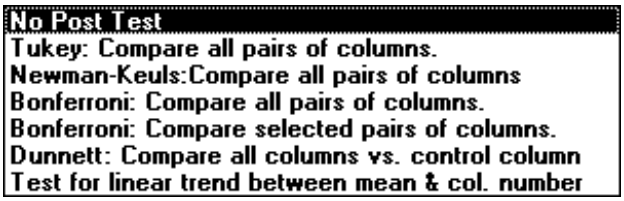
You should only choose a one-tail P value when you believe the following:

- Before collecting any data, you can predict which group will have the larger mean (if the means are in fact different).
- If the other group ends up with the larger mean, then you should be willing to attribute that difference to chance, no matter how large the difference.

Since those conditions are rarely met (and for other reasons), two-tail P values are usually more appropriate.

Which post test?

If you are comparing three or more groups, you may pick a post test to compare pairs of group means. Prism offers these choices:



Choosing an appropriate post test is not straightforward, and different statistics texts make different recommendations.

Ask yourself these questions:

1. Are the columns arranged in a natural order, perhaps time points or dose? If so, consider selecting the post test for linear trend.
2. Is one column the control group to which all other groups should be compared? If so, select Dunnett's test.
3. Based on experimental design, does it make sense to only compare certain pairs of columns? If so, select the Bonferroni test for selected pairs of columns, and select those pairs based on experimental design. It is not fair to first look at the results and then choose which columns you wish to compare. The decision must be based on experimental design (not results), and ideally should be made before the data are collected.

If you answered No to all three questions, you probably want to compare all pairs of columns. Prism offers three methods to do so – the methods of Bonferroni, Tukey, and Newman-Keuls .

The only advantage of the Bonferroni method is that it is well known and easy to understand. Its disadvantage is that it is too conservative, leading to P values that are too high and confidence intervals that are too wide. This is a minor concern when you compare only a few columns, but is a major problem when you have lots of columns. Don't use the Bonferroni test with more than five groups.

Choosing between the Tukey and Newman-Keuls test is not straightforward, and there appears to be no real consensus. The two methods are related, and the rationale for the differences is subtle. The methods are identical when comparing the largest group mean with the smallest. For other comparisons, the Newman-Keuls test yields lower P values. The problem is that it is difficult to articulate exactly what null hypotheses the P values are testing.

Interpreting t tests and non-parametric comparisons

Unpaired t test

P value

The unpaired t test compares two groups, based on the assumption that the two populations are Gaussian. The P value answers this question: If the populations really have the same mean, what is the chance that random sampling would result in means as far apart (or more so) than observed in this experiment?

If you chose a one-tail P value, you must have predicted which group would have the larger mean before collecting any data. Prism does not ask you to record this prediction, but assumes that it is correct. If your prediction was wrong, then ignore the P value reported by Prism and state that $P > 0.50$.

See "P values" on page 174.

t ratio

The t ratio is calculated by this equation:

$$t = \frac{\text{Sample Mean 1 - Sample Mean 2}}{\text{Standard Error of the difference}}$$

The numerator is the difference between the sample means. The denominator is the standard error of that difference, calculated by pooling the SEMs of the two groups. The P value is calculated from t and the size of the groups. If the difference is large compared to the SE of the difference, then the t ratio is large (or is a large negative number), and the P values is small.

CI for difference between mean

Prism reports the confidence interval for the difference between means. You set the degree of confidence (usually 95%) in the parameters dialog. If you accept the assumptions of the analysis, you can be 95% sure that the confidence interval includes the true difference.

R^2 value

R^2. This is the fraction of the overall variance (combining the groups) attributable to the difference between the group means. It compares the size of the difference between group means with the amount of scatter within the groups. A large value means that a large fraction of the variation is due to the treatment that defines the groups. (Note: Few other programs calculate this value following a t test.)

F test to compare variances

Prism tests whether the variances of the two groups are the same by calculating an F test. It reports the value of F and its degrees of freedom and the P value that answers this question: If the two populations really have the same variance, what is the chance that you'd randomly select samples whose variances are as different (or more different) than observed in your experiment.

If the P value is small, you first have to decide whether you wish to conclude that the variances of the two populations are different. Obviously the F test is based only on the values in this one experiment. You should base your conclusion, if possible, on all the information you know about the variable you are measuring, including data from other experiments.

If previous data make you conclude that the variances are really equal overall, then you should ignore the F test and interpret the t test results as usual.

If you conclude that the two populations have different variances, you have three choices:

- Conclude that the two populations are different or that the treatment had an effect. In some experimental contexts, the finding of different variances is as important as the finding of different means. If the variances are truly different, then the populations are different regardless of what the t test concludes about differences between the means. This may be the most important conclusion from the experiment.
- Transform the data to equalize the variances, then rerun the t test. Consider converting all values to their logarithms or reciprocals.
- Rerun the t test without assuming equal variances. Welch's modified t test does not make this assumption.

Note that the variance is the standard deviation squared. So the test of equal variances also tests whether the standard deviations are equal

Assumptions of unpaired t test

The unpaired t test assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each subject (or each experimental unit) was selected independently. See "Assumption of independence" on page 173.
- The populations follow a Gaussian distribution. This assumption is not too important if the samples are large and the sample sizes are equal.
- The two populations have the same standard deviation. This assumption does not apply to Welch's t test.

Paired t test

P value

The paired t test compares two paired groups. It calculates the difference between each set of pairs, and analyzes that list of differences based on the assumption that the differences in the entire population follow a Gaussian distribution.

The P value answers this question: If the treatment is really ineffective so the mean difference is really zero in the overall population, what is the chance that random sampling would result in a mean difference as far from zero (or further) than observed in this experiment?

If you chose a one-tail P value, you must have predicted which group would have the larger mean before collecting any data. Prism does not ask you to record this prediction, but assumes that it is correct. If your prediction was wrong, then ignore the P value reported by Prism and state that $P > 0.50$.

See "P values" on page 174.

t ratio

First the test calculates the difference between each set of pairs, keeping track of sign. If the value in column B is larger, then the difference is positive. If the value in column A is larger, then the difference is negative. The t ratio is the mean of these differences divided by the standard error of the differences. If the t ratio is large (or is a large negative number), the P value will be small.

CI for difference between means

Prism reports the confidence interval for the difference between means. You set the degree of confidence (usually 95%) in the parameters dialog. If you accept the assumptions of the analysis, you can be 95% sure that the confidence interval includes the true difference.

Test for adequate pairing

Prism tests the effectiveness of the pairing by calculating the Pearson correlation coefficient, r. Prism also reports a P value that answers this question: If the two groups really are not correlated at all, what is the chance that randomly selected subjects would have a correlation coefficient as large (or larger) than observed in your experiment (the P value is one-tail, as you are not interested in the possibility of observing a strong negative correlation).

Use this table to interpret the results.

r	P value	Conclusion
Positive	Small	This is what you expect. The pairing was effective.
Positive	Large	The pairing was not particularly effective. Consider revising your experimental procedures or using an unpaired test.
Negative		The pairing was counter productive! You expect the values of the pairs to move together – if one is higher, so is the other. Here the opposite is true – if one has a higher value, the other has a lower value. You should definitely review your procedures. These data are strange.

Assumptions of paired t test

The paired t test assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each pair was selected independently. See "Assumption of independence" on page 173.
- The population of differences follows a Gaussian distribution. This assumption is not too important if the samples are large.
- The pairing was based on information available before the experiment was begun. See "Paired and repeated measures tests" on page 191.

Mann-Whitney test

P value

The Mann-Whitney test, also called the rank sum test, is a nonparametric test that compares two unpaired groups. The P value answers this question: If the populations really have the same median, what is the chance that random sampling would result in medians as far apart (or more so) than observed in this experiment?

If you chose a one-tail P value, you must have predicted which group would have the larger mean before collecting any data. Prism does not ask you to record this prediction, but assumes that it is correct. If your prediction was wrong, then ignore the P value reported by Prism and state that $P > 0.50$.

See "P values" on page 174.

If your samples are small, Prism calculates an exact P value. If your samples are large, it approximates the P value from a Gaussian approximation. The term Gaussian has to do with the distribution of sum of ranks, and does not imply that your data need to follow a Gaussian distribution.

Intermediate results

To perform the Mann-Whitney test, Prism first ranks all the values from low to high, paying no attention to which group each value belongs to. If two values are the same, then they both get the average of the two ranks for which they tie. The smallest number gets a rank of 1. The largest number gets a rank of N, where N is the total number of values in the two groups. Prism then sums the ranks in each group, and reports the two sums. If the sums of the ranks are very different, the P value will be small.

Assumptions of Mann-Whitney test

The Mann-Whitney test assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each subject (or experimental unit) was selected independently. See "Assumption of independence" on page 173.
- The populations don't need to follow any particular distribution, but the two distributions should have about the same shape.

Wilcoxon test

P value

The Wilcoxon test is a nonparametric test that compares two paired groups. It calculates the difference between each set of pairs, and analyzes that list of differences. The P value answers this question: If the median difference in the entire population is zero (the treatment is ineffective), what is the chance that random sampling would result in a median as far from zero (or further) as observed in this experiment?

If you chose a one-tail P value, you must have predicted which group would have the larger median before collecting any data. Prism does not ask you to record this prediction, but assumes that it is correct. If your prediction was wrong, then ignore the P value reported by Prism and state that $P > 0.50$.

See "P values" on page 174.

If your samples are small, Prism calculates an exact P value. If your samples are large, it calculates the P value from a Gaussian approximation. The term Gaussian has to do with the distribution of sum of ranks, and does not imply that your data need to follow a Gaussian distribution.

Intermediate results

In calculating the Wilcoxon test, Prism first computes the differences between each set of pairs. Then it ranks the absolute values of the differences from low to high. Finally, it sums the ranks of the differences where column A was higher (positive ranks) and the sum of the ranks where column B was higher (it calls these negative ranks), and reports these two sums. If the two sums of ranks are very different, the P value will be small.

Test for effective pairing

Prism tests the effectiveness of the pairing by calculating the Spearman nonparametric correlation coefficient, r_s. Prism also reports a P value that answers this question: If the two groups really are not correlated at all, what is the chance that randomly selected subjects would have a correlation coefficient as large (or larger) than observed in your experiment (the P value is one-tail, as you are not interested in the possibility of observing a strong negative correlation).

Use this table to interpret the results.

r_s	P value	Conclusion
Positive	Small	This is what you expect. The pairing was effective.

Positive	Large	The pairing was not particularly effective. Consider revising your experimental procedures or using an unpaired test.
Negative		The pairing was counter productive! You expect the values of the pairs to move together – if one is higher, so is the other. Here the opposite is true – if one has a higher value, the other has a lower value. You should definitely review your procedures. These data are strange.

Assumptions of the Wilcoxon test

The Wilcoxon test assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each pair was selected independently. See "Assumption of independence" on page 173.
- The pairing was based on information available before the experiment was begun. See "Paired and repeated measures tests" on page 191.

Interpreting one-way ANOVA, and nonparametric comparisons

One-way ANOVA

P value

One-way ANOVA compares three or more unmatched groups , based on the assumption that the two populations are Gaussian. The P value answers this question: If the populations really have the same mean, what is the chance that random sampling would result in means as far apart (or more so) than observed in this experiment?

See "P values" on page 174.

R^2 value

This is the fraction of the overall variance (combining the groups) attributable to the difference between the group means. It compares the variability between group means with the variability within the groups. A large value means that a large fraction of the variation is due to the treatment that defines the groups. The R^2 value is calculated from the ANOVA table and equals between group sum-of-squares divided by the total sum-of-squares. (Note: Few other programs calculate this value following ANOVA.)

Bartlett's test for equal variances

ANOVA is based on the assumption that the populations all have the same variance. If your samples have more than five values, Prism tests this assumption with Bartlett's test. It reports the value of Bartlett's statistic and the P value that answers this question: If the populations really have the same vari-

ance, what is the chance that you'd randomly select samples whose variances are as different (or more different) than observed in your experiment.

Bartlett's test is very sensitive to deviations from a Gaussian distribution – more sensitive than is ANOVA itself. A low P value from Bartlett's test may be due to data that are not Gaussian, rather than due to unequal variances.

If the P value is small, you have to decide whether you wish to conclude that the variances of the two populations are different. Obviously Bartlett's test is based only on the values in this one experiment. You should base your conclusion, if possible, on all the information you know about the variable you are measuring, including data from other experiments.

If you conclude that the variances are really equal overall, then you should ignore Bartlett's test and interpret the ANOVA results as usual.

If you conclude that the populations have different variances, you have two choices:

- Conclude that the populations are different or that the treatment had an effect. In some experimental contexts, the finding of different variances is as important as the finding of different means. If the variances are truly different, then the populations are different, regardless of what ANOVA concludes about differences between the means. This may be the most important conclusion from the experiment.
- Transform the data to equalize the variances, then rerun ANOVA. Consider converting all values to their logarithms or reciprocals.

Note that the variance is the standard deviation squared. So the test of equal variances also tests whether the standard deviations are equal.

ANOVA table

The P value is calculated from the ANOVA table. The key idea is that variability among the values can be partitioned into variability between group means and variability within the groups. Variability within groups is quantified as the sum of the squares of the differences between each value and its group mean. This is the residual sum-of-squares. Total variability is quantified as the sum of the squares of the differences between each value and the grand mean. This is the total sum-of-squares. The variability between group means is calculated as the total sum-of-squares minus the residual sum-of-squares. This is called the between groups sum-of-squares.

Even if the null hypothesis is true, you expect values to be closer (on average) to their group means than to the grand mean. The calculation of the degrees of freedom and mean square account for this. See a statistics book for detail.

The end result is the F ratio. If the null hypothesis is true, you expect F to have a value close to 1.0. If F is large, the P value will be small.

Assumptions of one-way ANOVA

One-way ANOVA assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each subject (or each experimental unit) was selected independently. See "Assumption of independence" on page 173.
- The populations follow a Gaussian distribution. This assumption is not too important if the samples are large and the sample sizes are equal.
- The populations have the same standard deviation.

Repeated measures ANOVA

P value

One-way ANOVA compares three or more matched groups , based on the assumption that the differences between matched values are Gaussian. The P value answers this question: If the populations really have the same mean, what is the chance that random sampling would result in means as far apart (or more so) than observed in this experiment?

See "P values" on page 174.

R^2 value

This is the fraction of the overall variance (combining the groups) attributable to the difference between the group means, after adjusting for matching. It compares the size of the difference between group means with the amount of scatter within the groups. A large value means that a large fraction of the variation is due to the treatment that defines the groups. The R^2 value is calculated from the ANOVA table as between group sum-of-squares divided by the sum of between group sum-of-squares and residual sum-of-squares. There are three components to the total sum-of-squares. After subtracting the portion attributable to matching, R^2 is the fraction of the total sum-of-squares that is attributable to differences between groups.(Note: Few other programs calculate this value following ANOVA.)

Was the matching effective?

A repeated measures experimental design can be very powerful, as it controls for factors that cause variability between subjects. If the matching is effective,

the repeated measures test will yield a smaller P value than ordinary ANOVA. The repeated measures test is more powerful because it separates between-subject variability from within-subject variability. If the pairing is ineffective, however, the repeated measures test can be less effective because it has fewer degrees of freedom.

Prism tests whether the matching was effective and reports a P value that tests the null hypothesis that the population row means are all equal. If this P value is low, you can conclude that the matching is effective. If the P value is high, you can conclude that the matching was not effective and should consider using ordinary ANOVA rather than repeated measures ANOVA.

ANOVA table

The P value is calculated from the ANOVA table. With repeated measures ANOVA, there are three sources of variability: between columns (treatments), between rows (individuals) and random. The ANOVA table partitions the total sum-of-squares into those three components. It then adjusts for the number of groups and number of subjects (expressed as degrees of freedom) to compute two F ratios. One F ratio tests the null hypothesis that the population means are identical. The other tests the null hypothesis that the row means are identical (this is the test for effective matching). In both cases, the F ratio is expected to be near 1.0 if the null hypotheses are true. If F is large, the P value will be small.

Assumptions of repeated measures ANOVA

Repeated measures ANOVA assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each set of values (each row) was selected independently. See "Assumption of independence" on page 173.
- The random component of the variable follows a Gaussian distribution. This assumption is not too important if the samples are large.
- The matching was based on information available before the experiment was begun. See "Paired and repeated measures tests" on page 191.
- The correlation between every pair of columns is about the same. You will violate this assumption, for example, if two of the columns represent control measurements, and the other columns represent measurements during treatment. You expect the values in the two control columns to correlate more closely than do values in any other pair of columns. If you violate this assumption, the P value will be too low. For

more information on this assumption, look up "compound symmetry" in an advanced statistics book.

Kruskal-Wallis test

P value

The Kruskal-Wallis test is a nonparametric test to compare three or more unpaired groups. The P value answers this question: If the populations really have the same median, what is the chance that random sampling would result in medians as far apart (or more so) than observed in this experiment? For more details, see "P values" on page 174.

If your samples are small, Prism calculates an exact P value. If your samples are large, it calculates the P value from a Gaussian approximation. The term Gaussian has to do with the distribution of sum of ranks, and does not imply that your data need to follow a Gaussian distribution. With medium size samples, Prism can take a long time to calculate the exact P value. You can interrupt the calculations if an approximate P value is close enough.

Kruskal-Wallis statistic

The Kruskal-Wallis test first ranks all the values from low to high, regardless of group. It then sums the ranks in each group. If the sums are very different, the P value will be small. The program reports the Kruskal-Wallis statistic, which is computed from the sums of the ranks and the sample sizes.

Assumptions of the Kruskal-Wallis test

The Kruskal-Wallis test assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each subject (or each experimental unit) was selected independently. See "Assumption of independence" on page 173.
- The populations don't need to follow a Gaussian distribution, but the populations should have distributions of about the same shape.

Friedman test

P value

The Friedman test is a nonparametric test that compares three or more paired groups. The P value answers this question: If the populations really have the

same median, what is the chance that random sampling would result in medians as far apart (or more so) than observed in this experiment? For more details, see "P values" on page 174.

If your samples are small, Prism calculates an exact P value. If your samples are large, it calculates the P value from a Gaussian approximation. The term Gaussian has to do with the distribution of sum of ranks, and does not imply that your data need to follow a Gaussian distribution. With medium size samples, Prism can take a long time to calculate the exact P value. You can interrupt the calculations if an approximate P value is close enough.

Friedman statistic

The Friedman test first ranks the values in each matched set (each row) from low to high. Each row is ranked separately. It then sums the ranks in each group (column). If the sums are very different, the P value will be small. Prism reports the value of the Friedman statistic, which is calculated from the sums of ranks and the sample sizes.

Assumptions of the Friedman test

The Friedman test assumes:

- The data are representative of a population. See "Population vs. samples" on page 172.
- Each set of data (each row) was selected independently. See "Assumption of independence" on page 173.
- The matching was based on information available before the experiment was begun. See "Paired and repeated measures tests" on page 191.

Post tests

The main analysis of variance calculations compare all the groups at once. The post tests compare pairs of groups.

When interpreting the P values for the post tests, consider the entire group of tests as one family of comparisons. If the null hypothesis is true – all population means are equal – then there is only a 5% chance that you will randomly choose data such that any one or more of the post tests will yield a P values less than 0.05 by coincidence. The 5% chance does not apply to EACH comparison but to the ENTIRE family of comparisons.

With some post tests, Prism also presents a confidence interval for the difference between means for pair of columns. The degree of confidence (usually

95%) applies to the entire family. You can be 95% sure that ALL of the intervals contain the true difference between population means. There is only a 5% chance that any one (or more) of the intervals does not include the true difference between population means.

The post test for trend is different. It tests whether there is a correlation between column mean and column number. The P value answers this question: If there is no linear relationship between column mean and column number, what is the chance that random sampling would result in such a large (or larger) correlation.

Two-way analysis of variance

Introduction to two-way ANOVA

What questions does two-way ANOVA answer?

Two-way ANOVA, also called two factor ANOVA, determines how a response is affected by two different categorical variables. For example, you might measure a response to three different drugs in both men and women.

Two-way ANOVA simultaneously asks three questions:

1. Does the first variable systematically affect the results? In our example: Are the mean responses the same for all three drugs?
2. Does the second variable systematically affect the results? In our example: Are the mean responses the same for men and women?
3. Is there interaction between the two variables? In our example: Does each drug have the same effect for men as it does for women?

The term variance refers to the method used, not to the question you are trying to ask. ANOVA tests for differences between group means, not differences between group variances.

Entering data for two-way ANOVA

To enter data, follow these rules:

- Enter data for two or more data sets (columns). In the example, there would be one data set for each drug.
- Enter data for two or more rows. In the example, there would be one row for men and one for women.
- Enter single values, replicate values, or mean, N and SD (or SEM).

- You may leave some replicates blank, so long as you enter at least one value in each row for each data set. If you enter mean, SD/SEM and N, it is OK if the N values are not all the same.
- Enter X values as text, or leave X blank. The ANOVA calculations ignore the X values completely, and even ignore the order of the rows.

	X Values	A		B		C	
	X Labels	Control		Drug A		Drug B	
	X	Y1	Y2	Y1	Y2	Y1	Y2
1	Men	231.0	253.0	354.0		243.0	264.0
2	Women	284.0	271.0	423.0	465.0	290.0	303.0

Does it matter if some groups are bigger than others?

When groups have unequal numbers of replicates, statisticians say that ANOVA data are "unbalanced". Data can be unbalanced either because you designed the experiment with different numbers of replicates or because some data points were lost during the experiment.

ANOVA (one- and two-way) is based on the assumptions that the values in each group are sampled from a Gaussian population and that all populations have the same standard deviations. With balanced data, ANOVA is fairly robust to violations of these assumptions — the P values will be reasonably accurate even if the assumptions have been violated. With unbalanced data, ANOVA is much less robust. If the data are unbalanced and the assumptions are not true, the ANOVA results may be misleading. See "Missing values in ANOVA" on page 221.

While you may not be able to avoid an occasional missing data point, mildly unbalanced data shouldn't be a problem. However, since it is difficult to be sure of the validity of the ANOVA assumptions, you should avoid designing experiments with different numbers of replicates in different groups (unless you have consulted with a statistician).

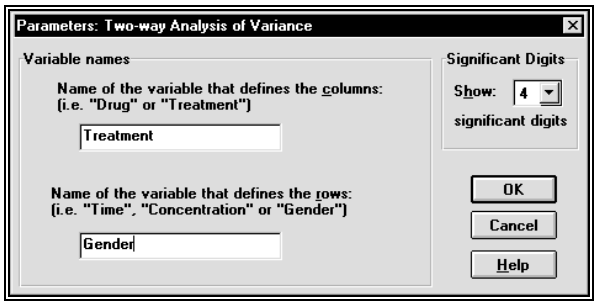
Does it matter if the data are entered as mean, SD (or SEM) and N rather than as individual values?

With ordinary one-way ANOVA, the results will be identical with raw data or summarized data.

With two-way ANOVA, the answer is "it depends". If your data are balanced, the results will be the same with either form of data input. If your data are unbalanced, Prism calculates the results differently depending on how you entered the data. If you entered raw data, Prism analyzes the data with the method most statisticians recommend (it performs multiple regression and reports the P values calculated from "Type III" sum-of-squares). If you entered

mean, SD (or SEM) and N, it is not possible to perform multiple regression, so Prism uses a simpler method (analysis of "unweighted means"). In some special cases, this simpler method gives exactly the same results as obtained by multiple regression. In most cases, however, the results will only be approximately correct. When data are unbalanced, you should enter individual replicates whenever possible.

Parameters dialog for two-way ANOVA



The Parameters dialog is very simple. Enter two names to make the output more clear. If you don't enter names, Prism will use the generic names "Column factor" and "Row factor".

Interpreting the results of two-way ANOVA

Because two-way ANOVA simultaneously addresses three questions, the results can be difficult to interpret.

How two-way ANOVA works

Two-way ANOVA partitions the overall variance of the outcome variable into four parts:

- Some variance is "explained" by systematic differences between columns. In the example, some variance is "explained" by differences between drugs.
- Some variance is "explained" by systematic differences between rows. In the example, some variance is "explained" by differences between men and women.
- Some variance is "explained" by an interaction between rows and columns. This means that some variance is "explained" because the difference between rows is not the same for all columns, and the difference between columns is not the same for all rows. In the example, some

variance is "explained" because the differences between men and women is not consistent for all drugs.

- The rest of the variance is due to scatter within the groups not related to systematic differences between rows and columns. This is called residual variance or error variance.

The ANOVA calculations compare each of the first three sources of variation with the residual variance. The P value is based on this comparison, adjusting for the number of rows and columns and the number of replicates.

Tests of the three null hypotheses

Two-way ANOVA simultaneously tests for column effects, row effects and interaction. For each, Prism reports the following:

- The fraction of the overall variance that is accounted for by this component. Interpret this value as you would an R^2 value from linear regression. The value is correct for the particular samples you have measured, but overestimates the true fraction for the overall population. If you had missing values, this fraction is an estimate.
- The P value. See the next section.
- A statement of statistical significance. Prism uses the conventional criteria: If P is less than 0.05, then it is deemed "significant". Otherwise the effect is "not significant". Don't over interpret this conclusion. Statistically "significant" results may or may not be scientifically interesting or important.

Interpreting the P values

1. **Interaction.** The null hypothesis is that there is no interaction between columns (data sets) and rows. More precisely, the null hypothesis states that any systematic differences between columns are the same for each row, and that any systematic differences between rows are the same for each column. If columns represent drugs and rows represent gender, then the null hypothesis is that the differences between the drugs are consistent for men and women.

   The P value answers this question: If the null hypothesis is true, what is the chance of randomly sampling subjects and ending up with as much (or more) interaction than you have observed. Often the test of interaction is the most important of the three tests.

   If you entered only a single value for each row/column pair, it is impossible to test for interaction between rows and columns. Instead, Prism *assumes* that there is no interaction, and continues with the other calculations. Depending on your experimental design, this assumption

may or may not make sense. The assumption cannot be tested without replicate values.

Note: If the interaction is statistically significant, it is difficult to interpret the row and column effects. Statisticians often recommend ignoring the tests of row and column effects when there is a significant interaction.

2. **Column factor.** The null hypothesis is that the mean of each column (totally ignoring the rows) is the same in the overall population, and that all differences we see between column means are due to chance. If columns represent different drugs, the null hypothesis is that all the drugs produced the same effect. The P value answers this question: If the null hypothesis is true, what is the chance of randomly choosing subjects and ending up with column means as different (or more so) than you have observed.
3. **Row factor.** The null hypothesis is that the mean of each row (totally ignoring the columns) is the same in the overall population, and that all differences we see between row means are due to chance. If the rows represent gender, the null hypothesis is that the mean response is the same for men and women. The P value answers this question: If the null hypothesis is true, what is the chance of randomly choosing subjects and ending up with row means as different (or more so) than you have observed.

The ANOVA table

The ANOVA table breaks down the overall variability between measurements (expressed as the sum of squares) into four components:

- Interaction between row and column. Variation between rows that is not the same for each column. Or, equivalently, variation between columns that is not the same for each row.
- Systematic variation between columns.
- Systematic variation between rows.
- Residual or error. Variation among replicates not related to systematic differences between rows and columns.

The table also shows the degrees of freedom and mean square for each component, as well as the F ratio. Each F ratio is the ratio of the mean-square value to the residual mean-square. If the null hypothesis is true, the F ratio is likely to be close to 1.0. If the null hypothesis is not true, the F ratio is likely to be greater than 1.0.

See any statistics book for help interpreting this table. Most scientists will skip over the ANOVA table, and focus on the tests of the three null hypotheses.

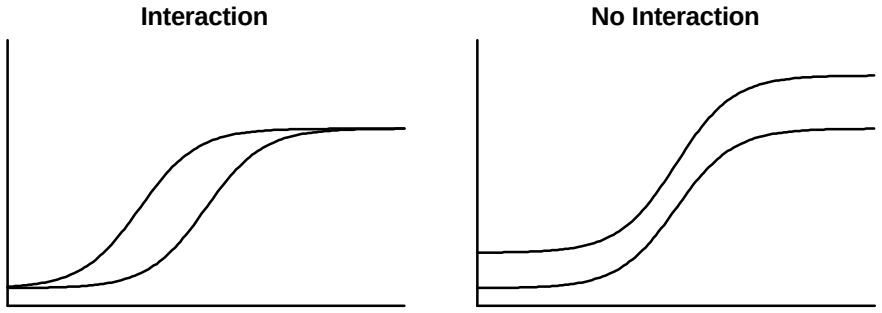
Interpreting two-way ANOVA when comparing curves

When X values represent time or concentration, you will interpret the results of two-way ANOVA somewhat differently. Note, however, that the two-way ANOVA calculations completely ignore the X values.

Interaction

The null hypothesis can be stated this way: The difference between the Y values of the curves is the same at all values of X. More simply, the null hypothesis is that the curves are "parallel".

Beware of the term "parallel". Pharmacologists consider two dose-response curves "parallel" when two drugs have similar effects at very low and very high concentrations, but different (and horizontally parallel) effects at moderate concentrations. This pattern would reject the null hypothesis of no interaction, because the difference between Y values in the middle of the curves is very different than the difference at the ends. For two dose-response curves to show "no interaction", they would have to be vertically displaced so the vertical difference between the curves is the same for all values of X.



Column factor

The null hypothesis can be stated this way: Averaging over all concentrations (or times), the average Y value of each data set is identical. If the different data sets represent different experimental treatments, then the null hypothesis is that none of the experimental treatments affected the outcome.

Note that the calculations completely ignore the X values. If you use two-way ANOVA to compare dose-response curves or kinetic experiments, the calcu-

lations pay no attention to dose or time. You would get exactly the same results if you randomly shuffled the X values, or even if you left the X column entirely blank.

Row factor

The null hypothesis is that the curves are horizontal — that dose (or time) does not systematically alter the response. In most cases, you already know that this null hypothesis is false, or you wouldn't have done the experiment. The P value will almost always be very low, and is rarely of interest.

Assumptions of two-way ANOVA

Two-way ANOVA is based on four assumptions:

1. All subjects must be randomly selected from, or at least representative of, the larger populations. This assumption underlies all statistical analysis.
2. The observations within each sample must be independent of one another. The relationships between all the observations in a group should be the same. You don't have four independent measurements if you measure two animals twice, if the four animals are from two separate litters, or if two animals were assessed in one experiment in January and two others were assessed in another experiment in June.
3. There is no matching or blocking of subjects.
4. The data are sampled from populations that approximate a Gaussian distribution, and the SD of all populations is identical.

The first two assumptions are common to most statistical tests, and are usually not a problem.

The third assumption may not be true for your data. If the subjects are matched in some way, you should perform repeated measures two-way ANOVA (not performed by Prism).

The fourth assumption causes the most concern. Fortunately, ANOVA is fairly robust to departures from this assumption, especially when there are no missing values. If your data are clearly not Gaussian, you may be able to transform the data to make the distribution more Gaussian. For example, if the outcomes are counts, transform to their square roots. If the outcomes are measurements with a skewed distribution, it might help to transform to their logarithm or reciprocal. See "Advantages of transforming the data first" on page 188.

If you really don't think your data are sampled from a Gaussian distribution (and no transform will make the distribution Gaussian), you should consider performing nonparametric two-way ANOVA. Prism does not offer this test.

How Prism calculates two-way ANOVA

Read any advanced statistics book to learn how to calculate two-way ANOVA. There are three complexities.

Model I (fixed effects) vs. Model II (random effects) ANOVA

When calculating two-way ANOVA, you must differentiate between fixed-factor independent variables and random-factor independent variables.

The difference between fixed factors and random factors is best explained through examples. Assume that you are comparing three different species at three different time points. Species is considered to be a fixed factor if you are interested in those three particular species. Species would be a random factor if you were interested in differences between species in general, and randomly selected those three particular species. Time is considered to be a fixed factor if you chose those three time points to span the interval you are interested in. Time would be a random factor if you picked those three time points at random. Since this is not likely, time is considered to be a fixed factor.

When Prism calculates two-way ANOVA, it assumes that both row and column variables are fixed factors. Prism cannot deal with random factors. Since most biological experiments deal with fixed-factor variables, this is rarely a limitation.

When both row and column variables are fixed factors, the analysis is called Model I ANOVA. When both row and column variables are random factors, the analysis is called Model II ANOVA. When one is random and one is fixed, it is termed mixed effects (Model III) ANOVA. Prism only calculates Model I two-way ANOVA.

Entering data as mean, SD (or SEM) and N

With ordinary one-way ANOVA, the results will be identical with raw data or summarized data. Enter whatever format is more convenient for you. However, repeated measures ANOVA requires raw data.

With two-way ANOVA, there is no problem if your data are balanced (same sample size for each condition). The results will be the same (and correct) regardless of whether you enter raw data or mean, SD (or SEM) and N. If your data are unbalanced, Prism calculates the results differently depending on

how you entered the data (see next section). If you entered mean, SD (or SEM) and N, it is not possible to perform the correct calculations, so Prism uses a simpler method (analysis of "unweighted means"). In some special cases, this simpler method gives exactly the same results as obtained by analysis of the raw data. In most cases, however, the results will only be approximately correct. When data are unbalanced, you should avoid entering summarized data but instead enter individual replicates whenever possible.

Missing values in ANOVA

If you entered the same number of replicates everywhere, calculation of two-way ANOVA is straightforward. If some values are missing, the data are said to be unbalanced, and standard ANOVA calculations cannot be performed. Prism calculates unbalanced two-way ANOVA in one of two ways:

If you entered replicate values

If you entered individual replicate values, Prism converts the ANOVA problem to a multiple regression problem, and then displays the results as ANOVA. In other words, Prism uses the General Linear Model to calculate the results. Statisticians agree that this is the best method to analyze unbalanced two-way ANOVA (Prism reports the P values calculated from "Type III" sum-of-squares).

With unbalanced data, there are several ways to calculate the sum-of-squares values in the ANOVA table. Prism performs multiple regression three times — each time presenting columns, rows and interaction to the multiple regression procedure in a different order. Although it calculates each sum-of-squares three times, Prism only displays the sum-of-squares for the factor entered last into the multiple regression equation.

With unbalanced data, the three sum-of-square values reported by Prism are determined from three different runs of multiple regression. The sum of the three may not add up to the total sum-of-squares.

If you entered mean, SD (or SEM) and N

If you entered mean, SD (or SEM) and N, it is not possible to perform multiple regression, so Prism uses a simpler method (analysis of "unweighted means"). This method ignores the differences in sample size, and calculates ANOVA as if all the groups had a sample size equal to the geometric mean of all the sample sizes. In some special cases, this simpler method gives exactly the same results as obtained by multiple regression. In most cases, however, the results will only be approximately correct. If your data are almost balanced (just one or a few missing values), the approximation is a good one. When data are unbalanced, you should enter individual replicates whenever possible.

For more information on analyzing two-way ANOVA, see SA Glantz and BK Slinker, Primer of Applied Regression and Analysis of Variance, McGraw-Hill, 1990 or SE Maxwell and HD Delaney, Designing Experiments and Analyzing Data, Brooks/Cole Publishing Company, 1990.

Note: Prism cannot calculate repeated-measures two-way ANOVA and cannot calculate post tests following two-way ANOVA.

Single replicates

Prism can perform two-way ANOVA even if you have entered only a single replicate for each column/row pair. This kind of information does not provide any information for testing interaction. Instead, Prism assumes that there is no interaction, and only tests for row and column effects. If this assumption is not valid, then the P values for row and column effects won't be meaningful.

Survival curves

Introduction to survival curves

What are survival curves?

Survival curves plot percent survival as a function of time. Prism creates survival curves using the method of Kaplan and Meier, and calculates the 95% CI for fractional survival at any particular time. Prism can also compare two or more survival curves using the log-rank test.

The term "survival" curve is a bit misleading, as "survival" curves can plot time to any well-defined end point such as occlusion of a vascular graft, first metastasis, or rejection of a transplanted kidney. The event does not have to be dire. The event could be restoration of renal function, discharge from a hospital, or graduation. The rest of this chapter is written assuming that the event is death.

The end point must be a one time event. Recurring events should not be analyzed with survival curves.

Censored survival data

It would be very easy to create a survival curve if you knew when each subject died. But you don't. In most studies, some surviving subjects are not followed for the entire span of the curve.

- Some subjects are still alive at the end of the study, but were not followed for the entire span of the curve. Many studies enroll patients over a period of several years. The patients who enroll late are not followed for as many years as patients who enroll early. Imagine a study that enrolls patients between 1985 and 1989, and that ends in 1991. Patient A enrolled in 1989 and is still alive at the end of the study. Even though the study lasted six years, we only know that patient A survived for at least 2 years.
- Some drop out of the study early. Perhaps they moved to a different city or wanted to take a medication disallowed on the protocol. Patient

B enrolled in 1986 but moved to another city (and stopped following the protocol) in 1988. We know that this subject survived at least two years on the protocol, but can't evaluate survival after that.

In either case, you know that the subject survived up to a certain time, but have no useful information about what happened after that. Information about these patients is said to be censored.

Before the censored time, you know these subjects were alive and following the experimental protocol. After they are censored, you can't use any information on the subjects. Either we don't have information beyond the censoring day (because the data wasn't, or can't be, collected) or we have information but can't use it (because the patient no longer was following the experimental protocol).

The word "censor" has a negative ring. It sounds like the subject has done something bad. Not so. It's the data that has been censored, not the subject! Every subject either dies or is censored.

When calculating and comparing survival curves, Prism automatically accounts for censored data.

Entering data for survival curves

Enter time until censoring or death in the X column using any convenient unit. X is usually entered in days. Time zero is not some specified calendar date; rather it is the time that each subject entered the study. In many clinical studies, time zero spans several calendar years as patients are enrolled.

Note: You must enter X as a duration. You cannot enter dates.

Format the data sheet for single Y values (no replicates; no error bars). Enter one code (usually Y=1) into the Y columns for rows where the subject died (or the event occurred) at the time shown in the X column. Enter another code (usually Y=0) into the rows where the subject was censored at that time (see the previous section for a definition of censored).

Note: The term death is used for convenience. Survival curves can be used to plot time to any nonrecurring event. The event does not have to be death.

Each Y column represents a different treatment group. We suggest entering the data as follows: Place the X values for the subjects for the first group at the top of the table with the Y codes in the first Y column. Then place the X values for the second group of subjects beneath those for the first group. Place their Y codes in the second Y column, leaving the first column blank. In this example, rows 1-14 were for group A and 15-28 were for group B.

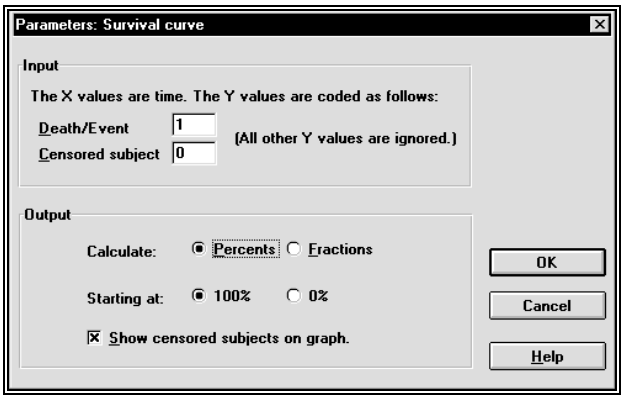
If the treatment groups are intrinsically ordered (perhaps increasing dose) maintain that order when entering data. Make the progression from column A to column B to column C follow the natural order of the treatment groups. See "Logrank test for trend" on page 227.

10	46.0	1	
11	64.0	0	
12	78.0	0	
13	124.0	1	
14	127.0	0	
15	9.0		1
16	26.0		0
17	43.0		1
18	64.0		1

Prism is very flexible about how you enter survival data:

- It is not necessary to enter the X values in order or to sort X values later.
- It is OK to enter Y codes for more than one group on the same row.

Parameters for survival curves



Codes

Ordinarily, Y = 0 indicates a censored subject and Y = 1 indicates a death. You may define other codes (use digits, not letters).

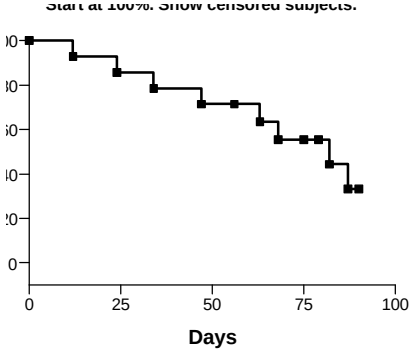
Results

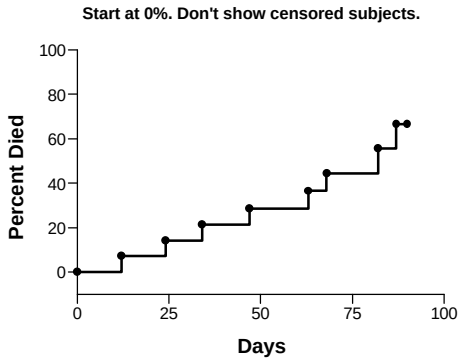
Choose whether you wish to report results as fractions or percents. Also choose whether you want the curve to start at 100% (fractional survival) or at

0% (fractional deaths). If the event is a positive one (i.e. restoration of renal function), you'll probably want to start at 0% so the graph runs uphill.

Censored subjects

Decide whether you wish censored subjects to appear on the survival curves. The analysis will be identical either way; only the appearance of the graph will differ. Censored subjects appear as a data point in a flat part of the curve.





Interpreting survival curves

The fraction (or percent) survival at each time

For each X value (time) Prism shows the fraction still alive (or the fraction already dead, depending on your option). It also shows standard error. This table contains the numbers used to graph survival vs. time. Prism calculates survival fractions using the product limit or Kaplan-Meier method.

On a second output view, Prism displays the number of patients still at risk at each time. As subjects die and are censored, the number still being followed decreases. This table also shows the median survival time for each group.

Logrank test to compare survival curves

If you entered two or more data sets on the table, Prism compares the survival curves and presents the comparison on a third output view.

Prism compares survival curves using the logrank test. If you entered two data sets, the log-rank test is equivalent to the Mantel-Haenszel test. This test generates a P value testing the null hypothesis that the survival curves are identical in the overall populations. In other words, the null hypothesis is that

the treatments did not change survival. The P value answers this question: If the null hypothesis is true, what is the probability of randomly selecting subjects whose survival curves are as different (or more so) than was actually observed?

The P value is two-sided. You may justify using a one-sided P value (also called a one-tail P value) in the following circumstances:

- You predicted which group would survive longer before collecting any data.
- If the other group survived longer, you would attribute the difference to chance, no matter how large the difference is.

Prism does not calculate one-sided P values from surival curves, although you can calculate it yourslef. The one-sided P value is half the two-sided P value. See "One- vs. two-tail P values" on page 177.

Logrank test for trend

If you entered three or more data sets, Prism also calculates the logrank test for trend. This test is only meaningful if the data sets were entered in a logical order, perhaps corresponding to dose or age. If the data sets are not ordered (or not equally spaced), then you should ignore the results of the logrank test for trend. The logrank test for trend calculates a P value testing the null hypothesis that there is no linear trend between column number and median survival.

Assumptions of survival analysis

Your results are not valid unless the following are true:

- Random sample. If your sample is not randomly selected from a larger population, then you must assume that your sample is representative of that population.
- Independent observations. Choosing any one subject in the population should not affect the chance of choosing any other particular subject.
- Consistent entry criteria. Typically, subjects are enrolled over a period of months or years. In these studies, it is important that the starting criteria don't change during the enrollment period. Imagine a cancer survival curve starting from the date that the first metastasis was detected. What would happen if improved diagnostic technology detected metastases earlier? Even with no change in therapy or in the natural history of the disease, survival time will apparently increase (patients die at the same age they otherwise would, but are diagnosed at an earlier age and so live longer with the diagnosis).

- Consistent criteria for defining survival. If the curve is plotting time to death, then there is no ambiguity. If the curve is plotting time to some other event, it is crucial that the event be assessed consistently throughout the study.

  If the curve is plotting deaths due to a particular form of cancer, you need to decide what to do with patients who die of another cause, say an automobile accident. Some investigators count these as deaths, others count them as censored subjects. Both approaches are sensible, but the approach should be decided before the study begins.

- Time of censoring is unrelated to survival. The survival of the censored patients must be identical (on average) to the survival of the remainder. If there are many censored subjects, this assumption is critical to the validity of the results.

  There is no reason to doubt that assumption for patients still alive at the end of the study. Patients who dropped out of the study are a different matter. You must ask why the patients left the study. A survival curve would be misleading, for example, if many patients quit the study because they were too sick to come to clinic, or because they felt too well to take medication.

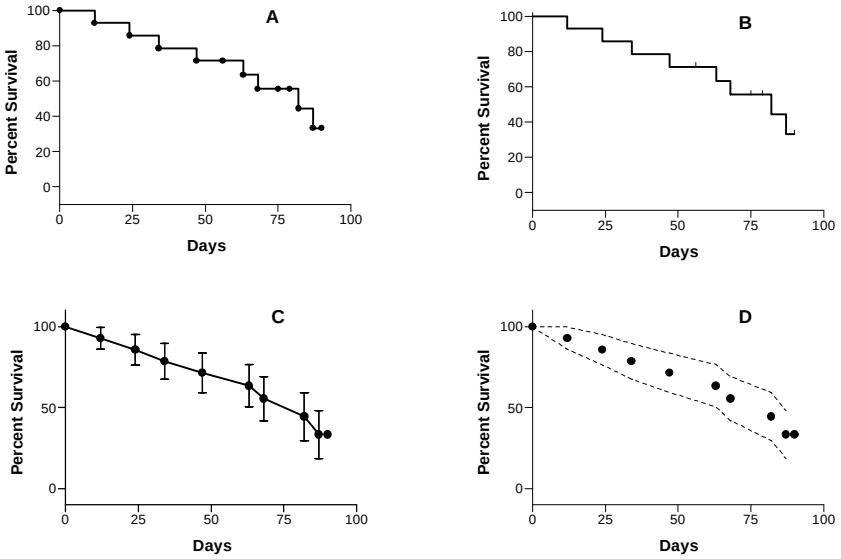
- Average survival does not change during the course of the study. Many survival studies enroll subjects over a period of several years. The analysis is only meaningful if you can assume that average survival of the first few patients is not different than the average survival of the last few subjects. If the nature of the disease or the treatment changes during the study, the results will be difficult to interpret.

Graphing survival curves

Prism automatically creates a graph of survival vs. time. There are several ways you can customize this graph. You have to make these decisions:

- Downhill (start at 100%) or uphill (start at 0%)? Make this decision in the Parameters dialog for survival curves.
- Show percents or fractions? Make this decision in the Parameters dialog for survival curves.
- Show censored subjects? Prism can create data points for deaths and censored subjects or just for deaths. Censored subjects appear as data points in a flat part of the curve. Make this decision in the Parameters dialog for survival curves.
- Which symbol? Choose conventional symbols (i.e. circles or squares) or ticks on the Symbols dialog. The second to the last choice for symbol shape is an upward pointing tick. See example B below.

- Show standard error of survival? Prism calculates the standard error of the fractional survival at the time of each death and can plot these as error bars. However, error bars are unattractive when superimposed on staircase lines. To show the standard error, click on the error bar option box in the Symbols dialog.
- Show as staircase or straight (point-to-point lines)? Survival curves are conventionally shown as staircases, but this makes it difficult to include error bars. Point-to-point lines are used less commonly, but allow for error bars. Choose in the connecting line style list in the Symbols dialog.



In example A, the data points were shown as circles and connected by a staircase. The censored subjects are shown as circles in the horizontal part of the staircase. There are no error bars. This kind of graph is commonly used.

In example B, the data points were shown as ticks (the second to the last choice in the list of symbols). You only see ticks for censored subjects, because the other ticks are superimposed on the staircase. There are no error bars. This is probably the most common appearance for survival curves.

Note: You can t graph error bars with a staircase line because the error bars would be superimposed on the staircase.

In example C, the data are shown as circles, and connected by straight (point-to-point) lines. Error bars show plus and minus one SE.

In example D, the data are shown as circles connected by straight lines. Error bars show plus and minus one SE. The error bars are shown as bands one SE above or below the points. To draw error bands, select the last choice (- - - -) on the list of error bar styles in the Symbols dialog.

Note: Some statisticians think that is misleading to graph survival data with point-to-point lines, because the percent survival in your sample actually follows a staircase pattern.

Contingency tables

What is a contingency table?

Here is an example contingency table. Subjects with HIV infection were divided into two groups and given placebo or AZT. The result was recorded as disease progression or no progression (from New Eng. J. Med. 329:297-303, 1993).

	Disease progression	No progression	Total
AZT	76	399	475
Placebo	129	332	461
Total	205	731	936

Note these features of a contingency table, also called a cross tabulation table.

- The values in the table are the number of subjects. Table with averages, percentages or rates are not contingency tables.
- The two columns are mutually exclusive. A subject can be in one or the other, but not both.
- The two rows are also mutually exclusive.

Contingency tables are used to tabulate the results of four kinds of experiments:

- Experimental. Subjects are randomly divided into groups that become the rows of the table. Each group is given an alternative treatment. The columns of the table represent alternative outcomes. The example above is an experimental study.
- Prospective study. Subjects are recruited based on exposure to risk factors, and then are followed to observe the outcome. The experimenter assigns rows based on selection criteria and assigns columns based on the outcome .

- Cross-sectional study. A large group of subjects is recruited without regard to exposure or outcome. The investigators then divide the subjects into groups based on exposure to a risk factor and compare the prevalence of disease in the groups.
- Retrospective or case-control studies. Subjects are recruited based on the outcome (different columns). The investigators then look back at exposure to possible risk factors (or protective factors) which are shown as different rows. The investigator asks whether there is an association of the risk factor with the disease.

Contingency tables can tabulate the results of some basic science experiments. The rows represent alternative treatments, and the columns tabulate alternative outcomes.

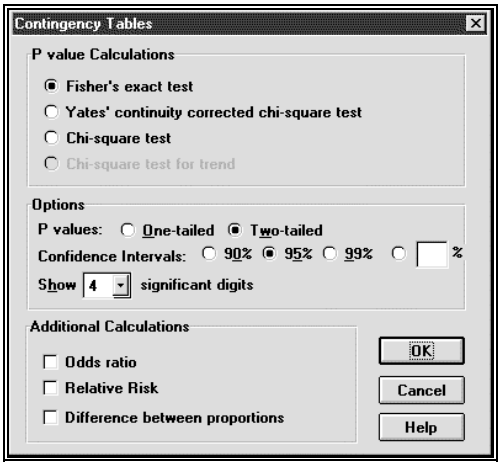
Entering contingency tables into Prism

Prism is not able to cross-tabulate data to create a contingency table. You must enter the contingency table itself.

	X Labels	A	B
	X Labels	Progression	No Progress
	X	Y	Y
1	AZT	76.0	399.0
2	Placebo	129.0	332.0

If your experimental design matched patients and controls, you should not analyze your data with contingency tables. Instead you should use McNemar's test. This test is not offered by Prism, but it is included in the companion program GraphPad StatMate.

Analyzing contingency tables



P value calculation

Prism offers three methods for calculating a P value from tables with two rows and two columns.

- Fisher's exact test. This test calculates a P value that is exactly correct.
- The chi-square test. The only advantage of the chi-square test is that it is easier to calculate by hand, and so is better known.
- Yates' continuity corrected chi-square test. This correction is designed to make the approximate results from a chi-square test more accurate with small samples. Statisticians disagree about whether to use it. If you always use Fisher's exact test, you don't have to think about Yates' correction.

If you have very large numbers (thousands) in your table, Prism will automatically perform the chi-square test even if you select Fisher's test, because the Fisher's test calculations are lengthy with large samples. With large samples, the chi-square test is very accurate and the Yates' continuity correction has negligible effect.

If your table has more than two rows or two columns, Prism always calculates the chi-square test. You have no choices. Extensions to Fisher's exact test have been developed for larger tables, but Prism doesn't offer them.

If your table has two columns and three or more rows, you may also select the chi-square test for trend. This calculation tests whether there is a linear trend between row number and the fraction of subjects in the left column. It

only makes sense when the rows are arranged in a natural order (i.e. age, dose, time) and are equally spaced.

One or two-sided P value?

I recommend that you always choose a two-tail P value, unless you have a good reason to pick a one-tail P value. See "One- vs. two-tail P values" on page 177.

For analysis of contingency tables, the concept of "tails" to P values is not quite appropriate, so we refer to two-sided P values rather than two-tail P value. The distinction is subtle and not worth worrying about.

Relative risk, Odds ratio, and Difference between proportions

There are three ways to quantify the results of a contingency table with two rows and two columns.

- Difference between proportions. In the example, disease progressed in 28% of the placebo-treated patients and in 16% of the AZT-treated subjects. The difference is 28% - 16% = 12%.
- Relative risk. The ratio is 16%/28% = 0.57. A subject treated with AZT has 57% the chance of disease progression as a subject treated with placebo. The word "risk" is not always appropriate. Think of the relative risk as being simply the ratio of proportions.
- Odds ratio. This is a more difficult concept. There isn't much point in calculating an odds ratio for experimental or prospective studies. When analyzing case-control retrospective studies, however, you cannot meaningfully calculate the difference between proportions or the relative risk. The odds ratio is used to summarize the results of these kinds of studies. See a biostatistics or epidemiology book for details.

If your data came from an experimental, cross-sectional or prospective study, you can pick any or all of these. The odds ratio is rarely used with these kind of data.

If your data came from a retrospective case-control study, you should pick only the odds ratio.

Interpreting the results

P value

The P value answers this question: If there is no association between row and column variables in the overall population, what is the chance of observing so much association in an experiment of this size.

In the example, the P value from Fisher's test is less than 0.0001. If there was no association between AZT treatment and the fraction of patients whose disease progresses, there is less than a 0.01% of observing such a strong association (or stronger) in a study of this size.

Confidence intervals

You can calculate the difference between proportions, relative risk, or odds ratio for the particular subjects you happened to sample. You can be 95% sure that the confidence interval contains the true value in the overall population.

Prism approximates the confidence interval of the relative risk and odds ratio using the methods of Woolf and Katz respectively. If any of the four values in the contingency table are zero, Prism adds 0.5 to all values before calculating the relative risk or odds ratio and the confidence intervals.

Assumptions

To interpret the results, you must assume:

- The data are randomly selected from (or are at least representative of) a larger population.
- Each subject was obtained independently.
- No pairing or matching.

Correlation

Introduction to correlation

Correlation is used when you have measured two variables in each subject, and wish to know what degree the two vary together. When the two variables vary together, statisticians say that there is a lot of covariation or correlation. The direction and magnitude of correlation is quantified by the correlation coefficient, r.

Prism calculates the correlation coefficient, r, and its 95% confidence interval. It also calculates a P value that answers this question: If the two variables really aren't related at all in the overall population, what is the chance that you would obtain a correlation coefficient as far from zero as observed in your experiment from randomly selected subjects?

Comparison of correlation and linear regression

The methods of correlation and linear regression are related, and used in similar situations. When deciding which of these methods to use, consider these points:

- The value of r^2 and the P value will be the same with linear regression and correlation.
- It only makes sense to interpret the confidence interval of r when X values were measured, and not controlled. If you control X (time, dose, concentration) you should perform linear regression.
- It only makes sense to interpret linear regression when you can clearly define which variable is X and which is Y. It matters. Linear regression finds the best line that predicts Y from X. It finds the line by minimizing the sum of the square of the vertical distances of the points from the regression line. Therefore it matters which variable is X and which is Y.
- In some situations it makes sense to perform both linear regression and correlation.

Calculating correlation with Prism

To calculate correlation:

1. Enter the X and Y values. Y values should be single values. If you enter replicate values, or mean and SD (or SEM) only the mean values will be considered.
2. Press the Analyze button.
3. Choose Correlation from the Statistical analyses section.
4. Complete the choices on the dialog.

Nonparametric correlation

The standard correlation calculations are based on the assumption that both X and Y values are sampled from populations that follow a Gaussian distribution, at least approximately. With large samples, this assumption is not too important.

If you don't wish to make the Gaussian assumption, select nonparametric (Spearman) correlation instead. Spearman correlation is based on ranking the two variables, and so makes no assumption about the distribution of the values.

One- vs. two-tail P value?

Start with the null hypothesis that the two populations are not correlated and that the observed correlation between the samples is due to chance.

- The two-tail P value answers this question: Assuming the null hypothesis, what is the chance that randomly selected samples would have a correlation coefficient as far from zero as observed in this experiment with either a positive or a negative correlation coefficient?
- To interpret a one-tail P value, you must predict whether the correlation will be positive or negative before collecting any data. The one-tail P value answers this question: Assuming the null hypothesis, what is the chance that randomly selected samples would have a correlation coefficient as far from zero as observed in this experiment and with the sign (positive or negative) you predicted in advance?

A one-tail P value is appropriate only when previous data, physical limitations or common sense tell you that the correlation, if any, can only go in one direction. The issue is not whether you expect a correlation to exist – that is what you are trying to find out with the experiment. The issue is whether you should interpret a positive correlation and a negative correlation the same. See "One- vs. two-tail P values" on page 177.

You should only choose a one-tail P value when you believe the following:

- Before collecting any data, you can state with certainty that there is either no correlation or the correlation will either be positive or negative (you specify which).
- If you data end up showing a correlation in the "wrong" direction, then you should be willing to attribute that correlation to chance, no matter how large it is.

Interpreting correlation results

Meaning of r

The correlation coefficient, r, ranges from -1 to 1.

- If $r = 0$, the two variables do not vary together at all.
- If r is positive, the two variables tend to increase or decrease together.
- If r is negative, one variable increases as the other decreases.
- If $r = 1$ or $r = -1$, a graph of the data forms a perfect straight line.

Perhaps the best way to interpret the value of r is to square it to calculate r^2. Statisticians call the quantity the *coefficient of determination*, but scientists call it *r squared*. It is has a value that ranges from zero to one, and is the fraction of the variance in the two variables that is shared. For example, if $r^2 = 0.59$, then 59% of the variance in X can be explain by (or goes along with) variation in Y. Likewise, 59% of the variance in Y can be explained by (or goes along with) variation in X.

If r is far from zero, there are four possible explanations:

- The X variable helps determine the value of the Y variable.
- The Y variable helps determine the value of the X variable.
- Both X and Y are influenced by another variable.
- X and Y don't really correlate at all, and you just happened to observe such a strong correlation by chance. The P value determines how often this could occur.

Meaning of the P value

The P value answers this question: If the two variables really aren't correlated at all in the overall population, what is the chance that you would obtain a correlation coefficient as far from zero as observed in your experiment from randomly selected subjects?

To interpret a one-tail P value, you must have predicted in advance whether you expect the correlation to be positive (two variables vary together) or negative (one variable goes up as the other goes down). Prism does not ask

you to record this prediction, but calculates the one-tail P value assuming that it was correct. If your prediction was wrong, then ignore the P value reported by Prism and state that $P > 0.50$.

Assumptions of correlation

- Subjects are randomly sampled from, or are at least representative of, a larger population.
- X and Y are paired. You've measured two variables for each subject.
- Subjects are independent. This means that the relationship between all the subjects should be the same. You would violate this assumption if you choose half the subjects from one population and half from another. You would also violate this assumption if you made triplicate measurements on 12 experimental preparations and entered the data as 36 XY pairs (you should enter 12 XY pairs, where each is the average of the replicates).
- X and Y must be measured independently. The calculations are not valid if X and Y are intertwined. You'd violate this assumption if you correlate midterm exam scores with overall course score, as the midterm score is one of the components of the overall score.
- X values were measured, not controlled. If you controlled X values (i.e. concentration, dose or time) you should calculate linear regression rather than correlation.
- All covariation must be linear. The correlation would not be meaningful if Y increases as X increases up to a point, and then Y decreases as X increases further.
- Gaussian distributions. The X and Y values must be sampled from populations that follow Gaussian distributions. Spearman nonparametric correlation does not make this assumption.

Linear regression

Introduction to linear regression

What is linear regression for?

Linear regression is used to analyze the relationship between two variables, which we will label X and Y. For each subject, you know both X and Y and you want to find the best straight line through the data. There are two reasons you might want to do this:

- Because the slope and/or intercept have a scientific meaning.
- To create a standard curve to find new values of X from Y, or Y from X.

What can Prism do?

Prism determines and graphs the best-fit linear regression line, including 95% confidence interval bands. You may also force the line through a particular point (usually the origin), calculate residuals, calculate a runs test, or compare the slopes and intercepts of two or more regression lines.

Calculating linear regression with Prism

Prism makes it very easy to calculate and graph linear regression.

To calculate linear regression:

1. Start either from a data table or from a graph of your data.
2. Press Analyze and select Linear regression.
3. Make selections on the Parameters dialog. See the next section.
4. If you started from a graph, Prism will superimpose the linear regression line. Click on Results to see numerical results.
5. If you started from a data table, Prism will show you the numerical results first. Click on Graphs to see a graph.

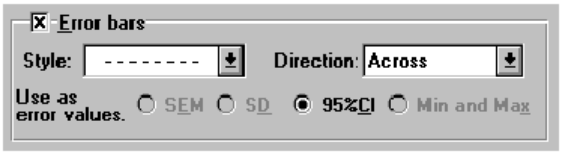
After calculating linear regression, Prism automatically graphs the regression line with the original data. Prism defines the regression line as a table of XY values. You can see this table as one of the output views of linear regression, although you will rarely find it of interest. Depending on the parameters you set, this table may also include the 95% CI for the regression line.

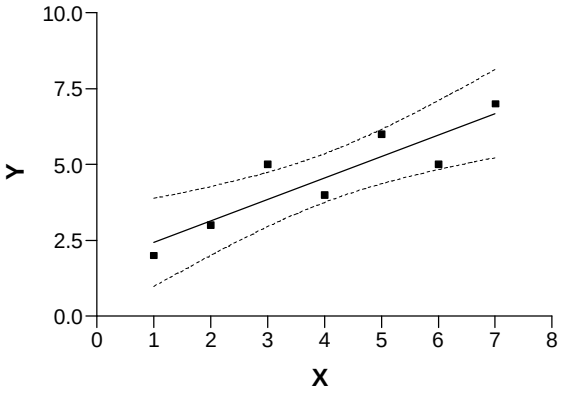
To automatically graph the curve with the data points, Prism adds the table of XY values to the list of data sets included on the graph. The original data table and the table containing the regression line are separate tables included on one graph.

When you customize the graph using the Symbols dialog, note that the regression lines are separate data sets from the original data. Graph the regression lines using no symbols and with straight lines connecting the "points". Graph the data as symbols with no connecting straight lines.

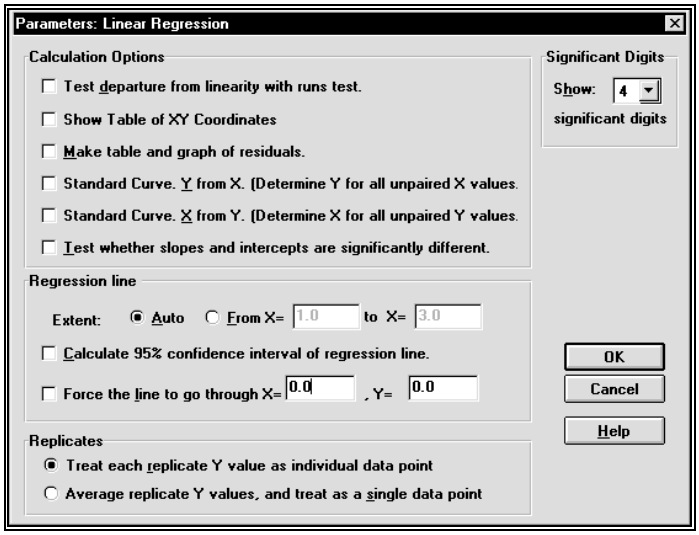
To graph the 95% CI of the linear regression line:

1. On the Parameters dialog, check the option "Calculate 95% CI of regression line".
2. Go to the graph.
3. Click on Change and choose Symbols & Lines.
4. In the Symbols dialog, choose the data set containing the regression line (not the original data).
5. Click the option box for error bars. Choose the - - - - - style.





Parameters for linear regression



Calculation options

- Test departure from linearity with runs test. Linear regression is based on the assumption that the relationship between X and Y is linear. If you select this option, Prism tests that assumption with the runs test. See "Runs test" on page 264.
- Show table of XY coordinates. You'll rarely want to see this table, but it can be useful if you want to export the XY coordinates of the regression line (and its confidence interval) to another program.
- Make table and graph of residuals. See “Residuals” on page 245.
- Standard curve calculations. See "Reading unknowns from standard curves" on page 349.
- Compare slopes and intercepts. See "Comparing slopes and intercepts" on page 246.

Regression line options

- Extend of regression line. Check auto to create a line that starts at the smallest X value in your data set and extends to the largest. Or enter the minimum and maximum X values. This determines the appearance of the regression line, but doesn't affect any calculated values.

- Calculate 95% confidence intervals. To graph the 95% CI of the regression line, you need to check this option to calculate the necessary values. Then check the error bar option box in the Symbols dialog to include the confidence bands on the graph.
- Force line through a specified point whose X and Y coordinates you enter (usually 0,0).

How to treat replicates

If you collected replicate Y values at every value of X, there are two ways to analyze the data:

- Treats each replicate as a separate point.
- Average the replicate Y values, and treat the mean as a single point.

Deciding which approach to use can be difficult. This topic is discussed in depth in the nonlinear regression chapter. See "Average replicates?" on page 256.

Interpreting the results of linear regression

Slope and intercept

Prism displays the values of the slope and Y-intercept with standard errors and 95% confidence intervals. If the assumptions of linear regression are true, then you can be 95% certain that confidence interval contains the true population values of the slope and intercept.

Goodness of fit

Prism assesses goodness-of-fit by reporting $s_{y.x}$ and r^2

The value r^2 is a fraction between 0.0 and 1.0, and has no units. When r^2 equals 0.0, there is no linear relationship between X and Y. In this case, the best-fit line is a horizontal line going through the mean of all Y values, and knowing X does not help you predict Y. When $r^2 = 1.0$, all points lie exactly on a straight line with no scatter. If you know X, you can predict Y exactly. With most data, r^2 is between 0.0 and 1.0.

You can think of r^2 as the fraction of the total variance of Y that is "explained" by variation in X. The value of r^2 (unlike the regression line itself) would be the same if X and Y were swapped. So r^2 is also the fraction of the variance in X that is "explained" by variation in Y. Therefore, r^2 is the fraction of the variation that is shared between X and Y.

The value $s_{y.x}$ is a bit harder to understand. It is the standard deviation of the vertical distances of the points from the line. Since the distances of the points

from the line are termed residuals, $s_{y.x}$ is the standard deviation of the residuals. Its value is expressed in the same units as Y.

Is the slope significantly different than zero?

Prism reports the P value testing the null hypothesis that the overall slope is zero. The P value answers this question: If there were no linear relationship between X and Y overall, what is the probability that randomly selected points would result in a regression line as far from horizontal as we observed? The P value is calculated from an F test, and Prism also reports the value of F and its degrees of freedom.

Residuals and the runs test

Residuals are the vertical distances of each point from the regression line. The X values in this table are identical to the X values you entered. The Y values in this table are the residuals — the vertical distances of each point from the regression line. A positive value means the point is above the line; a negative value means the point is below the line.

If you create a table of residuals, Prism automatically makes a new graph containing the residuals and nothing else. It is easier to interpret the graph than the table of numbers.

If the assumptions of linear regression have been met, the residuals will be randomly scattered above and below the line at Y=0. The scatter should not vary with X. You also should not see large clusters of adjacent points that are all above or all below the Y=0 line.

The runs test determines whether your data differ significantly from a straight line.

A run is a series of consecutive points that are either all above or all below the regression line. In other words, a run is a series of points whose residuals are either all positive or all negative.

If the data follow a curve rather than a line, the points will tend to be clustered together above or below the line. There will be too few runs. The P value answers this question: If the data points are randomly scattered around a straight line, what is the chance of finding as few (or fewer) runs than you observed. If there are fewer runs than expected, the P value will be low, suggesting that your data follow a curve rather than a straight line.

These topics are discussed in more detail in the nonlinear regression chapter. See "Residuals and the runs test" on page 271.

Comparing slopes and intercepts

If you entered more than one data set and selected the option on the Parameters dialog, Prism will test whether the slopes and intercepts are significantly different.

Prism compares slopes first. It calculates a P value testing the null hypothesis that the slopes are all identical in the overall populations. The P value answers this question: If the slopes were all identical overall, what is the chance of randomly selecting data points and ending up with slopes as different (or more so) than actually observed. If the P value is less than 0.05, Prism concludes that the lines are significantly different. In that case, there is no point in comparing the intercepts. The P value is two-sided.

If the P value for comparing slopes is greater than 0.05, Prism concludes that the slopes are not significantly different and calculates an overall (pooled) slope. Since the slopes are assumed to be identical, there are two possibilities. Either the lines are identical, or they are different but parallel. Prism calculates a second P value testing the null hypothesis that the lines are identical. If this P value is low, you will conclude that the lines are not identical, but rather that they are distinct but parallel. If this second P value is high, you can conclude that there is no evidence that the lines are not identical.

For more information on comparing regression lines, see Chapter 18 of J. H. Zar, Biostatistical Analysis, Second edition 1984, Prentice-Hall.

Standard Curve

To read unknown values from a standard curve, you must enter unpaired X or Y values below the X and Y values for the standard curve.

Depending on which option(s) you selected in the Parameters dialog, Prism calculates Y values for all the unpaired X values and/or X values for all unpaired Y values and places these on new output views.

See "Reading unknowns from standard curves" on page 349.

Assumptions of linear regression

The linear regression model is based on these assumptions:

- The relationship between X and Y can be graphed as a straight line. In many experiments, the relationship between X and Y is curved. Linear regression is not appropriate for this type of data.
- The variability of values around the line follows a Gaussian distribution. Even though no biological variable follows a Gaussian distribution exactly, it is sufficient that the variation be approximately Gaussian.

- The variability is the same everywhere. In other words, replicate Y values have the same SD everywhere, regardless of the value of X. The assumption that the SD is the same everywhere is termed *homoscedasticity*.
- You know X exactly. The model only provides for variability in the Y variable. The model assumes that X values are exact. This is rarely the case, but it is sufficient to assume that any imprecision in measuring X is very small compared to the variability in Y.
- Random sampling. Like all statistical tests, linear regression assumes that each subject (or each XY data pair) was randomly sampled from a very large population. At a minimum, you need to assume that the subjects are representative of the entire population.
- Independent subjects. Each subject (or each XY data pair) was selected independently. Picking one subject from the population should not influence the chance of picking another.

Note that X and Y are asymmetrical. Switching X and Y would result in a different regression line.

Default options for linear regression

If you always want to choose the same options for linear regression, you should define your preferences in the Analysis Options dialog. The settings in this dialog become the default settings whenever you perform linear regression.

The Analysis Options dialog sets these defaults for linear regression:

- Test departure from linearity with runs test?
- Make table and graph of residuals?
- Test whether slopes and intercepts are different?
- Calculate and graph 95% CI of regression line?

To change the analysis options:

1. Pull down the Edit menu.
2. Choose Options.
3. Then choose Analysis options.

Note: Changing the analysis options changes the default settings for future linear regression analyses. It will not change analyses you have already performed.

How linear regression works

In general, the goal of linear regression is to find the line that best predicts Y from X. Linear regression does this by finding the line that minimizes the sum of the squares of the vertical distances of the points from the line.

Why minimize the square of the distances?

- Distances are squared because it is better to have two points sort of close to the line (say 5 units each) than to have one very close (1 unit) and one further (9 units). If the scatter of points around the line is Gaussian, the former is far more likely than the latter.
- A more rigorous answer (for those who have studied statistics intensively) is that minimizing the sum-of- squares results in the same line that would be given by maximum likelihood calculations.

Note that linear regression does not *test* whether your data are linear (except for the runs test). It assumes that your data are linear, and finds the slope and intercept that make a straight line best fit your data.

Introduction to nonlinear regression

Introduction to nonlinear regression

Nonlinear regression is a powerful tool for analyzing scientific data, especially in pharmacology and physiology. Because it is a topic ignored by most statistics books, we explain nonlinear regression in more detail than the other analyses performed by Prism. This chapter explains the principles of nonlinear regression, the next explains how easily you can perform nonlinear regression with Prism, and the following chapter helps you interpret the results.

The goal of nonlinear regression is to fit a model to your data. The program finds the best-fit values of the variables in the model (perhaps rate constants, affinities, receptor number, etc.) which you can interpret scientifically. In most cases, the primary goal is to obtain those values and a secondary goal is to draw a graph of the fit curve.

In some situations, your only goal is to draw a curve. You don't care about models or equations, and don't want to obtain best-fit values. You just want a smooth curve through your points either for artistic reasons or to use as a standard curve (see "Reading unknowns from standard curves" on page 349). You may still use nonlinear regression in these situations, or you may use these alternatives:

- Polynomial regression. See "Polynomial equations" on page 294.
- Cubic spline or LOWESS curve. See "Creating a LOWESS or cubic spline curve" on page 161.
- A program that fits your data to thousands of equations and picks the best.

This chapter (and the next two) assumes that your goal is primarily to obtain the best-fit values of the variables – to fit a model to your data.

A note on terminology. A model *is a formal presentation of a chemical or physiological idea. To be useful for nonlinear regression, the model must be expressed as an equation that defines Y, the outcome you measure, as a function of X and one or more variables that you want to fit. We use the term* variable *to refer to the terms in the equation you want to fit. In the context of nonlinear regression, the term variable does not refer to X and Y. Some programs and books use the word parameters rather than variables.*

Why you should use nonlinear regression

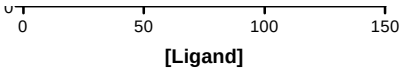
Linear regression of transformed data is less accurate

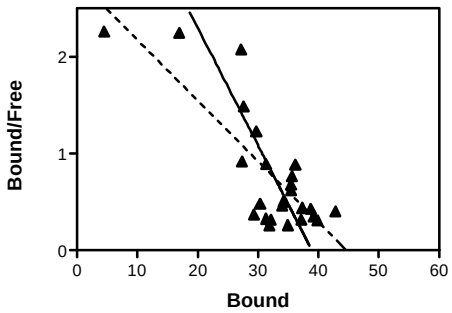
Before the age of microcomputers, nonlinear regression was not readily available to most scientists. Instead, scientists transformed their data to make a linear graph, and then analyzed the transformed data with linear regression. Examples include Lineweaver-Burke plots of enzyme kinetic data, Scatchard plots of binding data, and logarithmic plots of kinetic data.

These methods are outdated, and should not be used to analyze data. One problem is that the linear transformation distorts the experimental error. Linear regression assumes that the scatter of points around the line follows a Gaussian distribution and that the standard deviation is the same at every value of X. These assumptions are usually not true with the transformed data. A second problem is that some transformations alter the relationship between X and Y. For example, in a Scatchard plot the value of X (bound) is used to calculate Y (bound/free), and this violates the assumptions of linear regression. For more information, see "Assumptions of linear regression" on page 246.

Since the assumptions of linear regression are violated, the results of linear regression are incorrect. The values derived from the slope and intercept of the regression line are not the most accurate determinations of the variables in the model. Considering all the time and effort you put into collecting data, you want to use the best possible analysis technique. Nonlinear regression produces the most accurate results.

This figure shows the problem of transforming data. The left panel shows data that follows a rectangular hyperbola (binding isotherm). The right panel is a Scatchard plot of the same data. The solid curve on the left was determined by nonlinear regression. The solid line on the right shows how that same curve would look after a Scatchard transformation. The dotted line shows the linear regression fit of the transformed data. The transformation amplified and distorted the scatter, and thus the linear regression fit does not yield the most accurate values for B_{max} and K_d.





Transformations can be very useful when used appropriately. When analyzing data, follow these rules:

- You should transform your data when the transformation makes the variability more consistent and more Gaussian.
- You should not transform data when the transformation makes the variability less consistent and less Gaussian.
- You should not perform transforms (such as the Scatchard transform) that destroy the relationship between X and Y.
- You should not transform the data merely to make it linear. Since nonlinear regression is easy, there is no reason to force your data into a linear form.

Although it is usually inappropriate to analyze transformed data, it is often helpful to display data after a linear transform. Many people find it easier to visually interpret transformed data. This makes sense because the human eye and brain evolved to detect edges (lines) — not to detect rectangular hyperbolas or exponential decay curves. Even if you analyze your data with nonlinear regression, it may make sense to display transformed data.

Don't relegate scientific decisions to a computer program

The goal of nonlinear regression is to fit a model to your data. The program finds the best-fit values of the variables in the model (perhaps rate constants, affinities, receptor number, etc.) which you can interpret scientifically. Choosing a model is a scientific decision. You should base your choice on your understanding of chemistry or physiology (or genetics, etc.). The choice should not be based solely on the shape of the graph.

Some programs (not available from GraphPad) automatically fit data to hundreds or thousands of equations and then present you with the equation(s) that fit the data best. Using such a program is appealing because it frees you from the need to choose an equation. The problem is that the program has no understanding of the scientific context of your experiment. The equations that

fit the data best are unlikely to correspond to scientifically meaningful models. You will not be able to interpret the best-fit values of the variables, and the results are unlikely to be useful for data analysis.

This kind of approach is very useful in three situations:

- Your only goal is to plot an attractive curve.
- You wish to create a standard curve for interpolating unknown values.
- You need an equation to use within a computer simulation.

In all three situations, it doesn't matter whether the equation corresponds to a biological, chemical or physical model. What matters is that the equation accurately predicts Y from X within the range of your data.

This approach can be useful in some situations. Don't use it when the goal of curve fitting is to fit the data to a model based on chemical, physical, or biological principles. Don't use a computer program to avoid making a scientific decision.

The results of polynomial regression are often impossible to interpret scientifically

Beware of the term "curve fitting". The term is often used to refer not to nonlinear regression, but rather to polynomial regression. This method fits data to a polynomial equation: $Y = AX + BX + CX^2 + DX^3$...Programmers prefer polynomial regression, because it is so much easier to program. That's why it is built in to so many spreadsheet and graphics programs. But few biological or chemical models are described by polynomial equations, so polynomial regression is of limited usefulness to scientists.

Cubic spline is not a data analysis method

Cubic spline curves are smooth curves that go through every data point. In some cases, a cubic spline curve can look attractive on a graph and work well as a standard curve for interpolation. The curve does not correspond to any equation (or rather the equation differs for every pair of points) so cubic spline is not useful in data analysis.

How nonlinear regression works

Comparison of linear and nonlinear regression

A line is described by a simple equation that calculates Y from X, slope and intercept. The purpose of linear regression is to find values for the slope and intercept that define the line that comes closest to the data. More precisely, it

finds the line that minimizes the sum of the square of the vertical distances of the points from the line.

The goal of minimizing the sum-of-squares in linear regression can be achieved quite simply. A bit of algebra (shown in many statistics books) derives equations that define the slope and intercept. Put the data in, and the answers come out. There is no chance for ambiguity.

Nonlinear regression is more general. It can fit data to any equation that defines Y as a function of X and one or more variables. It finds the values of those variables that generate the curve that comes closest to the data. More precisely, the goal is to minimize the sum of the squares of the vertical distances of the points from the curve.

Except for a few special cases, it is not possible to directly solve the equation to find the values of the variables that minimize the sum-of-squares. Instead nonlinear regression requires an iterative approach.

Iterations in nonlinear regression

Here are the steps that every nonlinear regression program follows:

1. Start with an initial estimated value for each variable in the equation.
2. Generate the curve defined by the initial values. Calculate the sum-of-squares (the sum of the squares of the vertical distances of the points from the curve).
3. Adjust the variables to make the curve come closer to the data points. There are several algorithms for adjusting the variables. The most commonly used method was derived by Levenberg and Marquardt (often called simply the Marquardt method).
4. Adjust the variables again so that the curve comes even closer to the points.
5. Keep adjusting the variables until the adjustments make virtually no difference in the sum-of-squares. See "Convergence criteria" on page 265.
6. Report the best-fit results. The precise values you obtain will depend in part on the initial values chosen in step 1 and the stopping criteria of step 5. This means that repeat analyses of the same data will not always give exactly the same results.

Decisions you need to make when fitting curves with nonlinear regression

When you use a program for nonlinear regression, you must make the following decisions. See the next chapter for details on using Prism.

Choose a model

To use nonlinear regression, you must first define a mathematical model based on theory. The first step is to choose a model. For example, many kinds of binding data are explained by the law of mass action (see "Law of mass action" on page 310). The next step is to express the model as an equation that defines Y as a function of X and one or more variables. Some programs (not Prism) also let you define the model as a differential equation that defines dY/dX as a function of one or more variables.

Choosing a model is a scientific decision, not a statistical one. The model needs to make sense in scientific terms. See "Prism and equations" on page 287.

You may also fit two different models to your data, and then use statistical methods (F test) to compare them. See "Comparing two equations" on page 277.

Prepare data for nonlinear regression

When preparing data for nonlinear regression, keep these points in mind:

- It matters which variable is X and which is Y. X should be the variable you control or manipulate. Y is the variable you measure. Nonlinear regression finds the curve that lets you best predict Y from X.
- Use reasonable units. In pure mathematics, it doesn't matter whether you express your results as 1 picomolar or 10^{-12} molar, as 1 nanovolt or 10^{-9} volts. When computers do the calculating, however, it can matter. Calculation problems such as round off errors are far more likely when the values are very high or very low. We recommend that you scale your data to avoid values less than 10^{-4} or greater than 10^{4}.
- Don't smooth. You lose information when you smooth data, and this won't get you a better fit.
- If you are fitting data to a sigmoidal dose-response or competitive binding curve, enter the X values as the logarithm of concentration, rather than the concentration itself.

Estimate initial values

Nonlinear regression is an iterative procedure. The program must start with estimated values for each variable that are in the right “ball park” — say within a factor of five of the actual value. It then adjusts these initial values to improve the fit. It then adjusts the values again and again until the improvement is tiny.

If you have “clean” data that clearly define a curve, then it usually doesn't matter if the initial values are fairly far from the correct values. You'll get the same answer no matter what initial values you use, unless the initial values are very far from correct.

Initial values matter more when your data have a lot of scatter, don't span a large enough range of X values to define a full curve, or don't really fit the model. In these cases, you may get different answers depending on which initial values you use. See “Is the fit a local minimum?” on page 275.

You'll find it easy to estimate initial values if you have looked at a graph of the data, and understand the model and what all the variables mean. Remember, you just need an estimate. It doesn't have to be very accurate. If you are having problems estimating an initial value:

- Check that you have chosen a model that makes scientific sense.
- Make sure you understand what each variable in the equation means.
- Put away your data, and spend an hour or two generating curves using the model. Change the variables one at a time, and see how they influence the shape of the curve.

Prism automatically provides initial values if you choose a built-in equation. If you use a user-defined equation, you can define rules for obtaining initial values from the range of the X and Y values. Once you define these rules, Prism will automatically determine the initial values in the future. See "Defining rules for initial values in nonlinear regression" on page 301.

Constants

You don't have to fit every variable in the equation. In many situations it makes sense to fix some of the variables to constant values. For example, you might want to define the bottom plateau of a dose-response curve or an exponential decay curve to equal zero.

Weighting

In general, the goal of nonlinear regression is to find the values of the variables in the model that make the curve come as close as possible to the points. Usually this is done by minimizing the sum of the squares of the vertical distances of the data points from the curve. This is appropriate when you

expect that the scatter of points around the curve is Gaussian and unrelated to the Y values of the points. (Note to those who have studied advanced statistics: If those assumptions are true, minimizing the sum-of-squares is equivalent to finding the maximum likelihood estimate of the variables).

With many experimental protocols, you don't expect the experimental scatter to be the same, on average, for all points. Instead, you expect the experimental scatter to be a constant percentage of the Y value. If this is the case, points with high Y values will have more scatter than points with low Y values. When the program minimizes the sum of squares, points with high Y values will have a larger influence while points with smaller Y values will be relatively ignored. You can get around this problem by minimizing the sum of the square of the relative distances. This procedure is termed *weighting* the values by $1/Y^2$. Because it prevents large points from being over-weighted, the term *unweighting* seems more intuitive.

It is also possible to weight the data in other ways. The goal, always, is to end up with a measure of goodness-of-fit that weights all the data points equally.

Average replicates?

If you collected replicate Y values at every value of X, there are two ways to analyze the data:

- Treats each replicate as a separate point.
- Average the replicate Y values, and treat the mean as a single point.

Deciding which approach to use can be difficult.

The advantage of the first approach is that you have more data points and thus more degrees of freedom. However, you should only use that approach when the experimental error of each replicate is no more closely related to the other replicates than to other data points. Here are two examples where you should analyze each replicate:

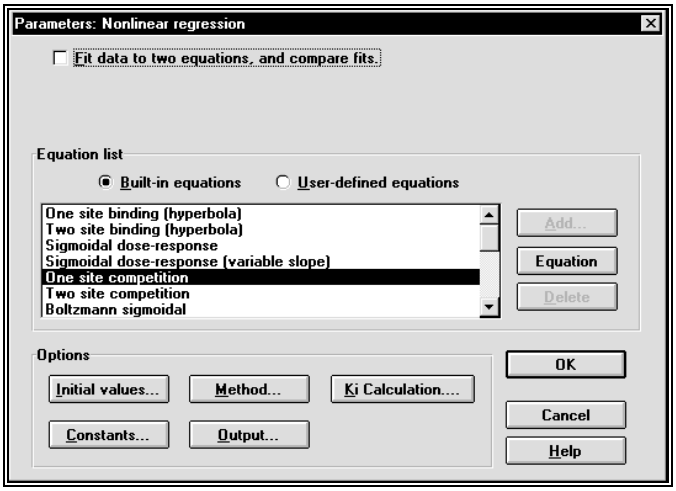
- You are doing a radioligand binding experiment. All the data were obtained from one tissue preparation and each replicate was determined from a separate incubation (separate test tube). The sources of experimental error are the same for each tube. If one value happens to be a bit high, there is no reason to expect the other replicates to be high as well.
- You are doing an electrophysiology study. You apply a voltage across a cell membrane and measure conductance. Each data point was obtained from a separate cell. The possible sources of experimental error are independent for each cell. If one cell happens to have a high conductance, there is no reason to expect the replicate cells (those that you apply the same voltage to) to also have high conductance.

You should not treat each replicate as a separate point when the experimental error of the replicates are related. You should average the replicates instead, and analyze the averages. Here are two examples where you should average the replicates:

- The experiment was only performed with a single replicate at each value of X, and you measure radioactivity as Y. Each tube is counted three times, and the three counts are treated as replicates. Any experimental error while conducting the experiment would appear in all the replicates. The replicates are not independent.
- The experiment is a dose-response curve. At each dose, you use a different animal but measure the response three times. The three measurements are not independent. If an animal happens to respond more than the others, that will affect all the replicates. The replicates are not independent.

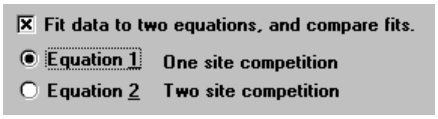
Nonlinear regression with Prism

Performing nonlinear regression



To perform nonlinear regression:

1. You may start either from a table of data or from a graph. Press the Analyze button and select Nonlinear regression.
2. Pick an equation. Choose either a built-in equation, or one that you have entered yourself. See "Choosing or entering equations" on page 287.
3. To fit to two equations simultaneously, select the option box "Fit data to two equations" and then select both equations. See "Comparing two equations" on page 277.



The F test to compare equations is only strictly valid when the simpler equation is a special case of the more complicated equation. For example, you can compare a one-site vs. two-site binding curve. You should not pick two unrelated equations. Prism does not enforce this rule.

3. If necessary, set other options. In many situations, you can skip these options.

- Initial values. Prism automatically supplies initial values for all variables, so you rarely have to think about them. Press this button to inspect and alter the initial values. See "Initial values for nonlinear regression" on page 261.
- Constants. Fix the value of one or more variables in the equation. See "Constants for nonlinear regression" on page 262.
- Method. Nonlinear regression is a complicated procedure, and Prism offers many options. See "Method options for nonlinear regression" on page 263.
- Output. Select optional output such as residuals, standard curve calculations or a summary sheet. See "Output options for nonlinear regression" on page 265.
- Calculating K_i. When fitting curves to competitive binding data, calculate the K_i automatically. See "Ki" on page 320.

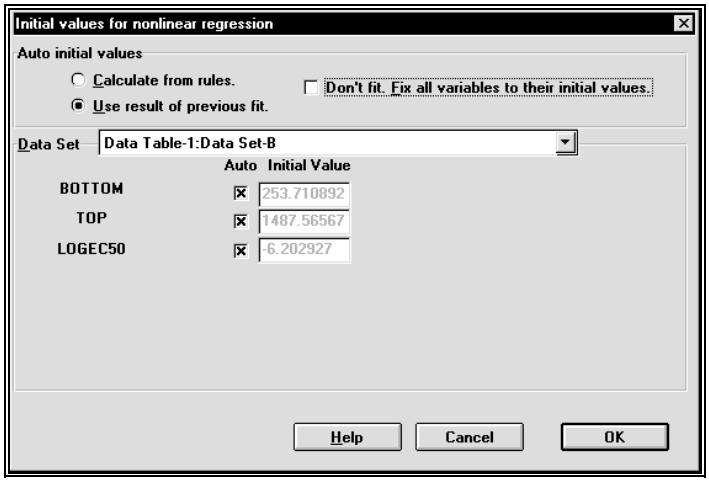
Graphing curves

After you perform nonlinear regression, Prism automatically graphs the resulting curve with the original data. When Prism defines a curve, it actually creates a table of XY values. You can create this table as one of the output views of nonlinear regression, although you'll rarely find it of interest.

To automatically graph the curve with the data points, Prism adds the table of XY values to the list of data sets included on the graph. The original data table and the table containing the curve are separate tables included on one graph.

When you change the appearance of the graph using the Symbols dialog, note that curves are separate data sets from the original data. Graph the curve using no symbols and with straight lines connecting the "points". Graph the data as symbols without connecting straight lines.

Initial values for nonlinear regression



Nonlinear regression is an iterative procedure. The program must start with estimated values for each variable that are in the right "ball park" — say within a factor of five of the actual value.

Prism automatically provides initial values for each variable, calculated from the range of your data. If you select a built-in equation, the rules for calculating the initial values are built-in to the program. If you enter a user-defined equation, you define the rules. See "Defining rules for initial values in nonlinear regression" on page 301.

To view and change the initial values:

1. Press the Initial values button on the Nonlinear Regression Parameters dialog.
2. Select a data set from the drop down list.
3. Look at the list of values.
4. If you've already run the curve fit and are changing the options, Prism will set the initial values to the results of the previous fit rather than to automatically calculated values. This usually makes the fit go faster. If it causes problems, click the option box "Calculate from rules" rather than "Use result of previous fit".
5. If you want to change the initial value of a variable, deselect the AUTO check box next to a variable and enter the new initial value.

Note: To see the curve generated by the initial values, select Don t fit. Fix all variables to their initial values .

How much difference do initial values make?

If you have "clean" data that clearly define a curve, then it usually doesn't matter if the initial values are fairly far from the correct values. You'll get the same answer no matter what initial values you use, unless the initial values are very far from correct.

Initial values matter more when your data have a lot of scatter, don't span a large enough range of X values to define a full curve, or don't really fit the model. In these cases, you may get different answers depending on which initial values you use. See "Is the fit a local minimum?" on page 275.

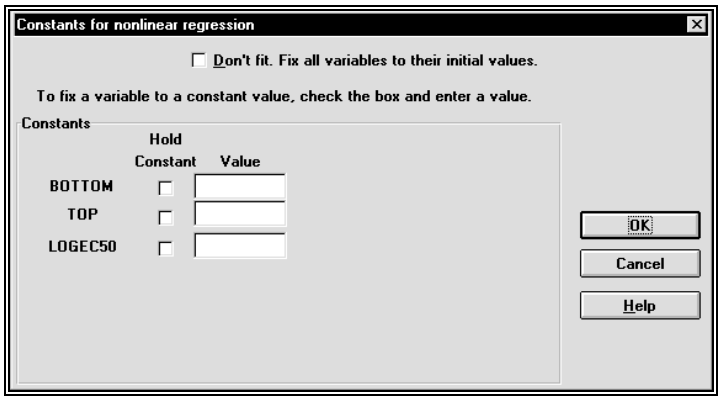
Constants for nonlinear regression

When performing nonlinear regression, you don't have to fit each variable in the equation. Instead, you may fix one or more of the variables to constant values.

It is often helpful to define constants when you have only a few data points. For example, you might fix the bottom plateau of a sigmoidal curve or exponential decay to zero.

Remember that Prism has no "common sense". Prism does not know how you did the experiment. It doesn't know that a curve has to plateau at zero unless you tell it.

Prism will let you set all the variables to constant values. If you do that, Prism will generate a theoretical curve based on the constants you enter. There is nothing to fit. When you first use a user-defined equation, you may find this useful to make sure that your rules for initial values really do generate a curve that is somewhere near the points.

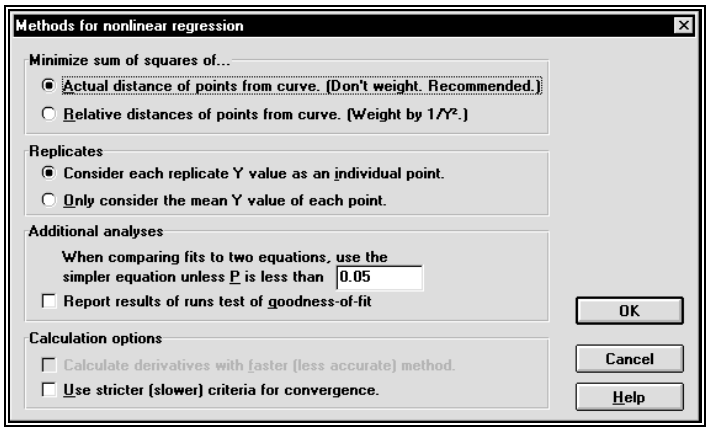


To set constants:

1. Click on the Constants button in the Nonlinear Regression Parameters dialog to bring up the Constants dialog.
2. Check the "constant" box in front of the variables you want to hold constant and enter their values. As a shortcut, you can enter the value first and Prism will automatically select the check box.
3. It is not possible to enter different constant values for each data set. If you set a variable to a constant value, it will have that value for every data set analyzed.

Note: To see the curve generated by the initial values, select Don t fit. Fix all variables to their initial values .

Method options for nonlinear regression



Prism provides these options for adjusting the method it uses to calculate nonlinear regression.

Weighting method

Prism normally minimizes the sum of the squares of the vertical distances of the data points from the curve. This is appropriate when you expect that the scatter of points around the curve is unrelated to the Y values of the points.

With many experimental protocols, you expect the experimental scatter to be a constant percentage of the Y value. If this is the case, points with high Y values will have more scatter than points with low Y values. When the program minimizes the sum of squares, points with high Y values will have a larger influence while points with smaller Y values will be relatively ignored. You can get around this problem by minimizing the sum of the square of the

relative distances. This procedure is termed *weighting* the values by $1/Y^2$. Because it prevents large points from being over-weighted, the term *un-weighting* seems more intuitive.

For a discussion of weighting in analysis of radioligand binding data, see "Weighting" on page 318.

Replicates

If you entered your data with replicate Y values, you have two choices:

- Treat each replicate as a separate point.
- Average the replicates and analyze the mean values.

See "Average replicates?" on page 256.

Threshold P value

When you compare two different equations, Prism performs an F test and chooses the best equation based on the resulting P value. The simpler equation (fewer variables) is chosen unless the P value is less than the threshold P value you enter here. Most investigators leave this set to 0.05. See "Comparing two equations" on page 277.

Runs test

If you check this box, Prism will calculate a runs test of goodness-of-fit. The runs test determines whether your data systematically deviate from the predictions of the model. See “Residuals and the runs test” on page 271.

Derivatives

The nonlinear regression algorithm repeatedly evaluates the partial derivative of the equation with respect to each variable. This is the slowest part of nonlinear regression.

If you choose a built-in equation, Prism uses analytical derivatives built-in to the program. There is no choice for you to make.

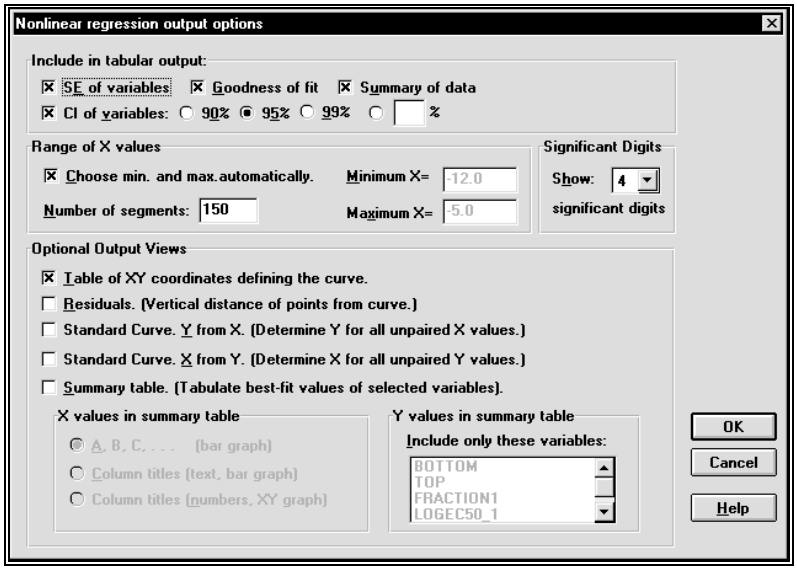
If you enter your own equation, Prism evaluates the derivatives numerically using one of two methods. One of the methods is slower, but potentially more accurate. In most cases, the results will be identical with both methods. We recommend that you use the slow but accurate method to validate results with a new equation, and then switch to the quick method for routine analyses. For details, see “How Prism calculates derivatives” on page 268.

Convergence criteria

Prism stops iterating and declares the results to have converged when two iterations in a row change the sum-of-squares by less than 0.01%. If you check the box for "strict convergence criteria", Prism will continue the iterations until three consecutive iterations each reduce the sum-of-squares by less than 0.0001%. Using the stricter criteria slows the calculations by a second or two, and usually has a negligible effect on the results.

We recommend that you use the slow method only when you are having difficulty fitting an equation, or to validate your results. Use the standard method for routine analyses.

Output options for nonlinear regression



Include in tabular output

Choose which parts of the output you wish to see. We recommend that you leave all the boxes checked to get the most complete output.

Also choose between 90%, 95% or 99% confidence intervals for the fit variables. By convention, 95% confidence intervals are reported most often.

Number of significant digits

Choose the number of significant digits used to report results. This is especially useful if you copy the results to hot linked table objects on graphs or layouts. See "Embedding data and results tables" on page 113.

Range of X values and number of line segments

Ordinarily Prism draws the curve within the range of data defined by each data set. In other words, it draws the curve from the X position of the first (lowest X) data point to the last data point. You may enter different limits.

Curves are defined by many short line segments. You decide how many segments to use. If you increase the number of segments, graphs of curves will appear smoother but calculations will take longer.

Prism initially creates all curves with 150 line segments. Increasing the number will improve the accuracy of standard curve calculations, and may make the curve appear smoother (especially if it has many inflection points).

If you entered two or more data sets that span different ranges of X values, the resolution is an estimate. It is impossible for each curve to have the same number of line segments when they share the same X values, but different X ranges.

Residuals

Residuals are the vertical distances of the data points from the curve. Prism will create a table of residuals. The X values are the same as those in your original data and the Y values are replaced by the vertical distance of the points from the curve. If the residual is positive, it means the point is above the curve. If the residual is negative, it means the point is below the curve.

Prism also creates a graph of the residuals. See "Residuals and the runs test" on page 271.

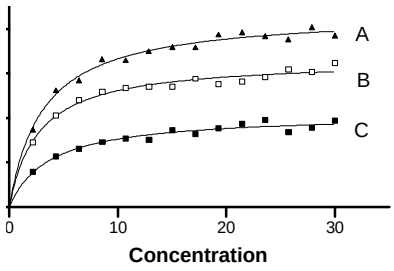
Standard curve calculations

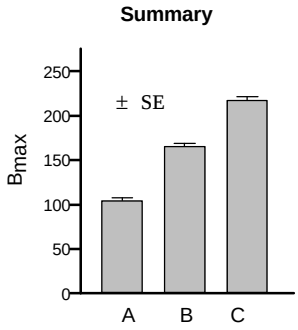
You may use the curve generated by nonlinear regression as a standard curve and read off unknown values. Check the appropriate option(s), Y from X, or X from Y. For more information, see "Reading unknowns from standard curves" on page 349.

Summary table and graph

If you are analyzing a data table with more than one data set, the table of results is rather lengthy. You may summarize key results on a separate sum-

mary table that displays the best-fit value of one (or more) variable for each data set.





To make a summary table:

1. From the Parameters dialog, press the Output button.
2. Check the box to create a summary table.
3. Choose how you want the X values of this table to appear. You have three choices: A, B, C for a bar graph; Data set (column) titles for a bar graph; Data set (column) titles converted to numbers for an XY graph.
4. Select one or more variables to include on the summary table. The table and graph will be much easier to understand if you only select a single variable.

Prism automatically makes a graph from the summary table. Depending on your choices in the dialog, this may be a bar graph or an XY graph. It shows the best-fit value of a selected variable for each data set on the table.

If it is an XY graph, you may analyze it further with linear or nonlinear regression. For example, the summary graph may show the best-fit value of a rate constant as a function of concentration (obtained from the column titles of the original data). You can fit a line or curve to that graph.

Note: When fitting two equations to each data set, the summary table includes the results for the second equation only. Since this may not be helpful, we suggest that you only make summary tables when fitting a single equation.

Default options for nonlinear regression

The Parameters dialog for nonlinear regression has many optional settings. If you always want to choose the same options, you should define your prefer-

ences in the Analysis Options dialog. The settings in this dialog become the default settings whenever you perform nonlinear regression.

The Analysis Options dialog gives you these choices for nonlinear regression:

- Minimize sum-of-square of absolute distances or relative distances?
- Report results of runs test of goodness-of-fit?
- Use stricter (slower) criteria for convergence?
- Make table and graph of residuals?
- Number of line segments to generate curves.

To change the analysis options:

1. Pull down the Edit menu.
2. Choose Options. Then choose Analysis options.

Note: Changing the analysis options changes the default settings for future nonlinear regression analyses. It will not change analyses you have already performed.

How Prism calculates derivatives

While performing nonlinear regression, Prism repeatedly evaluates the partial derivative of your equation with respect to each variable. This is the most time consuming part of nonlinear regression.

Ordinarily, Prism uses Richardson's method to evaluate the derivatives. This method is accurate but slow. Check the option box to use a quicker, but potentially less accurate, method. If you select this method, nonlinear regression will progress approximately twice as fast (it depends on your equation).

You can choose the method in two places. While entering the equation, choose the default method for that equation. Override that default for a particular fit in the methods options dialog.

Richardson's method (the slow method) calculates the derivative by determining Y after both increasing and decreasing the value of the variable a bit. The faster method only determines Y after increasing the value of the variable, and thus performs fewer calculations. In most cases, the results of the two methods are virtually identical.

We recommend that you use the standard (slow) method to validate results with a new equation, and then switch to the quick method for routine analyses.

Interpreting nonlinear regression results

Assumptions of nonlinear regression

The results of nonlinear regression are meaningful only if these assumptions are true (or nearly true):

- The model is correct. Nonlinear regression adjusts the variables in the equation you chose to minimize the sum-of-squares. It does not attempt to find a better equation.
- The variability of values around the curve follow a Gaussian distribution. Even though no biological variable follows a Gaussian distribution exactly, it is sufficient that the variation be approximately Gaussian.
- The SD of the variability is the same everywhere, regardless of the value of X. The assumption is termed *homoscedasticity* . If the SD is not constant but rather is proportional to the value of Y, you should weight the data to minimize the sum-of-squares of the relative distances. See "Weighting method" on page 263.
- The model assumes that you know X exactly. This is rarely the case, but it is sufficient to assume that any imprecision in measuring X is very small compared to the variability in Y.
- The errors are independent. The deviation of each value from the curve should be random, and should not be correlated with the deviation of the previous or next point. If there is any carryover from one sample to the next, this assumption will be violated.

Variables, standard errors, and confidence intervals

Along with the best-fit value of each variable in the equation, Prism reports its standard error and 95% confidence interval.

By itself, the standard error is difficult to interpret. It is used to calculate the 95% confidence interval, which is easier to interpret.

This is what the CI is supposed to mean: If all the assumptions of nonlinear regression are true, there is a 95% chance that the true value of the variable lies within the interval. More precisely, if you perform nonlinear regression many times (on different data sets) you expect the confidence interval to include the true value 95% of the time, but to exclude the true value the other 5% of the time.

Three factors can make the confidence interval too narrow:

- The CI is based only on the scatter of data points around the curve within this one experiment. If you repeat the experiment many times, you'll introduce new sources of error, so the variability between the experiments is likely to be greater than predicted from the CI based on one experiment.
- The CI can only be interpreted if you accept the assumptions of nonlinear regression. See "Assumptions of nonlinear regression" on page 269.
- The confidence intervals from linear regression are calculated using straightforward mathematical methods. If you accept the assumptions of linear regression, then you can interpret the 95% CI of slope and intercept quite rigorously. It is not straightforward to calculate the 95% CI of variables from nonlinear regression – mathematical shortcuts are needed. These shortcut intervals (the ones reported by Prism) are sometimes referred to as asymptotic confidence intervals. In some cases these intervals can be too narrow (too optimistic).

Because of these problems, you shouldn't interpret the confidence intervals too rigorously. Rather than focusing on the CI reported from analysis of this one experiment, you should repeat the experiment several times.

Sum-of-squares, $s_{y.x}$, and R^2

The sum-of-squares (SS) is the sum of the square of the vertical distances of the points from the curve. Nonlinear regression works by varying the values of the variables to minimize the sum-of-squares. It is expressed in the square of the units used for the Y values.

The value $s_{y.x}$ is the standard deviation of the vertical distances of the points from the line. Since the distances of the points from the line are called residuals, $s_{y.x}$ is the standard deviation of the residuals. Its value is expressed in the same units as Y.

The value R^2 is a measure of goodness of fit. It is a fraction between 0.0 and 1.0, and has no units. When R^2 equals 0.0, the best-fit curve fits the data no better than a horizontal line going through the mean of all Y values. In this

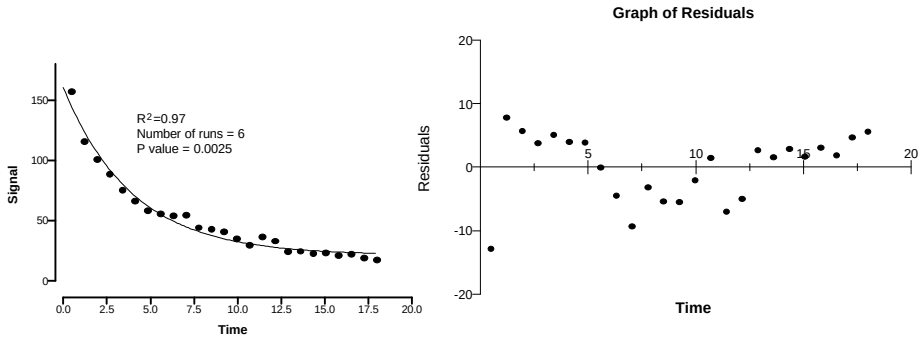
case, knowing X does not help you predict Y. When $R^2 = 1.0$, all points lie exactly on the curve with no scatter. If you know X you can calculate Y exactly.You can think of R^2 as the fraction of the total variance of Y that is explained by the model (equation). Mathematically, it is defined by this equation (the denominator is the standard deviation of Y squared, which is the same as the variance of Y).

$$R^2 = 1.0 - \frac{SS}{s_y^2}$$

Residuals and the runs test

A residual is the distance of a point from the curve. A residual is positive when the point is above the curve, and is negative when the point is below the curve. The residual table has the same X values as the original data, but each Y value is replaced by the vertical distance of the point from the curve.

If you selected the residuals output option, Prism creates a graph of the residuals. An example is shown below. If you look carefully at the curve on the left, you'll see that the data points are not randomly distributed above and below the curve. There are clusters of points all above or all below. This is much easier to see on the graph of the residuals on the right. The points are not randomly scattered above and below the X axis.



The runs test determines whether your data differ significantly from the equation you selected. A run is a series of consecutive points that are either all above or all below the regression curve. Another way of saying this is that a run is a series of points whose residuals are either all positive or all negative.

If the data points are randomly distributed above and below the regression curve, it is possible to calculate the expected number of runs. If there are fewer runs than expected, it may mean that the regression model is wrong.

The P value from the runs test answers this question: If the data really follow the linear or nonlinear equation used to create the line or curve, what is the chance of obtaining as few (or fewer) runs as observed in this experiment? If the P value is small, you'd be inclined to conclude that the data really don't follow the model.

The P values are always one-tail, asking about the probability of observing as few runs (or fewer) than observed. If you observe more runs than expected, the P value will be higher than 0.50.

If the runs test reports a low P value, you should suspect that the data don't really follow the equation you have selected.

In the example above, the equation does not adequately match the data. There are only six runs, and the P value for the runs test is tiny. This means that the data systematically deviate from the curve. Most likely, the data were fit to the wrong equation.

How to tell if the nonlinear regression fit is any good

Before accepting the results that Prism (or any curve fitting program) gives you, ask yourself the following questions:

Did the fit converge on a solution?

Nonlinear regression stops its iterations when it can't improve the fit by adjusting the values of any of the variables. At that point, the program is said to have *converged* on the best-fit. In some cases, the program gets stuck. It doesn't know whether the fit would improve by increasing or decreasing the value of a variable. When this happens, the program stops and says that it was unable to converge on a solution. No results are reported.

Does the curve come close to the points?

In rare cases, the fit may be far from the data points. This may happen, for example, if you picked the wrong equation. Look at the graph to make sure this didn't happen.

Also look at the R^2 value. It is the fraction of the overall variance in Y that is "explained" by the model. See "R " on page 270. If R^2 is low, the curve does not come close to the points. If R^2 is high, you can conclude that the curve comes closer to the points than would a horizontal line through the mean Y value. But don't over interpret a high R^2 . It does not mean that you have chosen the equation that best describes the data. It also does not mean that the fit is unique — other values of the variables may generate a curve that fits just as well.

Are the results scientifically plausible?

Prism fits curves and displays the results. It is up to you to figure out what they mean. Before accepting the results, ask yourself if the results make any sense.

The mathematics of curve fitting sometimes yields results that make no scientific sense. For example with noisy or incomplete data, Prism can calculate negative rate constants, fractions greater than 1.0, and negative K_d values. Its up to you to realize that these are nonsense.

If the results make no scientific sense, you should conclude that the fit is no good, regardless of R^2 and regardless of how close the curve comes to the points. Try a simpler equation, or try fixing some variables to constant values.

Also check that the best-fit values of the variables make sense in light of the range of the data. Don't trust the results if the top plateau of a sigmoid curve is far larger than the highest data point. Don't trust the results if an EC_{50} value is not within the range of your X values.

Do the data systematically deviate from the curve?

If the data really follow the model described by your equation, the data points should randomly bounce above and below the curve. The distance of the points from the curve should also be random, and not be related to the value of X.

The best way to look for systematic deviations of the points from the curve is to inspect a graph of the residuals and to look at the runs test. See "Residuals and the runs test" on page 271. With a good fit, the residuals should be randomly distributed between positive and negative values and the P value from the runs test will be high.

If the runs test reports a low P value, you should suspect that the data don't really follow the equation you have selected.

Are the confidence intervals wide?

Prism reports the standard error of each variable, and its 95% confidence interval. You can be approximately 95% sure that the true value of the variable lies within the confidence interval.

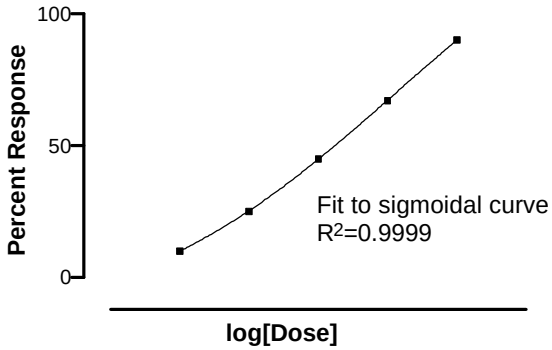
The confidence interval will be very wide (i.e. the standard error will be very large) when the fit is not unique. This means that curves generated from other values of the variables would fit nearly as well.

Confidence intervals are wide in these circumstances:

- You have not collected data over a wide enough range of X values. See the first example below.

- You have not collected data in an important part of the curve. See the second example below.
- The data are very scattered.
- The equation contains redundant variables. For example, the confidence intervals would be very wide if you fit this equation: Y = A + B + C*X. This equation describes a line, but the intercept is defined by the sum of A plus B. There is no way for the program to know how to apportion the value between A and B, so both will have very wide confidence intervals.

Example 1. Data not collected over a wide range of X.



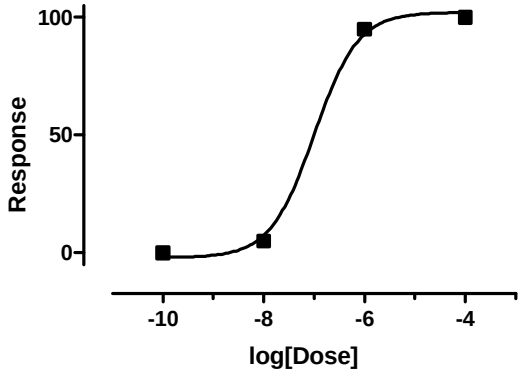
This best-fit dose response curve has wide confidence intervals. The 95% CI for the EC_{50} extends over six orders of magnitude!

The explanation is simple. The data were fit to a sigmoidal equation with four variables: the top plateau, the bottom plateau, the slope, and the EC_{50} (the log[Dose] when response = 50%). But the data do not form plateaus at either the top or the bottom, so the program is unable to fit unique values for the plateaus. The information is simply not in the data. Since the data do not define zero and one hundred, the value for the EC_{50} is very imprecise and the confidence interval is wide. The fit is not unique. You could find other values of the variables that fit the data equally well.

In this example, it might make scientific sense to set the bottom plateau to 0% and the top plateau to 100% (if the plateaus were defined by other controls not shown on the graph). If you did this, the equation would fit fine and the confidence interval would be narrow.

Note that the problem with the fit is not obvious by inspecting a graph, because the curve goes very close to the points. The value of R^2 (0.9999) is also not helpful. That value just tells you that the curve comes close to the points, but does not tell you whether the fit is unique.

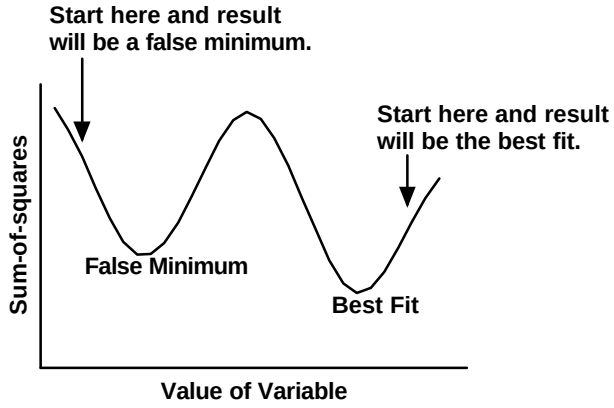
Example 2. Data missing in an important part of the curve.



This dose-response curve also has wide confidence intervals. Even if you constrain the bottom to be zero and the top to be 100 and the slope to equal 1.0, the 95% CI for the EC_{50} extends over almost an order of magnitude. The problem is simple. The EC50 is the concentration at which the response is half-maximal, and this example has no data near that point.

Is the fit a local minimum?

The nonlinear regression procedure adjusts the variables in small steps in order to improve the goodness-of-fit. If Prism converges on an answer, you can be sure that altering any of the variables a little bit will make the fit worse. But it is theoretically possible that large changes in the variables might lead to much better goodness-of-fit. Thus the curve that Prism decides is the "best" may really not be the best.



Think of latitude and longitude as representing two variables Prism is trying to fit. Now think of altitude as the sum-of-squares. Nonlinear works iteratively to reduce the sum-of-squares. This is like walking downhill to find the bottom of the valley. When nonlinear regression has converged, changing any variable increases the sum-of-squares. When you are at the bottom of the valley, every direction leads uphill. But there may be a much deeper valley over the ridge that you are unaware of. In nonlinear regression, large changes in variables might decrease the sum-of-squares.

This problem (called finding a local minimum) is intrinsic to nonlinear regression, no matter what program you use. You will rarely encounter a local minimum if your data have little scatter, you collected data over an appropriate range of X values, and you have chosen an appropriate equation.

To continue the analogy, the confidence intervals for the variables are very wide when the bottom of the valley is very flat. You can walk a great distance without changing elevation. You can change the values of the variables a great deal without changing the goodness-of-fit.

To test for the presence of a false minimum:

1. Note the values of the variables and the sum-of-squares from the first fit.
2. Make a large change to the initial values of one or more variables and run the fit again.
3. Repeat step 2 several times.
4. Ideally, Prism will report nearly the same sum-of-squares and same variables regardless of the initial values. If the values are different, accept the ones with the lowest sum-of-squares.

What to do when the fit is no good?

The last section explained how to identify a bad fit. Briefly, a fit is bad when:

- The fit did not converge.
- The results make no sense.
- The confidence intervals are wide.

If you encounter any of these situations, here are some things to try.

Potential problem	Solution
The equation simply does not describe the data.	Try a different equation.

The initial values are too far from their correct values.	Enter different initial values. If you are using a user-defined equation, check the rules for initial values.
The range of X values is too narrow to define the curve completely.	If possible, collect more data. Otherwise, hold one of the variables to a constant value.
You have not collected enough data in a critical range of X values.	Collect more data in the important regions.
Your data are very scattered and don't really define a curve.	Try to collect less scattered data. If you are combining several experiments, normalize the data for each experiment to an internal control.
The equation includes more than one component, but your data don't follow a multicomponent model.	Use a simpler equation.
Your numbers are too large.	If your Y values are huge, change the units. Don't use values greater than about 10^4.
Your numbers are too small.	If your Y values are tiny, change the units. Don't use values less than about 10^{-4}.

Comparing two equations

Sometimes you don't know which of two equations is more appropriate for your data. You want to fit both equations, and let the program compare the results. For example, you might want to fit a competitive binding curve to models with both one and two binding sites. Or you might want to fit a dissociation kinetics curve to exponential decay equations with both one and two phases.

Goodness of fit is quantified by the sum-of-squares. Therefore you might imagine that you can simply define the "best" equation as the one that gives the smaller sum-of-squares. That rule makes sense when both equations have the same number of variables.

Most often, however, you wish to compare equations with different numbers of variables. If the more complicated equation fits worse than the simpler equation, then you should clearly stick with the simpler equation. However, the curve generated by the more complicated equation (the one with more variables) will nearly always come closer to the points because it has more inflection points (it wiggles more).The question is whether this decrease in sum-of-squares is worth the "cost" of the additional variables (loss of degrees

of freedom). The F test addresses this question by calculating a P value that answers this question: If the simpler model is really correct, what is the chance that you'd randomly obtain data that fits the more complicated model so much better? If the P value is low, you conclude that the more complicated model is significantly better than the simpler model.

The results of the F test are only strictly valid when the simpler equation is a special case of the more complicated equation. For example, you can compare a one-site vs. two-site binding curve.

How the F test works

First fit the more complicated model (Model 2) and calculate its goodness-of-fit as the sum-of-squares. Now fit the simpler model (Model 1). Even if this simpler model is correct, you expect it to fit worse (have the higher sum-of-squares) because it has fewer inflection points (more degrees of freedom). In fact, statisticians can prove that the relative increase in the sum of squares is expected to equal the relative increase in degrees of freedom. In other words, if the simpler model is correct you expect that:

$$(SS1 - SS2) / SS2 \approx (DF1 - DF2) / DF2$$

If the more complicated model is correct, then you expect the relative increase in sum-of-squares (going from complicated to simple model) to be greater than the relative increase in degrees of freedom:

$$(SS1 - SS2) / SS2 > (DF1 - DF2) / DF2$$

The F ratio quantifies the relationship between the relative increase in sum-of-squares and the relative increase in degrees of freedom.

$$F = \frac{(SS1 - SS2) / SS2}{(DF1 - DF2) / DF2}$$

If the simpler model is correct you expect to get an F ratio near 1.0. If the ratio is much greater than 1.0, there are two possibilities:

- The more complicated model is correct.
- The simpler model is correct, but random scatter led the more complicated model to fit better. The P value tells you how rare this coincidence would be.

The P value answers this question: If model 1 is really correct, what is the chance that you'd randomly obtain data that fits model 2 so much better? If the P value is low, you conclude that model 2 is significantly better than model 1.

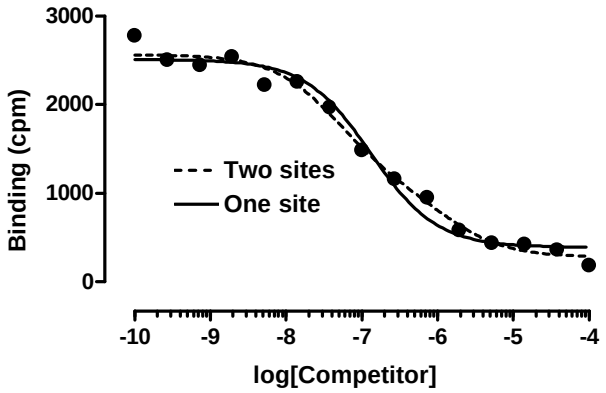
The equation is usually presented in this more conventional form.

$$F = \frac{(SS1 - SS2) / (DF1 - DF2)}{SS2 / DF2} \qquad DFn = (Df1 - DF2), \quad DFd = DF2$$

If you are extremely familiar with analysis of variance, you'll appreciate that the F ratio is determined from this analysis of variance table.

Source of variation	Sum-of-squares	df	MS
Difference	SS1 - SS2	DF1 - DF2	$\frac{SS1 - SS2}{DF1 - DF2}$
Model 2 (complicated)	SS2	DF2	SS2/DF2
Model 1 (simple)	SS1	DF1	

Example



This graph compares a one-site and two-site competitive binding curve. The results are shown here:

	Two-site	One-site	% Increase
Degrees of freedom	10	12	20.00%
Sum-of-squares	129800	248100	91.14%

In going from the two-site to the one-site model, we gained two degrees of freedom because the one-site model has two fewer variables. Since the two-site model has 10 degrees of freedom (15 data points minus 5 variables), the degrees of freedom increased 20%. If the one-site model were correct, you'd expect the sum-of-squares to also increase about 20% just by chance. In fact the sum-of-squares increased 91%. The percent increase was 4.56 times higher than expected (91.1/20.0 = 4.56). This is the F ratio (F = 4.56), and it corresponds to a P value of 0.039. If the one-site model is correct, there is only a 3.9% chance that you'd randomly obtain data that fits the two-site

model so much better. Since this is below the traditional threshold of 5%, you'd probably conclude that the two-site model fits significantly better than the one-site model. Here is how Prism reports the comparison:

Comparison of Fits	
DFn, DFd	2, 10
F	4.557
P value	0.03919
Best Fit Equation	Eq. 2

Comparing fits to two data sets

The previous section discussed how to compare the fits of two different equations to one set of data. Here we discuss how to compare the fit of one equation to two different sets of data, for example comparing fits to data from control and treated preparations. Although this is a common situation, there is no clear consensus for how to compare fits to different groups. Three approaches are discussed below.

1. Compare the results of repeated experiments.

If you repeat the experiment several times, you can compare the best-fit value of a variable in control and treated preparations using a paired t test (or the analogous Wilcoxon nonparametric test).

For example, here are the results of a competitive binding curve performed in two groups of cells. The table shows the logKi values.

Experiment	Control	Treated
1	-6.13	-6.53
2	-6.39	-6.86
3	-5.92	-6.31

Compare the results using a paired t test using Prism (or InStat). The t ratio is 16.7, and the P value is 0.0036 (two-tail). If the treatment did not alter the logKi, there is only a 0.36% chance that you would observe such a large difference (or larger) between logKi by chance.

Notes:

- We compare logKi values, not Ki values. When doing a paired t test, a key assumption is that the distribution of differences (treated - control) follow a Gaussian distribution. Since a competitive binding curve (similar to a dose response curve) is conducted with X values (concentration) equally spaced on a log scale, the uncertainty of X is reasonably symmetrical (and perhaps Gaussian) when expressed on a log scale. It is

equally likely that the observed logKi is 0.1 log units too high or 0.1 log units too low. In contrast, the uncertainty in Ki is not symmetrical.

- Prism can do the paired t test. Copy the results from each nonlinear regression results page onto a new data table for the t test.

2. Compare the results within one experiment. Simple approach.

When Prism reports the best-fit value of each variable for each data set, it also reports the standard error of the estimates. You can compare the best-fit values between two data sets using a t test.

For example, here are the results of fitting control and treated data to a competitive binding curve:

	X Labels	A	B
		Control	Treated
	X	Y	Y
1	Equation 1		
2	Variables		
3	BOTTOM	111.7	115.9
4	TOP	1206	1213
5	LOGEC50	-5.902	-6.290
6	EC50	1.251e-006	5.124e-007
7	Std. Error		
8	BOTTOM	22.09	23.09
9	TOP	20.29	29.87
10	LOGEC50	0.04986	0.06167
11	95% Confidence Intervals		
12	BOTTOM	63.53 to 159.8	65.55 to 166.1
13	TOP	1161 to 1250	1149 to 1279
14	LOGEC50	-6.011 to -5.794	-6.424 to -6.156
15	EC50	9.737e-007 to 1.606e-006	3.761e-007 to 6.982e-007
16	Goodness of Fit		
17	Degrees of Freedom	12	12
18	R^2	0.9928	0.9890
19	Absolute Sum of Squares	19560	29320
20	Sy.x	40.37	49.43

Nonlinear regression fit three variables, Top, Bottom, and LogEC50. We only care about comparing the LogEC50 values. To compare with a t test, enter the best fit value and its SE into a new data page:

	A			B		
	Control			Treated		
	Y	SEM	N	Y	SEM	N
1	-6.08	0.0677	13	-6.29	0.0617	13
2						

There is no way to transfer the values automatically – you need to copy the values manually. The only trick is deciding what value to enter for N. Follow this logic:

- For nonlinear regression, the number of degrees of freedom (df) equals the number of data points minus the number of variables fit. In this example, there were 15 data points, and three variables were fit. So there are 12 df.
- For an ordinary t test, the number of df for each sample equals one less than the number of data points.
- The t test calculations are based on the numbers of degrees of freedom. There is no way to enter degrees of freedom into Prism – instead you enter N. Prism is programmed to always compute the df as N-1. When comparing the results of nonlinear regression, enter N as the number of degrees of freedom plus 1. Prism will subtract 1, and make the df correct. In this example, enter $N = 12 + 1 = 13$.

To learn how to perform the t test using Prism, see "Comparing groups (t tests, ANOVA, etc.)" on page 187. Here are the results for the example:

	X Labels	A
	Parameter	Value
	X	Y
1	Table Analyzed	Data Table-2 Columns A and B
2		
3	Unpaired t test	
4	P value	0.0309
5	P value summary	*
6	Are means signif. different? (alpha= 0.05)	Yes
7	One- or two-tailed P value?	Two-tailed
8	t, df	t=2.292 df=24
9		
10	How big is the difference?	
11	Mean ± SEM of column A	-6.080 ± 0.06771 N=13
12	Mean ± SEM of column B	-6.290 ± 0.06167 N=13
13	Difference between means	0.2100 ± 0.09159
14	95% confidence interval	-0.3990 to -0.02097
15	R squared	0.1797

The P value is 0.0309. If the treatment really didn't alter the EC50, there is only a 3.09% chance that you would observe this large of a difference (or more) by coincidence. Since the P value is so low, you conclude that the two EC50 values are statistically significantly different.

Notes:

- This method only uses data from one experiment. The SE value is a measure of how precisely you have determined the logEC50 in this one experiment. It is not a measure of how reproducible the experiment is.

Despite the impressive P value, I wouldn't trust these results until the experiment is repeated.

- The t test assumes that the uncertainty in the values of the variables follows a Gaussian distribution. This assumption is not necessarily true with the SE values that emerge from nonlinear regression. The only way to assess the validity of this assumption is to simulate many sets of data, fit each with nonlinear regression, and examine the distribution of best-fit values. I have done this informally with many commonly used equations, and it seems that the assumption is reasonable in many cases.
- Compare LogEC50, not EC50. You want to express the variables in a form that makes the uncertainty as symmetrical and Gaussian as possible. Since a competitive binding curve (similar to a dose response curve) is conducted with X values (concentration) equally spaced on a log scale, the uncertainty of X is reasonably symmetrical (and perhaps Gaussian) when expressed on a log scale. It is equally likely that the observed logKi is 0.1 log units too high or 0.1 log units too low. In contrast, the uncertainty in Ki is not symmetrical.

3. Compare the results within one experiment. More complicated approach.

The method of the previous section only compared the value of the logEC50. This section describes a more general method to compare entire curves to ask whether the data sets differ at all. The idea is to first fit the two curves separately, and then combine the values and fit one curve to all the data.

Follow these steps:

1. Fit the two data sets separately. We did this in the previous section.
2. Total the sum-of-squares and df from the two fits. For this example the total sum of squares equals $19560 + 29320 = 48880$, and the total df equals $12 + 12 = 24$. Since these are the results of fitting the two data sets separately, label these values $SS_{separate}$ and $DF_{separate}$.
3. Combine the two data sets into one. For this example, the combined data set has 30 XY pairs, with each X value appearing twice.
4. Fit the combined data set to the same equation. Note the SS and df. For this example, $SS = 165200$, and $df = 27$ (30 data points minus three variables). Call these values $SS_{combined}$ and $Df_{combined}$.
5. You expect $SS_{separate}$ to be smaller than $SS_{combined}$ even if the curves are really identical simply because the separate fits have more degrees of freedom. The question is whether the SS values are more different than

you'd expect to see by chance. To find out, calculate the F ratio using this equation:

$$F = \frac{\left(SS_{combined} - SS_{separate}\right) / \left(DF_{combined} - DF_{separate}\right)}{SS_{separate} / DF_{separate}}$$

For this example, F = 14.03.

6. Determine the P value from F. There are $DF_{combined}$ - $DF_{separate}$ degrees of freedom in the numerator, and $DF_{separate}$ degrees of freedom in the denominator. GraphPad StatMate and InStat can calculate the P value (from F and the two df values), or you can use tables in the back of most statistics books.
7. For this example, the P value is <0.0001. If the treatment were really ineffective, there is less than a 0.01% chance that the two curves would differ as much (or more) than observed in this experiment. Since the P value is low, you'll conclude that the curves are really different.

Notes:

- This method only uses data from one experiment. Despite the impressive P value, I wouldn't make a strong conclusion until the experiment is repeated.
- This method compares the curves overall. It doesn't tell you which variable(s) are different. Differences might be due to something trivial like a different baseline, rather than something important like a different rate constant.

Advantages and disadvantages of the three methods

If you have repeated the experiment several times, I recommend that you use the first method. There are two advantages:

- Compared to the other methods, this method is far easier to understand and communicate to others.
- The entire test is based on the consistency of the results between repeat experiments. Since there are usually more causes for variability between experiments than within experiments, it makes sense to base the comparison on differences between experiments.

The disadvantage of the first method is that you are throwing away information. The calculations are based only on the best-fit value from each experiment, and ignore the SE of those values presented by the curve fitting program.

If you have performed the experiment only once, then you probably ought to repeat the experiment. Regardless of what statistical results you obtain, you

shouldn't trust results from a single experiment. If you want to compare results in a single experiment, you can use method 2 or method 3.

The advantage of method 2 is that it focuses your thinking on a single variable. Generally, you care mostly about one variable (i.e. a rate constant or EC_{50}), and care less about the others. Method 2 compares the variable of interest.

Method 3 is the most general method. Since the method compares the entire curve, it does not force you to decide which variable(s) you wish to compare. This is both its advantage and disadvantage.

Choosing or entering equations

Prism and equations

When generating a theoretical curve or performing nonlinear regression, you must select or enter an equation.

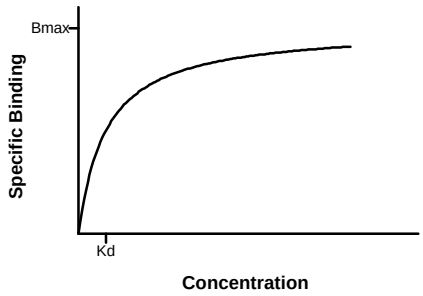
Many equations are built-in to Prism. These equations are commonly used in science, and are especially well suited to analysis of biochemical or pharmacological data. The next section lists these equations and describes their usefulness.

You may also enter your own equations into Prism's versatile equation editor. See "Entering your own equation" on page 295.

Built-in equations

One site binding (hyperbola)

$$Y = \frac{B_{max} \cdot X}{K_d + X}$$



This curve is known as a rectangular hyperbola, binding isotherm, or saturation binding curve. Y is zero initially, and increases to a maximum plateau value B_{max}.

This equation describes the equilibrium binding of a ligand to a receptor as a function of increasing ligand concentration. X is the concentration of the ligand, and Y is the specific binding. B_{max} is the maximum number of binding sites, expressed in the same units as the Y axis (I. e. radioactive counts per minute, sites per cell, or fmol of receptor per mg of tissue). K_d is the equilibrium dissociation constant, expressed in the same units as the X axis (concentration). When the drug concentration equals K_d, half the binding sites are occupied at equilibrium.

Note that Y should be the specific binding, not the total binding. If you wanted to fit total binding, write a user-defined equation with the term E*X added to the equation. To learn how Prism can subtract nonspecific from total binding, see "Subtracting (or dividing by) baseline values" on page 157.

This equation also describes the activity of an enzyme as a function of substrate concentration. In this case, B_{max} is really V_{max}, the maximum enzyme activity. K_d is really K_m, the Michaelis-Menton constant.

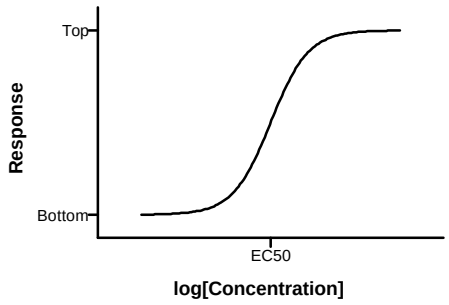
Two site binding

$$Y = \frac{B_{max1} \cdot X}{K_{d1} + X} + \frac{B_{max2} \cdot X}{K_{d2} + X}$$

This equation is an extension of the one site binding curve. It shows the binding of a ligand to two receptors with different affinities. It also describes the enzyme activity as a function of substrate concentration when two isozymes are present.

Sigmoidal dose-response

$$Y = Bottom + \frac{(Top - Bottom)}{1 + 10^{LogEC50 - X}}$$



This is a general equation for a dose-response curve. It shows response as a function of the logarithm of concentration. X is the logarithm of agonist concentration and Y is the response.

BOTTOM is the Y value at the bottom plateau; TOP is the Y value at the top plateau, and LogEC50 is the X value when the response is halfway between BOTTOM and TOP. LogEC50 is the logarithm of the EC_{50}, the concentration that gives a response halfway between BOTTOM and TOP. This variable is sometimes called ED_{50} or IC_{50}.

This equation assumes a standard slope, where the response goes from 10% to 90% of maximal as X increases over about two log units. The next equation allows for a variable slope.

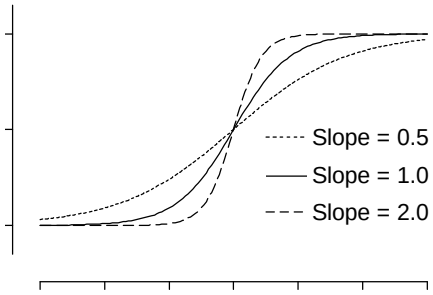
Sigmoidal dose-response (variable slope)

$$Y = Bottom + \frac{(Top - Bottom)}{1 + 10^{(LogEC50 - X)\cdot HillSlope}}$$

This equation extends the previous equation, but allows for a variable slope. This equation is also called a four-parameter logistic equation.

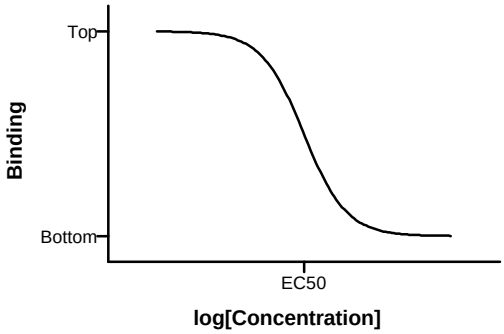
The variable HILLSLOPE controls the slope of the curve. This variable, which is unitless, is called the Hill slope, the slope factor, or the Hill coefficient. If it is positive, the curve increases as X increases. If it is negative, the curve decreases as X increases. When HILLSLOPE = 1, this equation generates a standard dose-response curve, identical to the previous equation. When HILLSLOPE is less than 1.0, the curve is more shallow. When HILLSLOPE is greater than 1.0, the curve is steeper.

BOTTOM is the Y value at the bottom plateau; TOP is the Y value at the top plateau, and LogEC50 is the X value when the response is halfway between BOTTOM and TOP. LogEC50 is the logarithm of the EC_{50}, the concentration that gives a response halfway between BOTTOM and TOP. This variable is sometimes called ED_{50} or IC_{50}.



One site competition

$$Y = Bottom + \frac{(Top - Bottom)}{1 + 10^{X - LogEC50}}$$



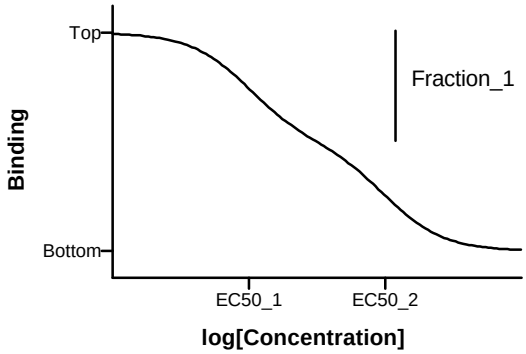
This equation describes the competition of a ligand for receptor binding. It is identical to the sigmoidal dose-response curve with HILLSLOPE = -1.0.

The variable LogEC50 is the concentration of the competitor required to compete for half the specific binding. We use the term EC50 to be consistent with the equations for the other sigmoidal curves. The term IC50 is used more frequently ("E" stands for effective; "I" stands for inhibitory).

If you enter specific, as opposed to total, binding, you may want to define BOTTOM to be a constant equal to zero. If you enter percent specific binding, you may want to set TOP to be a constant equal to 100.

Two site competition

$$Y = Bottom + (Top - Bottom)\left[\frac{Fraction\_1}{1 + 10^{X - LogEC50\_1}} + \frac{1 - Fraction\_1}{1 + 10^{X - LogEC50\_2}}\right]$$

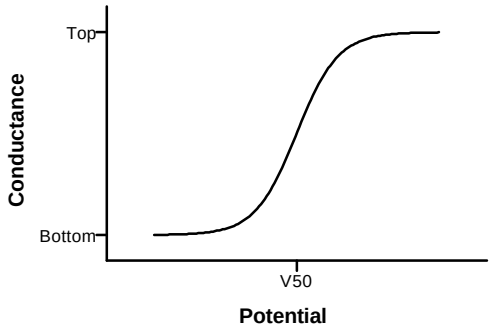


This equation describes the competition of a ligand for two types of receptors. The radioligand has identical affinities for both receptors, but the competitor has a different affinity for each.

Y is binding (total or specific) and X is the logarithm of the concentration of the unlabeled ligand. FRACTION_1 is the fraction of the receptors that have an affinity described by LogEC50_1. The remainder of the receptors have an affinity described by LogEC50_2. If LogEC50_1 is smaller than LogEC50_2, then Fraction_1 is the fraction of high affinity sites. If LogEC50_1 is larger than LogEC50_2, then Fraction_1 is the fraction of low affinity sites.

Boltzmann sigmoid

$$Y = Bottom + \frac{(Top - Bottom)}{1 + \exp\left(\frac{V50 - X}{Slope}\right)}$$



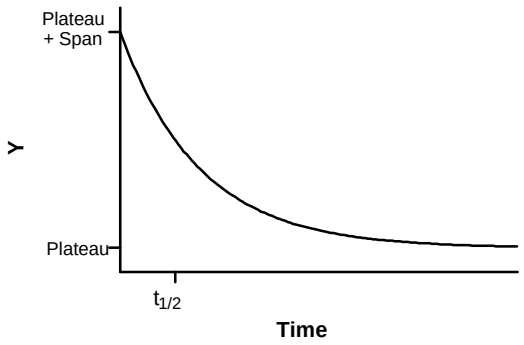
This equation describes voltage dependent activation of ion channels. It describes conductance (Y) as a function of the membrane potential (X). With a large negative potential, conductance equals BOTTOM. With a large positive potential, the conductance equals TOP. V50 is the potential where conductance is halfway between BOTTOM and TOP.

For channels that activate upon depolarization, SLOPE is positive. Under appropriate experimental conditions, you can use SLOPE to calculate the valence (charge) of the ion moving across the channel. SLOPE equals RT/zF, where R is the universal gas constant, T is temperature in °K, F is the Faraday, and z is the valence. Since RT/F ≈ -26 mV at 25°C, z = -26/SLOPE. If SLOPE has a larger value, the curve is more shallow.

BOTTOM is commonly made a constant equal to 0.0. If you also make TOP a constant equal to 1.0, then Y can be viewed as the fraction of channels that are activated.

One phase exponential decay

$$Y = Span \cdot e^{-K \cdot X} + Plateau$$



This equation describes the kinetics such as the decay of radioactive isotopes, the elimination of drugs, and the dissociation of a ligand from a receptor.

X is time, and Y may be concentration, binding, or response. Y starts out equal to SPAN + PLATEAU and decreases to PLATEAU with a rate constant K. The half-life of the decay is 0.6932/K. SPAN and PLATEAU are expressed in the same units as the Y axis. K is expressed in the inverse of the units used by the X axis. In many circumstances, plateau equals zero. When fitting data to this equation, consider fixing plateau to a constant value of zero.

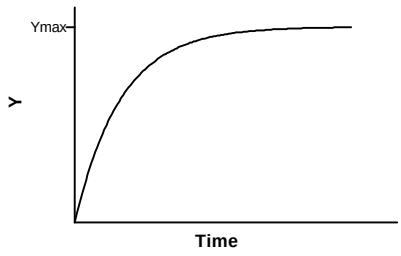
Two phase exponential decay

$$Y = Span1 \cdot e^{-K_1 \cdot X} + Span2 \cdot e^{-K_2 \cdot X} + Plateau$$

This equation describes a two phase exponential decay. Y starts out equal to Span1 + Span2 + PLATEAU and decays to PLATEAU with fast and slow components. The two half-lives are 0.6932/K1 and 0.6932/K2.

One phase exponential association

$$Y = Ymax \cdot \left(1 - e^{-K \cdot X}\right)$$



This equation describes the pseudo-first order association kinetics of the interaction of a ligand and its receptor, or a substrate and an enzyme. Y is binding or enzyme activity. X is time.

Y starts out equal to zero and increases to a maximum plateau equal to YMAX. When X equals 0.6932/K, Y equals 0.5*YMAX.

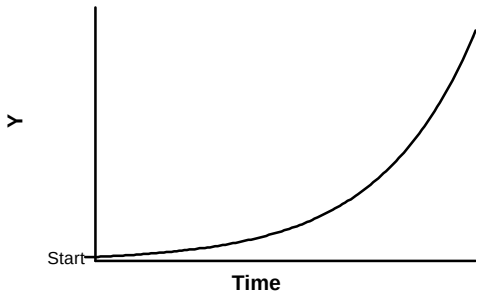
Two phase exponential association

$$Y = Ymax_1 \cdot \left(1 - e^{-K_1 \cdot X}\right) + Ymax_2 \cdot \left(1 - e^{-K_2 \cdot X}\right)$$

This is an extension of the exponential association to two phases.

Exponential growth

$$Y = Start \cdot e^{K \cdot X}$$



This describes an exponential growth curve. Y is population size (perhaps cell number) and X is time. Y starts out equal to START and increases geometrically with a doubling time equal to 0.6932/K.

Note: It is difficult to fit data to this equation with nonlinear regression, because a tiny change in the initial values will drastically alter the sum-of-squares. You may need to override the initial values provided by Prism.

Power series

This versatile equation has many uses.

$$Y = A \cdot X^B + C \cdot X^D$$

Polynomial equations

Prism offers first, second, third and fourth order polynomial equations. Although few chemical or pharmacological models are described by polynomial equations, these equations are often used to fit standard curves. The higher order equations have more inflection points.

Unlike all other equations, you don't have to worry a bit about initial values when fitting data to polynomial equations. You will get exactly the same answer no matter what the initial values are.

First order: $Y = A + B \cdot X$

This equation defines a straight line. Choosing this equation in nonlinear regression is exactly the same as performing linear regression. There are three advantages to using this equation as part of nonlinear regression:

- You can simultaneously fit the data to a line and to another equation. Prism will compare the two with an F test.
- Prism can create a summary table of slopes or intercepts for each data set.
- You can choose to weight the values so as to minimize the sum of the square of the relative distances of the points from the line rather than the absolute distances.

Second order: $Y = A + B \cdot X + C \cdot X^2$

Third order: $Y = A + B \cdot X + C \cdot X^2 + D \cdot X^3$

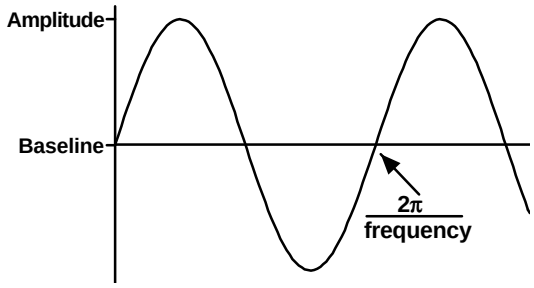
Fourth order: $Y = A + B \cdot X + C \cdot X^2 + D \cdot X^3 + E \cdot X^4$

If you want to use a higher order polynomial equation, you can enter it as a user-defined equation.

Sine wave

$$Y = Baseline + Amplitude \cdot sin(Frequency \cdot X + Offset)$$

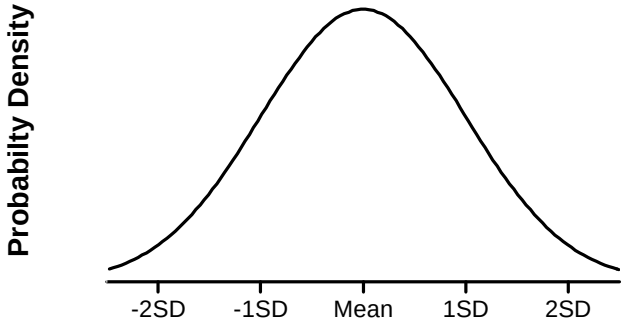
X is in radians.



Gaussian distribution

$$Y = \left(\frac{Area}{SD\sqrt{2p}} \right) e^{-0.5\left(\frac{X - Mean}{SD}\right)^2}$$

Cumulative probability distribution of a Gaussian bell-shaped distribution with specified mean and SD. The area under the entire curve is AREA. A standard probability distribution is scaled so that AREA equals 1.0.



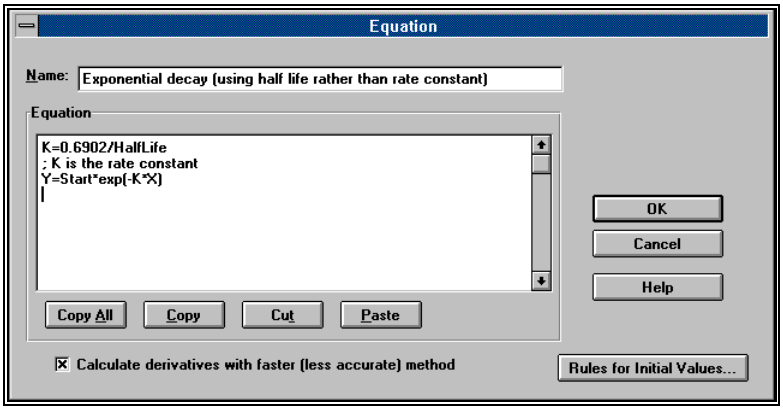
Entering your own equation

Entering equations

Prism's built-in equations are commonly used in many fields of science. If you need to fit data to another equation, you may add it to Prism's list.

To enter your own equation:

1. On the nonlinear regression (or generate curve) Parameters dialog, select "User-defined equations".
2. Prism shows a list of user-defined equations already entered. If you haven't entered any equations yet, the list is empty.
3. Press the button labeled "Add".
4. On the Equation dialog, enter a descriptive name. Then enter the equation itself.



Rather than starting from scratch, you might find it helpful to edit one of the built-in equations. To do this, copy a built-in equation to the clipboard and then paste it into the Equation dialog.

Tip: It is not possible to simultaneously fit to both a built-in equation and a user-defined equation. Instead copy a built-in equation to the user-defined list.

To copy and paste a built-in equation:

1. Start from the Nonlinear Regression or Generate Equation Parameters dialog.
2. Select a built-in equation, and press the button labeled "Equation".
3. Press the button labeled "Copy All".
4. Cancel from that dialog.
5. Select user-defined equations. Then press "Add".
6. Enter an equation name. Then move the insertion point to the Equation block.
7. Press the button labeled "Paste".
8. Edit the equation.

Entering one line equations

Enter a one-line equation using syntax similar to the Basic computer language. Start with "Y=" then enter the rest of the equation. Note the following:

- Use an asterisk (*) to indicate multiplication and a caret (^) to indicate power. You can also indicate multiplication implicitly with a space or parentheses.

- Use parentheses as necessary. To increase readability, use brackets [like this] or braces {like this} as well as parentheses. Prism interprets parentheses, brackets, and braces identically.
- Prism does not distinguish between upper and lower case letters.
- Variable names must not be longer than 13 characters.
- If you want to use two words to describe a variable, separate with the underscore character, for example Half_Life. Don't use a space, hyphen or period.

Three examples:

```
Y=Bmax*X/(Kd + X)
Y=A(X+1)
Y=Start*exp[(-0.693/Half_Life)*K]
```

Using intermediate variables in multiline equations

Use intermediate variables to simplify longer equations. Note the following:

- Define an intermediate variable before you use it.
- To enter comments in your equation, type a semicolon (;) and then add comments. You may add a comment after an equation or on a line by itself.
- To enter a long line, type a backslash (\) at the end of the first line, then press RETURN and continue. Prism treats the two lines as one.
- The last line in the equation (excepting comments) must start with "Y=".

Example:

```
K=0.693/HalfLife ;rate constant
Y=Start*exp(-K*X)
```

Because K is on the left of the equals sign, Prism recognizes that it is an intermediate variable rather than a variable to be fit by nonlinear regression. In contrast, HalfLife and START are not defined within the equation, so Prism recognizes that they are true variables to be fit by nonlinear regression.

Functions you can use in user-defined equations

Function	Explanation
ABS(k)	Absolute value. If K is negative, multiply by -1.
ARCCOS(k)	Arccosine. Result is in radians.
ARCSIN(k)	Arcsine. Result is in radians.
ARCTAN(k)	Arctangent. Result is in radians.

COS(k)	Cosine. k is in radians.
DEG(R)	Converts R radians to degrees.
EXP(k)	e to the kth power.
IF(condition, J, K)	If the condition is true, then the result is J. Otherwise the result is K. The condition is in the form A<B or A=B.
INT(k)	Truncate fraction. INT(3.5)=3 INT(-2.3) = -2
LN(K)	Natural logarithm.
LOG(K)	Log base 10.
MAX(J,K)	Maximum of two values.
MIN(J,K)	Minimum of two values.
J MOD K	The remainder (modulus) after dividing J by K.
RAD(D)	Converts D degrees to radians.
SGN(K)	If K>0, SGN(K)=1. If K<0, SGN(K)=-1. If K=0, SGN(K)=0.
SIN(K)	Sine. K is in radians.
SQRT(K)	Square root.
TAN(K)	Tangent. K is in radians.

When you enter your equations, you can use any of the functions listed above. You cannot use any of those names for your variables.

Most of those functions are self-explanatory. Later sections give more information on the MIN, MAX, and IF functions.

Constraining variables with MAX and MIN

You can use the MIN and MAX functions to constrain a variable. In this example, the MAX function constrains the variable BOTTOM to a positive value.

```
Bottom=MAX(BASELINE,0.0)
;Don't let BOTTOM become negative
EXTENT=TOP - BOTTOM
Y=EXTENT*exp(-K*X) + BOTTOM
```

Look at the equation one line at a time.

1. The variable BOTTOM is set to the larger of the variable BASELINE and zero. This means that the variable BOTTOM can never be negative (because 0.0 is larger than any negative number). Note that Prism is indifferent to case: “Bottom” and “BOTTOM” are the same.

2. Since Bottom is to the left of the equals sign, it must be an intermediate variable. Since the value of an intermediate variable is defined from other variables, it is not fit by nonlinear regression.
3. Since BASELINE has not been defined, it is a true variable. It is either a constant whose value you will enter, or a variable that Prism will fit.
4. This is a comment. You can add comments anywhere by typing a semicolon and then entering your notes.
5. The variable EXTENT is defined as the difference between the variables TOP and BOTTOM.
6. Since EXTENT is to the left of the equals sign, it is an intermediate variable. We already know that BOTTOM is also an intermediate variable. Since TOP has not yet been defined, it must be a true variable.
7. Y is defined. The last line must start with "Y=" (comments may come later).
8. The equation has five variables, plus X and Y. EXTENT and BOTTOM are intermediate variables used to make the equation easier to follow. TOP, BASELINE, and K are true variables. Either you set them to constant values or Prism finds the best-fit values.

IF-THEN relationships

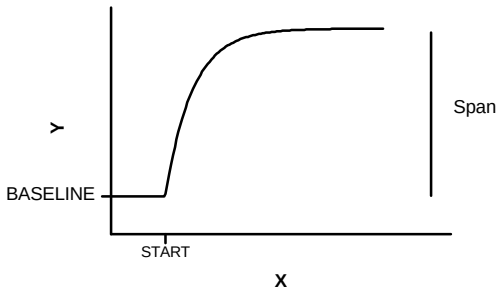
If you have done any programming, you already know that IF-THEN statements are very powerful as they allow branching logic. Prism implements this kind of logic using the IF function, which is very similar to IF functions used in spreadsheet programs.

IF is a function with three arguments. The syntax is:

```
IF (conditional expression, value if true, value if false)
```

The conditional expression has a form like "X<A" or "Ymax=Constraint" or "X<=A and X>=B". See the next section for more information on conditional expressions.

Example. The example below describes the curve shown in the figure.




```
DELTA_X=X-START
INCREASE=IF[(X<START),0.0,SPAN*(1-exp(-K*DELTA_X))]
;Before X=START, INCREASE is zero
;Then it is as an exponential association
Y=BASELINE+INCREASE
```

Here is a line by line description.

1. DELTA_X is defined as the difference between X and the variable START. It is the time elapsed since START. Since it is to the left of the equals sign, DELTA_X is an intermediate variable. Since START has not been defined, START is a true variable.
2. This statement says that the variable INCREASE equals zero when X is less than START. Otherwise INCREASE is set equal to SPAN*(1-exp(K*Delta_X)). The exponential increase starts when X=START. Before that time, INCREASE equals zero.
3. Comment. Comments always begin with semicolons.
4. Another comment. Use comments to remind yourself what the equations do.
5. Define Y as the sum of BASELINE and INCREASE. Since BASELINE hasn't been defined yet, it is a true variable.

This equation has one intermediate variable (INCREASE) and four true variables whose value can be fit: START, SPAN, K, and BASELINE.

Conditional expressions

When you use the IF function, you must enter a conditional expression as the first argument. If the expression is true, then the function evaluates to the second argument. Otherwise the function evaluates to the third argument.

A conditional expression compares two quantities, so you must use one of the following:

Symbol	Meaning
>	Greater than
<	Less than
>=	Greater or equal to
<=	Less than or equal to
=	Equal to
<>	Not equal to

You may connect two expressions with AND or OR, or precede an expression with NOT. Examples:

```
(A<B or A<C)
NOT(A<B AND A<C)
MAX>100
FRACTION<>1.0
```

The IF function requires a conditional expression. You may also use a conditional expression anywhere else in an equation. A conditional expression evaluates as 1.0 if true and 0.0 if false. Example:

```
Y=(X<4)*1 + (X>=4)*10
```

When X is less than 4, this evaluates to 1*1 + 0*10=1. When X is greater than 4, this evaluates to 0*1+1*10=10.

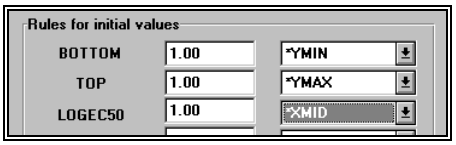
Defining rules for initial values in nonlinear regression

Before it can perform nonlinear regression, Prism must have initial values for each variable in the equation. If you define rules for generating the initial values, Prism automatically calculates the initial values. Otherwise, you must enter the initial values for every variable, for every data set, every time you fit data.

If you plan to use a user-defined equation more than once, we suggest that you define rules to generate initial values so you don't need to worry about initial values every time you use the equation.

To define rules for initial values for user-defined equations:

1. While entering or editing the user-defined equation, click on the button labeled "Rules for initial values".
2. On the Default Values dialog, enter the rule for finding the initial value of each variable. For each variable in the equation, enter a number in the first column and select a multiplier from the drop-down list in the second column.



All but one choice on the drop-down list are used to multiply or divide the number you entered by a value determined from range of the data: YMIN, YMAX, YMID, XMIN, XMAX, XMID, XMID/YMID, XMAX-XMIN, or YMAX-YMIN. The abbreviation YMIN is the minimum value of Y; YMAX is the maximum value, and YMID is the average of YMIN and YMAX. For example, if you enter "0.5" in the first column and select "YMAX" in the second column, Prism determines YMAX for each data set, and sets the initial value of the variable to half that value.

The first choice on the drop-down list is "(value)". This means that the value you entered is the initial value in all cases. The initial value does not depend on the range of the data.

Note: You can override the initial values provided by the rules. See *Initial values for nonlinear regression on page 261.*

Equations you cannot enter into Prism

Prism gives you a great deal of flexibility in entering equations. However, you cannot enter these kinds of equations:

- No implicit equations. Y must be defined as a function of X and one or more variables. The last line of the equation must start with "Y=", and Y must not be used elsewhere in the equation.
- No differential equations. You must define Y as a function of X and one or more variables. It is not sufficient to define the derivatives.
- No equations with more than one X variable. Prism does not calculate multiple regression.
- No equations with more than 16 variables.
- No discontinuous equations. Although you can enter discontinuous equations, the results of nonlinear regression may not be reliable. An equation is discontinuous when an infinitesimal change in X can create a huge change in Y.

Transferring equations with data files

Prism maintains a list of user-defined equations. Whenever you fit data, you can choose from equations you used before.

What happens when you want to transfer a file to another computer? There are no explicit commands to import or export equations. Prism handles the situation automatically. When you open a file that includes fits to user-defined equations, Prism uses these rules:

- Prism first reads the name of the equation from the file.
- If an equation with exactly the same name already exists in the equation list, the equation is read from the list even if it is different than the one saved on the file. This allows you to use a revised equation with stored data.
- If an equation with the same name has not already been stored, then Prism reads the equation from the file and adds it to the equation list.

Analyzing radioligand binding data

Introduction to radioligand binding

A radioligand is a radioactively labeled drug that can associate with a receptor, transporter, enzyme, or any protein of interest. Measuring the rate and extent of binding provides information on the number of binding sites, and their affinity and accessibility for various drugs.

There are three kinds of experimental protocols:

- Saturation binding experiments. Measure equilibrium binding of various concentrations of the radioligand. Analyze the relationship between binding and ligand concentration to determine the number of sites, Bmax, and ligand affinity, Kd.
- Competitive binding experiments. Measure equilibrium binding of a single concentration of radioligand at various concentrations of an unlabeled competitor. Analyze these data to learn the affinity of the receptor for the competitor.
- Kinetic experiments. Measure binding at various times to determine the rate constants for radioligand binding and dissociation.

Prism makes it easy to analyze and display data from all three kinds of experiments.

For more information on analysis of radioligand binding data, see:

- LE Limbird, Cell surface receptors: A short course on theory and methods, Martinus Nijhoff Publishers, 1986. (Second edition due in 1996.)
- HI Yamamura, et al, Methods in Neurotransmitter receptor analysis, Raven Press, 1990.

Radioactivity

Efficiency of detecting radioactivity

Efficiency is the fraction of the radioactive disintegration that are detected by the counter. Determine efficiency by counting a standard sample under conditions identical to those used in your experiment.

With ^{125}I, efficiencies are very high, over 90%. It depends on the geometry of the counter. The detector doesn't entirely surround the tube, so a few gamma rays (photons) miss the detector.

With ^{3}H, the efficiency of counting is much lower, often about 40%. The low efficiency is mostly a consequence of the physics of decay, and can not be improved by better instrumentation or better scintillation fluid. When a tritium atom decays, a neutron converts to a proton and the reaction shoots off an electron and neutrino. The energy released is always the same, but it is randomly partitioned between the neutrino (not detected) and an electron (that we try to detect). When the electron has sufficient energy, it will travel far enough to encounter a fluor molecule in the scintillation fluid. This fluid amplifies the signal and gives of a flash of light detected by the scintillation counter. The intensity of the flash (number of photons) is proportional to the energy of the electron. If the electron has insufficient energy, it is not captured by the fluor and is not detected. If it has low energy, it is captured but the light flash has few photons and is not detected by the instrument. Since the decay of many tritium atoms does not lead to a detectable number of photons, the efficiency of counting is less than 100%.

Efficiency of counting ^{3}H is reduced by the presence of any color in the counting tubes, if the mixture of water and scintillation fluid is not homogeneous, or if the radioactivity is trapped in tissue chunks (so emitted electrons don't travel into the scintillation fluid).

Specific radioactivity

When you buy radioligands, the packaging usually states the specific radioactivity as Curies per millimole (Ci/mmole). Since you measure counts per minute (cpm), the specific radioactivity is more useful when stated in terms of cpm. Often the specific radioactivity is expressed as cpm/fmol (1 fmol = 10^{-15} mole).

To convert from Ci/mmol to cpm/fmol, you need to know that 1 Ci equals 2.22×10^{12} disintegrations per minute. Use this equation to convert Z Ci/mmole to Y cpm/fmol when the counter has an efficiency (expressed as a fraction) equal to E.

$$Y\frac{\text{cpm}}{\text{fmol}} = Z\ \frac{\text{Ci}}{\text{mmole}} \cdot 2.22\text{x}10^{12}\ \frac{\text{dpm}}{\text{Ci}} \cdot 10^{-12}\ \frac{\text{mmole}}{\text{fmole}} \cdot E\frac{\text{cpm}}{\text{dpm}} = Z \cdot 2.22 \cdot E$$

$$Y\frac{\text{cpm}}{\text{fmol}} = Z\frac{\text{Ci}}{\text{mmole}} \cdot 2.22 \cdot E$$

You may also encounter the unit Becquerel, which equals one radioactive disintegration per second.

Calculating the concentration of the radioligand

Rather than trust your dilutions, you can accurately calculate the concentration of radioligand in a stock solution. Measure the number of counts per minute in a small volume of solution and use this equation. C is cpm counted, V is volume of the solution you counted in ml, and Y is the specific activity of the radioligand in cpm/fmol (calculated in the previous section).

$$\text{Concentration in pM} = \frac{\text{C cpm} \,/\, \text{Y cpm / fmol}}{\text{V ml}} \cdot \frac{0.001\ \text{pmol / fmol}}{0.001\ \text{liter / ml}} = \frac{C / Y}{V}$$

Radioactive decay

Radioactive decay is entirely random. A particular atom has no idea how old it is, and can decay at any time. The probability of decay at any particular interval is the same as the probability of decay during any other interval. If you start with N_0 radioactive atoms, the number remaining at time t is:

$$N_t = N_0 \cdot e^{-K_{decay} \cdot t}$$

K_{decay} is the rate constant of decay expressed in units of inverse time. Each radioactive isotope has a different value of K_{decay}.

The half-life ($t_{½}$) is the time it takes for half the isotope to decay. Half-life and decay rate constant are related by this equation:

$$t_{1/2} = \frac{\ln(2)}{K_{decay}} = \frac{0.693}{K_{decay}}$$

This table shows the half-lives and rate constants for commonly used radioisotopes. The table also shows the specific activity assuming that each molecule is labeled with one isotope. (This is often the case with ^{125}I and ^{32}P. Tritiated molecules often incorporate two or three tritium atoms, which increases the specific radioactivity.)

Isotope	Half life	Kdecay	Specific Radioactivity
^{3}H	12.43 years	0.056 $year^{-1}$	28.7 Ci/mmol
^{125}I	59.6 days	0.0116 day^{-1}	2190 Ci/mmol
^{32}P	14.3 days	0.0485 day^{-1}	9128 Ci/mmol
^{35}S	87.4 days	0.0079 day^{-1}	1493 cI/mmol

You can calculate radioactive decay from a date where you knew the concentration and specific radioactivity using this equation.

$$\text{Fraction Remaining} = e^{-K_{decay} \cdot Time}$$

For example, after ^{125}I decays for 20 days, the fraction remaining equals 79.5%. Although data appear to be scanty, most scientists assume that the energy released during decay destroys the ligand so it no longer binds to receptors. Therefore the specific radioactivity does not change over time. What changes is the concentration of ligand. After 20 days, therefore, the concentration of the iodinated ligand is 79.5% of what it was originally, but the specific radioactivity remains 2190 Ci/mmol. This approach assumes that the unlabeled decay product is not able to bind to receptors and has no effect on the binding. Rather than trust this assumption, you should always try to use newly synthesized or repurified radioligand for key experiments.

The Poisson distribution

The decay of a population radioactive atoms is random, and therefore subject to a sampling error. For example, the radioactive atoms in a tube containing 1000 cpm of radioactivity won't give off exactly 1000 counts in every minute. There will be more counts in some minutes and fewer in others, with the distribution of counts following a Poisson distribution. This variability is intrinsic to radioactive decay and cannot be reduced by more careful experimental controls.

After counting a certain number of counts in your tube, you want to know what the "real" number of counts is. Obviously, there is no way to know that. But you can calculate a range of counts that is 95% certain to contain the true average value. So long as the number of counts, C, is greater than about 50 you can calculate the confidence interval using this approximate equation:

$$\text{95\% Confidence Interval: } \left(C - 1.96\sqrt{C}\right) \text{ to } \left(C + 1.96\sqrt{C}\right)$$

GraphPad StatMate does this calculation for you using a more exact equation that can be used for any value of C. For example, if you measure 100 radioactive counts in an interval, you can be 95% sure that the true average num-

ber of counts ranges approximately between 80 and 120 (using the equation here) or more exactly between 81.37 and 121.61 (using StatMate).

When calculating the confidence interval, you must set C equal to the total number of counts you measured experimentally, not the number of counts per minute.

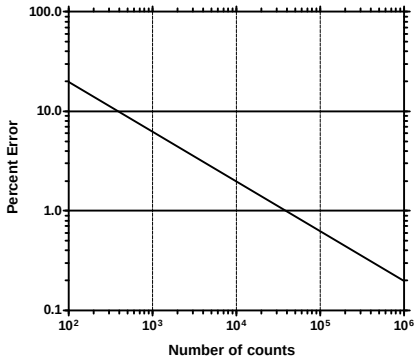
Example: You placed a radioactive sample into a scintillation counter and counted for 10 minutes. The counter tells you that there were 225 counts per minute. What is the 95% confidence interval? Since you counted for 10 minutes, the instrument must have detected 2250 radioactive disintegrations. The 95% confidence interval of this number extends from 2157 to 2343. This is the confidence interval for the number of counts in 10 minutes, so the 95% confidence interval for the average number of counts per minute extends from 216 to 234. If you had attempted to calculate the confidence interval using the number 225 (counts per minute) rather than 2250 (counts detected), you would have calculated a wider (incorrect) interval.

The Poisson distribution explains why it is helpful to count your samples longer when the number of counts is small. For example, this table shows the confidence interval for 100 cpm counted for various times. When you count for longer times, the confidence interval will be narrower.

	1 minute	10 minutes	100 minutes
Counts per minute (cpm)	100	100	100
Total counts	100	1000	10000
95% CI of counts	81.4 TO 121.6	938 TO 1062	9804 to 10196
95% CI of cpm	81.4 to 121.6	93.8 to 106.2	98.0 to 102.0

This graph shows percent error as a function of number of counts (C). Percent error is defined from the width of the confidence interval divided by the number of counts.

$$\text{Percent Error} = 100 \cdot \frac{1.96 \cdot \sqrt{C}}{C}$$



Law of mass action

Analysis of radioligand binding experiments is based on a simple model, called the law of mass action:

$$\textbf{Ligand} + \textbf{Receptor} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \textbf{Ligand-Receptor}$$

Binding occurs when ligand and receptor collide due to diffusion, and when the collision has the correct orientation and enough energy. The rate of association (number of binding events per unit of time) equals [Ligand]·[Receptor]·k_{on} .

Once binding has occurred, the ligand and receptor remain bound together for a random amount of time. The rate of dissociation (number of dissociation events per unit time) equals [ligand·receptor]·k_{off}. The probability of dissociation is the same at every instant of time. The receptor doesn't "know" how long it has been bound to the ligand. After dissociation, the ligand and receptor are the same as at they were before binding. If either the ligand or receptor are chemically modified, then the binding does not follow the law of mass action.

Equilibrium is reached when the rate at which new ligand·receptor complexes are formed equals the rate at which the ligand·receptor complexes dissociate. At equilibrium:

$$[\text{Ligand}] \cdot [\text{Receptor}] \cdot k_{on} = [\text{Ligand} \cdot \text{Receptor}] \cdot k_{off}$$

Rearrange that equation to define the equilibrium dissociation constant K_d.

$$\frac{[\text{Ligand}] \cdot [\text{Receptor}]}{[\text{Ligand} \cdot \text{Receptor}]} = \frac{k_{off}}{k_{on}} = K_d$$

This equation gives you a feel for what K_d means. When the ligand occupies half the receptors, the concentration of unoccupied receptors equals the concentration of occupied receptors: [Receptor] = [Ligand·Receptor].This can only be true when K_d equals [Ligand]. In other words, the K_d is the concentration of ligand which will bind to half the receptors at equilibrium.

Don't mix up K_d, the equilibrium dissociation constant, with k_{off}, the dissociation rate constant. They are not the same, and aren't even expressed in the same units.

Variable	Name	Units
k_{on}	Association rate constant or on rate constant	$M^{-1}min^{-1}$
k_{off}	Dissociation rate constant or off rate constant	min^{-1}
K_d	Equilibrium dissociation constant	M

The law of mass action predicts the fractional receptor occupancy at equilibrium as a function of ligand concentration. Fractional occupancy is the fraction of all receptors that are bound to ligand.

$$\text{Fractional occupancy} = \frac{[\text{Ligand} \cdot \text{Receptor}]}{[\text{Total Receptor}]} = \frac{[\text{Ligand} \cdot \text{Receptor}]}{[\text{Receptor}] + [\text{Ligand} \cdot \text{Receptor}]}$$

This equation is not useful, because you don't know the concentration of unoccupied receptor, [Receptor]. A bit of algebra creates a useful equation. (Multiply both numerator and denominator by [Ligand] and divide both by [Ligand·Receptor]. Then substitute the definition of K_d.) Now the equation is useful:

$$\text{Fractional occupancy} = \frac{[\text{Ligand}]}{[\text{Ligand}] + K_d}$$

This equation assumes equilibrium. To make sense of it, think about a few different values for [Ligand]. When [Ligand] = 0, the occupancy equals zero. When [Ligand] is very high (many times K_d) , the fractional occupancy approaches 100%. When [Ligand] = K_d, fractional occupancy is 50%.

Although termed a "law", the law of mass action is simply a model that can be used to explain some experimental data. Because it is so simple, the model is not useful in all situations.

- The model assumes that all receptors are equally accessible to ligands.
- The model ignores any states of partial binding. According to the model, receptors are either free or bound to ligand. It also doesn't allow for more than one affinity state.
- The model assumes that the ligand is not altered by binding.
- The model assumes that binding is reversible.

Despite its simplicity, the law of mass action has proven to be very useful in describing many aspects of receptor pharmacology.

Just because binding is constant over time does not mean the system is in equilibrium. Other reactions could be happening as well, especially when agonists are used. Many investigators use the term "steady state" to describe binding that has plateaued, and reserve the term "equilibrium" to describe the ideal model.

Total vs. free concentrations of ligand. Ligand depletion.

The equations that describe the law of mass action include the variable [Ligand] which is the free concentration of ligand.

In many experimental situations, you can assume that a very small fraction of the ligand binds to receptors (or to nonspecific sites). In these situations, you can assume that the free concentration of ligand is approximately equal to the concentration you added. This assumption vastly simplifies the analysis of binding experiments, and the standard analysis methods depend on this assumption.

In other situations, a large fraction of the ligand binds to the receptors (or nonspecifically). This means that the concentration of ligand free in the solution does not equal the concentration you added. The discrepancy is not the same in all tubes or at all times. The free ligand concentration is depleted by binding.

Many investigators use this rule of thumb. If less than 10% of the ligand binds, don't worry about ligand depletion. If more than 10% of the ligand binds, you have three choices:

- Change the experimental conditions. The simplest approach is to decrease the amount of receptor in the assay by using less tissue or fewer cells. The problem is that this will decrease the number of radioactive counts. The way around this is to increase the reaction volume without changing the amount of tissue. The problem with this approach is that it requires more radioligand, which is usually very expensive.
- Measure the free concentration of ligand in every tube. This is fairly straightforward if you use centrifugation or equilibrium dialysis, but is difficult if you use vacuum filtration to remove free radioligand.
- Use analysis techniques that adjust for the difference between the concentration of added ligand and the concentration of free ligand. The next chapter explains several such methods.

Analysis of saturation radioligand binding data

Saturation radioligand binding experiments measure specific radioligand binding at equilibrium at various concentrations of the radioligand. Analyze these data to determine receptor number and affinity. Because this kind of experiment used to be analyzed with Scatchard plots (more accurately attributed to Rosenthal), they are sometimes called "Scatchard experiments".

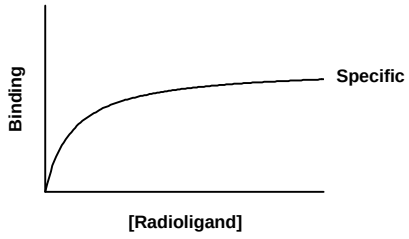
The analyses depend on the assumption that you have allowed the incubation to proceed to equilibrium. This can take anywhere from a few minutes to many hours, depending on the ligand, receptor, temperature, and other experimental conditions. The lowest concentration of radioligand will take the longest to equilibrate. When testing equilibration time, therefore, use a low concentration of radioligand (perhaps 10-20% of the K_D).

Nonspecific binding

In addition to binding to the receptors, radioligand also binds to other sites termed nonspecific sites. Nonspecific binding is detected by measuring radioligand binding in the presence of a saturating concentration of an unlabeled drug that binds to the receptors. Under those conditions, virtually all the receptors are occupied by the unlabeled drug so the radioligand can only bind to nonspecific sites. Subtract the nonspecific binding at a particular concentration of radioligand from the total binding at that concentration to calculate the specific radioligand binding to receptors.

What concentration of unlabeled drug should you use? You want to use enough to block virtually all the specific radioligand binding, but not so much that you cause more general physical changes to the membrane that might alter specific binding. A useful rule-of-thumb is to use the unlabeled compound at a concentration equal to 100 times its K_D for the receptors.

Nonspecific binding is usually (but not always) proportional to the concentration of radioligand (within the range it is used). Add twice as much radioligand, and you'll see twice as much nonspecific binding. The left figure shows a schematic of total and nonspecific binding. The figure on the right shows the difference between total and nonspecific binding – the specific binding.



Subtracting nonspecific binding automatically

Prism can subtract nonspecific from total binding to determine specific binding at each concentration.

To subtract nonspecific binding:

1. Enter concentrations in the X column. Enter total binding for the first condition in column A and the corresponding nonspecific binding in column B. For the second condition, enter total binding in column C, and nonspecific in D. You can enter up to 26 pairs of data sets on one sheet. Enter all values as cpm.
2. Click the Analyze button. From the Data manipulation panel, choose Remove baseline.
3. Don't change the default selections on the dialog. The baseline values are in columns B, D, etc. You want to subtract, not divide, the baseline. Prism will calculate a table with A-B, C-D, E-F, etc.
4. Since nonlinear regression is almost always a linear function of ligand concentration, check "Assume the baseline is linear". In this case, Prism first uses linear regression to find the line that best fits the nonspecific values, and then subtracts the nonspecific values predicted from this line. This option works even if you have not measured nonspecific binding at every concentration. If you leave the option box unchecked, Prism subtracts each nonspecific value from the corresponding total value.
5. Check "Make new graph".
6. Click OK. The specific binding of the first data set will appear in column A of the results table. The specific binding of the second data set (originally columns C and D) will be in column B.

This method is only valid when a small fraction of the ligand binds to the receptor. If this assumption is true, then the free concentration of ligand equals the added concentration in both the tubes used to measure total binding and the tubes used to measure nonspecific binding. If the assumption is not valid, then the free concentration of ligand will differ in the two sets of tubes. In this case subtracting the two values makes no sense, and determining specific binding is difficult. See the next chapter for more details.

Fitting a curve to determine B_{max} and K_d

Prism provides built-in equations for fitting saturation binding curves to one and two binding sites. Both equations assume X is concentration of radioligand (not the logarithm of concentration) and that each Y column contains specific binding data.

To fit a binding curve:

1. Start from a table where X is concentration of ligand in nM or pM and Y is specific binding. This can either be the original data table or a results table.
2. Click Analyze.
3. If you are starting from a results sheet, select "Analyze the results you are looking at".
4. From the curves and regressions section, choose Nonlinear regression.
5. On the nonlinear regression dialog, choose One site binding or Two site binding. Leave everything else at the default settings. The one-site equation is

$$Y = \frac{B_{max} \cdot X}{K_d + X}$$

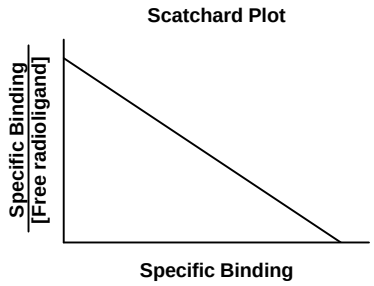
Press OK. You'll see the table of results. B_{max} is expressed in the same units as the Y values (i.e., cpm or fmol/mg). K_d is expressed in the same units as the X values (i.e. nM).

This analysis is based on these assumptions:

- Only a small fraction of the radioligand binds. The free concentration is almost identical to the concentration you added.
- There is no cooperativity. Binding of a ligand to one binding site does not alter the affinity of another binding site.
- Your experiment has reached equilibrium.
- Binding is reversible and follows the law of mass action.

Scatchard plots

In the days before nonlinear regression programs were widely available, scientists transformed data into a linear form, and then analyzed the data by linear regression. There are several ways to linearize binding data, including the methods of Lineweaver-Burke and Eadie-Hofstee. However, the most popular method to linearize binding data is to create a Scatchard plot (more accurately attributed to Rosenthal). In this plot, the X axis is specific binding and the Y axis is specific binding divided by free radioligand concentration.



It is possible to estimate the B_{max} and K_d from a Scatchard plot (B_{max} is the X intercept; K_d is the negative reciprocal of the slope). However, this transformation distorts the experimental error, and thus violates the assumptions of linear regression. The B_{max} and K_d values you determine by linear regression of Scatchard transformed data are likely to be far from their true values. You should analyze your data with nonlinear regression. Do not analyze your data with Scatchard plots. See "Linear regression of transformed data " on page 250.

After analyzing your data with nonlinear regression, however, it is often useful to display data as a Scatchard plot. The human retina and visual cortex are wired to detect edges (straight lines), not rectangular hyperbolas. Scatchard plots are often shown as insets to the saturation binding curves. They are especially useful when you want to show a change in B_{max} or K_d.

When making a Scatchard plot, you have to choose what units you want to use for the Y axis. Some investigators express both free ligand and specific binding in cpm so the ratio bound/free is a unitless fraction. While this is easy to interpret (it is the fraction of radioligand bound to receptors), a more rigorous alternative is to express specific binding in sites/cell or fmol/mg protein, and to express the free radioligand concentration in nM. While this makes the Y axis hard to interpret visually, it provides correct units for the slope (which is -1/K_D).

To transform specific binding data to a Scatchard plot:

1. Start from a table where X is concentration of ligand in nM or pM and Y is specific binding. This can either be the original data table or a results table.
2. Press Analyze.
3. If you are starting from a results sheet, select "Analyze the results you are looking at".
4. From the data manipulations section, choose Transforms.
5. Check the box to interchange X and Y.

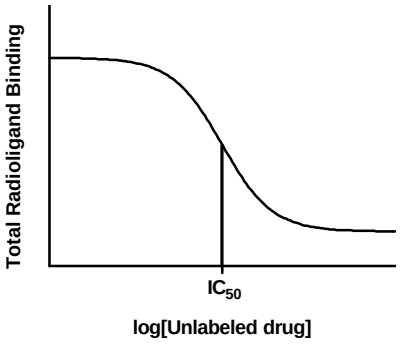
6. Check the box to transform Y and choose the Y transform: Y = X/Y.
7. Check the box to make a graph.
8. Click OK to see the table. Click the Graphs tab to see the graph.

Note: This method will only make a Scatchard plot from the first data set in the data table (because it is not possible to interchange X and Y with more than one Y value). Any other data sets on the table will be ignored. To make a Scatchard plot from other data sets: 1. While looking at the results sheet, press New analysis. 2. Select to duplicate the analysis you are looking at. 3. On the next dialog, choose a different data set to analyze.

Competitive binding data with one class of receptors

What is a competitive binding curve?

Competitive binding experiments measure the binding of a single concentration of labeled ligand in the presence of various concentrations of unlabeled ligand. Typically, the concentration of unlabeled ligand varies over at least six orders of magnitude.



The top of the curve is a plateau at a value equal to radioligand binding in the absence of the competing unlabeled drug. The bottom of the curve is a plateau equal to nonspecific binding. The concentration of unlabeled drug that produces radioligand binding half way between the upper and lower plateaus is called the IC_{50} (inhibitory concentration 50%) or EC_{50} (effective concentration 50%).

If the radioligand and competitor both bind reversibly to a single binding site, binding at equilibrium follows this equation (where Top and Bottom are the Y values at the top and bottom plateau of the curve).

$$Y = Bottom + \frac{(Top - Bottom)}{1 + 10^{X - LogEC50}}$$

Competitive binding experiments are often used in more complicated situations where this equation doesn't apply. See "Shallow competitive binding curves" on page 321 and "Homologous competitive binding curves " on page 334.

Entering data for competitive binding curves

Most investigators enter Y values as cpm. If you performed the experiment in triplicate, enter all three values and let Prism automatically plot the error bars.

Some investigators transform their data to percent specific binding. The problem with this approach is that you need to define how many cpm equal 100% binding and how many equal 0% specific binding. Deciding on these values is usually somewhat arbitrary. It is usually better to enter the data as cpm.

Enter the logarithm of the concentration of competitor into the X column. For example, if the competitor concentration varied from 1 nM to 1 mM, enter the X values from -9 to -3. A log axis cannot accommodate a concentration of zero (log(0) is undefined). Instead, enter a very low competitor concentration (in this example, -10).

If you prefer to enter concentrations, rather than the logarithm of concentration, transform the data before performing nonlinear regression.

To transform the X values from concentration to the log of concentration:

1. Enter the data with X as concentration and Y as binding.
2. Press Analyze.
3. From the data manipulation section, choose Transforms.
4. Check Transform X values.
5. Choose X = log(X).
6. If you want to transform Y from cpm to more useful units, check Transform Y values, choose Y = K*Y, and enter an appropriate value for K.

Decisions to make before fitting the data

Weighting

When analyzing the data, you need to decide whether to minimize the sum of the square of the absolute distances of the points from the curve or to

minimize the sum of the square of the relative distances. The choice depends on the source of the experimental error. Follow these guidelines:

- If the bulk of the error comes from pipetting, the standard deviation of replicate measurements will be, on average, a constant fraction of the amount of binding. In a typical experiment, for example, the highest amount of binding might be 2000 cpm with an SD of 100 cpm. The lowest binding might be 400 cpm with an SD of 20 cpm. With data like this, you should evaluate goodness-of-fit with relative distances. The details on how to do this are in the next section.
- In other experiments, there are many contributions to the scatter and the standard deviation is not related to the amount of binding. With this kind of data, you should evaluate goodness-of-fit using absolute distances, which is the default choice.
- You should only consider weighting by relative distances when you are analyzing total binding data. When analyzing specific binding (or data normalized to percent inhibition), you should evaluate goodness-of-fit using absolute distances, as there is no clear relationship between the amount of scatter and the amount of specific binding.

Constants

To find the EC_{50}, the concentration that blocks 50% of the binding, Prism needs to first define 100% and 0%.

Ideally your data span a wide range of concentrations of unlabeled drug, and clearly define the bottom or top plateaus of the curve. If this is the case, Prism can find the 0% and 100% values from the plateaus of the curve and you don't need to do anything special.

In some cases, your competition data may not define a clear bottom plateau, but you can define the plateau from other data. All drugs that bind to the same receptor should compete all specific radioligand binding and reach the same bottom plateau value. This means that you can define the 0% value (the bottom plateau of the curve) by measuring radioligand binding in the presence of a standard drug known to block all specific binding. If you do this, make sure that you use plenty of replicates to determine this value accurately. If your definition of the lower plateau is wrong, the values for the EC_{50} will be wrong as well. You can also define the top plateau as binding in the absence of any competitor.

If you have collected enough data to clearly define the entire curve, let the program fit all the variables and fit the top and bottom plateaus based on the overall shape of your data. If your data don't define a clear top or bottom plateau, you should define one or both of these values to be constants fixed to values determined from other data.

Fitting data to a one-site competitive binding curve

Follow these steps to fit data to a one-site competitive binding equation:

1. Press Analyze.
2. From the curves section, choose nonlinear regression.
3. Choose the one-site competitive binding equation
4. If you choose to minimize the sum of the relative distances (as percent of Y), click on the Methods option button and choose "Minimize relative distances".
5. If you want to fix the top and bottom plateaus to constant values, click the Constants button and enter the values.
6. Press the button “Ki Calculation”.
7. Enter the concentration of the radioligand you used and its K_d . Enter both values in nM (actually you can use any concentration units, as only the ratio of the two values matters). Enter these values as concentration, not as log(concentration). The Kd must be known from previous saturation binding experiments.

Interpreting competitive binding results

Assumptions

To interpret the results, you must make these assumptions:

- Only a small fraction of your radiolabeled ligand binds to the tissue so that the free concentration is virtually the same as the added concentration. The next chapter gives an alternative method that does not make this assumption.
- There is no cooperativity.
- Your experiment has reached equilibrium.
- Binding is reversible and follows the law of mass action.
- The K_d you entered is correct, and was determined under similar conditions.

K_i and EC_{50}

Prism first fits the curve to find the EC_{50}, the concentration of competitor that competes for half the specific binding. This is the same as the IC_{50}.

The value of the EC_{50} is determined by three factors:

- The affinity of the receptor for the competing drug, the K_i. If the affinity is high, the EC_{50} will be low. The subscript i is used to indicate that the competitor inhibited radioligand binding. You can interpret it as an

equilibrium dissociation constant, usually abbreviated K_d. It is the concentration of the competing ligand that will bind to half the binding sites at equilibrium in the absence of radioligand or other competitors.

- The concentration of the radioligand. If you choose to use a higher concentration of radioligand, it will take a larger concentration of unlabeled drug to compete for the binding.
- The affinity of the radioligand for the receptor (K_d). It takes more unlabeled drug to compete for a tightly bound radioligand (small K_d) than for a loosely bound radioligand (high K_d).

Prism calculates the K_i, using the equation of Cheng and Prusoff (Cheng Y., Prusoff W. H., Biochem. Pharmacol. 22: 3099-3108, 1973).

$$K_i = \frac{EC_{50}}{1 + \frac{[ligand]}{K_d}}$$

EC50 or log(EC50)?

The equation built-in to Prism is defined in terms of the log(EC_{50}), so Prism finds the best-fit value of the log(EC_{50}) along with its SE and 95% CI. Prism also reports the EC_{50} and its 95% CI. It does this by taking the antilog of the log(EC_{50}) and both ends of the 95% CI. Since the confidence interval is symmetrical on the log scale, it is not symmetrical when converted to EC_{50}.

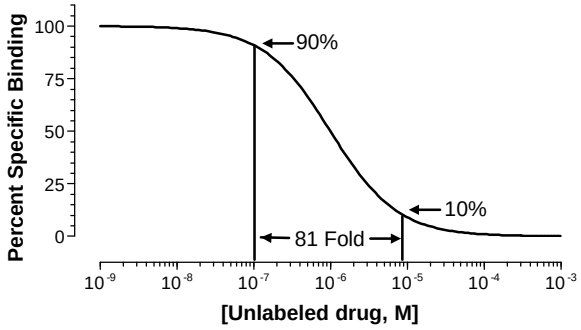
If the concentrations of unlabeled compound are equally spaced on a log scale, the uncertainty of the log(EC_{50}) will be symmetrical, but the uncertainty of the EC_{50} will not be. That is why the equation is written in terms of log(EC_{50}).

If you average together results from several experiments, it is better to average the log(K_i) values, rather than the K_i values.

Shallow competitive binding curves

The slope factor or Hill slope

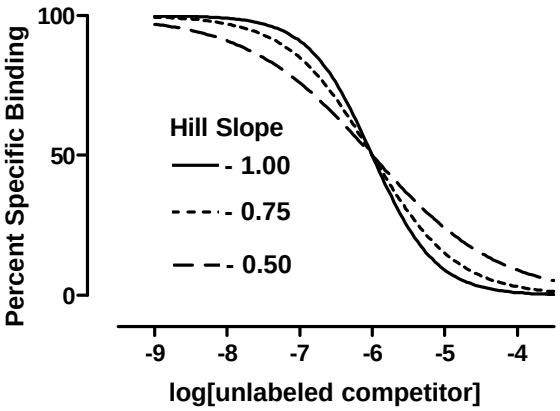
If the labeled and unlabeled ligand compete for a single binding site, the competitive binding curve will have a shape determined by the law of mass action. In this case, the curve will descend from 90% specific binding to 10% specific binding over an 81-fold increase in the concentration of the unlabeled drug. More simply, virtually the entire curve will cover two log units (100-fold change in concentration).



To quantify the steepness of a competitive binding curve, fit the data to the built-in equation "Sigmoid dose-response (variable slope)". Prism will fit the top and bottom plateaus, the IC_{50}, and the slope factor (also called Hill slope). A standard competitive binding curve that follows the law of mass action has a slope of -1.0. If the slope is more shallow, the slope factor will be a negative fraction (i.e. -0.85 or -0.60).

The slope factor describes the steepness of a curve. In most situations, there is no way to interpret the value in terms of chemistry or biology. If the slope factor is far from 1.0, then the binding does not follow the law of mass action with a single site.

Some investigators transform the data to create a linear Hill plot. The slope of this plot equals the slope factor. I don't see any advantage to displaying data in this manner. It does not aid data analysis (see "Linear regression of transformed data is less accurate" on page 250), and I don't think it helps visualize the findings.



Explanations for shallow binding curves include:

- Experimental problems. If the serial dilution of the unlabeled drug concentrations was done incorrectly, the slope factor is not meaningful.
- Curve fitting problems. If the top and bottom plateaus are not correct, then the slope factor is not meaningful. Don't try to interpret the slope factor unless the curve has clear top and bottom plateaus.
- Negative cooperativity. You will observe a shallow binding curve if the binding sites are clustered (perhaps several binding sites per molecule) and binding of the unlabeled ligand to one site causes the remaining site(s) to bind the unlabeled ligand with lower affinity.
- Heterogeneous receptors. The receptors do not all bind the unlabeled drug with the same affinity.
- Ternary complex model with limiting availability of G protein. This is a common explanation for shallow binding. See below.

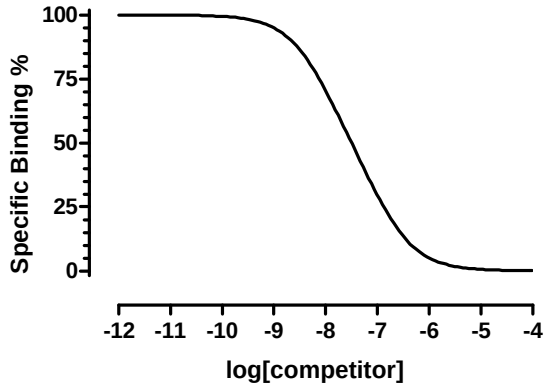
Competitive binding with two sites

Included in the list of built-in equations of Prism is "Competitive binding (two sites)". This equation fits data for the fairly common situation where:

- There are two distinct classes of receptors. For example, a tissue could contain a mixture of β_1 and β_2 adrenergic receptors.
- The unlabeled ligand has distinct affinities for the two sites.
- The labeled ligand has equal affinity for both sites. (If you are not willing to make this assumption, see "Competitive binding to two classes of receptor where each ligand has a different affinity for each sites" on page 346.)
- Binding has reached equilibrium.
- A small fraction of both labeled and unlabeled ligand bind. This means that the concentration of labeled ligand that you added is very close to the free concentration in all tubes.

This equation has five variables: the top and bottom plateau binding, the fraction of the receptors of the first class, and the IC_{50} of competition of the unlabeled ligand for both classes of receptors. If you know the K_d of the labeled ligand and its concentration, you (or Prism) can convert the IC_{50} values to K_i values.

When you look at the competitive binding curve, you will only see a biphasic curve in unusual cases where the affinities are very different. More often you will see a shallow curve with the two components blurred together. For example, this graph shows competition for two equally abundant sites with a ten fold (one log unit) difference in EC_{50}. If you look carefully, you can see that the curve is shallow (it takes more than two log units to go from 90% to 10% competition), but you cannot see two distinct components.



Comparing one- and two-site models

Prism can simultaneously fit your data to two equations and compare the two fits. This feature is commonly used to compare a one-site competitive binding model and a two-site competitive binding model. Since the model has an extra parameter and thus the curve has an extra inflection point, the two-site model almost always fits the data better than the one site model. And a three site model would fit even better. Before accepting the more complicated models, you need to ask whether the improvement in goodness of fit is more than you'd expect by chance. Prism answers this question with an F test that yields a P value that answers this question: If the one site model were really correct, what is the chance that randomly chosen data points would fit to a two-site model this much better (or more so) than to a one-site model. See "Comparing two equations" on page 277.

Before looking at Prism's comparison of the two equations, you should look at both fits yourself. Sometimes the two-site fit gives results that are clearly nonsense. For example, disregard a two-site fit when:

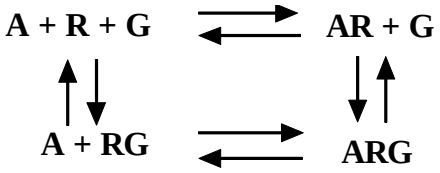
- The two IC_{50} values are almost identical.
- One of the IC_{50} values is outside the range of your data.
- The variable FRACTION is close to 1.0 or to 0.0. In this case, virtually all the receptors have the same affinity, and the IC_{50} value for the other site will not be reliable.
- The best-fit values for BOTTOM or TOP are far from the range of Y values observed in your experiment.

If the results don't make sense, don't believe them. Only pay attention to the results of the comparison when both fits are reasonable.

Agonist binding and the ternary complex

If you use an antagonist radioligand and compete with an unlabeled agonist you are likely to observe a shallow competitive binding curve. Receptors can interact with intracellular molecules in a way that alters agonist binding. One situation that has been well studied the competition of unlabeled agonist ligand for antagonist radioligands for binding to receptors linked to G proteins in membrane preparations lacking GTP.

The ternary complex model explains some of these data. This model shows binding of an agonist (A) to receptor (R) coupled to G protein (G):



Some features of the model:

- The left half of the picture shows the equilibrium between receptor (R) and G protein (G) in the absence of agonist (A).
- The right half shows the agonist-receptor complex binding to G to form the ternary complex (ARG).
- The top part shows the binding of agonist to receptors not linked to G.
- The bottom part shows the binding of agonist to receptors linked to G.
- Not shown is the radioligand L that binds with equal affinity to R and RG.

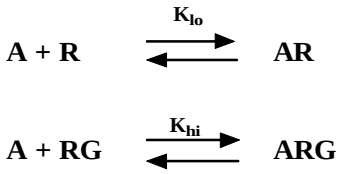
This model can explain the shape of agonist competition curves if the following are true:

- Agonist binds more tightly to the receptor-G complex than to receptor alone. In other words, A binds with low affinity to R but with high affinity to RG.
- The total concentration of G is limiting. If G is present in higher concentrations than R, this model does not predict shallow binding curves (Neubig et. al. Mol. Pharmacol. 28:475, 1985).
- GTP is omitted from the incubation. When GTP is present, as it is inside cells, the model is more complicated. Soon after the ARG complex forms, GTP binds to the G protein, and activates it. The activated G then dissociates from AR, and its α subunit dissociates from the $\beta\gamma$ subunits. Because the ARG complex is so transient, all you see in binding experiments is low affinity AR binding.

A few investigators have fit data to equations describing the ternary complex. Most do not, for these reasons:

- The model is too complicated to fit well. There are too many parameters to fit. You need to fit all four equilibrium constants, plus the relative concentration of receptor to G.
- The model is too simple to be useful. The ternary complex model only predicts a shallow competitive binding curve when the total concentration of G is less than or equal to the total concentration of receptors. But biochemical evidence in many systems demonstrates that G is present in vastly higher concentrations than receptors. Yet these systems demonstrate shallow competitive binding curves. To resolve this discrepancy, you must make the model still more complicated (either many of the G proteins are not available to bind to receptors, or many of the receptors are not available to bind to G proteins). For a review of these issues, see RR Neubig, Faseb J. 8:939-946, 1994.

Rather than fit data to the ternary complex model, most investigators fit data to a two-site binding model. The data fit well to the two site model, even though we know it is too simple. The two-site model is shown below – it is not the same as the ternary complex model.



The two-site model, as its name implies, assumes that there are two distinct kinds of receptors. In this context, one kind of receptor is coupled to G, and the other is not. This simpler model provides values for K_{lo}, K_{hi}, and relative numbers of the two kinds of receptors, and these can be compared after different treatments. Since we know that the two-site model is too simple (the ternary complex is a better model for many systems) you shouldn't interpret K_{hi} and K_{lo} strictly. Nonetheless, the ratio of the K_{hi} and K_{lo} is a useful empirical measure which often correlates with agonist efficacy.

This short section is hardly adequate to explain the complexities of the ternary complex model, but should be sufficient to keep you from making common mistakes in interpretation.

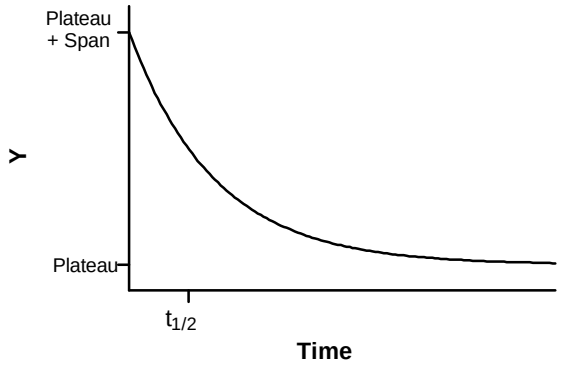
Dissociation binding data

How "off rate" experiments work

A dissociation binding experiment measures the "off rate" for radioligand dissociating from the receptor. Initially ligand and receptor are allowed to bind, perhaps to equilibrium. At that point, you need to block further binding of radioligand to receptor so you can measure the rate of dissociation. There are several ways to do this:

- If the tissue is attached to a surface, you can remove the buffer containing radioligand and replace with fresh buffer without radioligand.
- You can spin the suspension and resuspend in fresh buffer.
- Add a very high concentration of an unlabeled ligand. If the concentration is high enough it will instantly bind to nearly all the unoccupied receptors and thus block binding of the radioligand.
- Dilute the incubation by a large factor, perhaps a 20 fold dilution. This will reduce the concentration of radioligand by that factor. At such a low concentration, new binding of radioligand will be negligible. For this method to be useful, you need to use a low radioligand concentration to start with.

You then measure binding at various times after that to determine how rapidly the ligand falls off the receptors.



Binding follows this equation:

$$Y = Span \cdot e^{-K \cdot X} + Plateau$$

Variable	Meaning	Comment
X	Time	
Y	Total binding	
Span	Difference between binding at time zero and plateau.	Specific binding
Plateau	Binding that doesn't dissociate.	Nonspecific binding.
K	Dissociation rate constant often called k_{off}.	Expressed In units of inverse time
$t_{1/2}$	Half-life	$0.69302/k_{off}$

Analyzing dissociation data with Prism

To analyze dissociation binding data:

1. Enter the data with X equal to time after you initiated dissociation and Y equal to binding (usually total binding).
2. Perform nonlinear regression using the one phase exponential decay equation.
3. If you entered specific (rather than total) binding, make the variable PLATEAU a constant equal to zero. If you have entered total binding, leave the variable PLATEAU a variable to fit.
4. Look at the nonlinear regression results. The variable K is the dissociation constant (k_{off} or k_{-1}) expressed in units of inverse time. If you entered the X values as minutes, k_{off} is in units of min^{-1}. The results also show the half-life in units of time (minutes in this example).

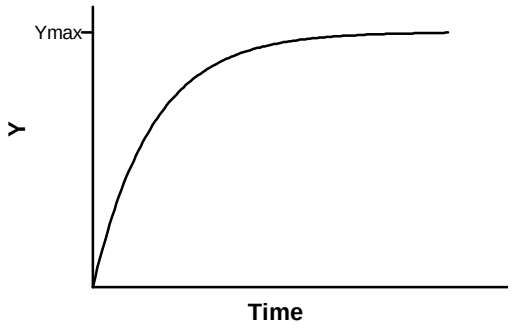
This analysis assumes that the law of mass action applies to your experimental situation. Dissociation binding experiments also let you test that assumption. Ask yourself these questions:

- Does all the specific binding dissociate? Is the binding truly reversible?
- Is the dissociation rate constant the same no matter how long you incubated the cells before initiating dissociation?
- Is the dissociation rate constant the same when you initiate dissociation by diluting and by adding unlabeled drug? If not, consider the possibility of cooperativity (binding sites are clustered, and binding of ligand to one binding site changes the affinity of the others).
- After dissociation, is the ligand chemically intact? Or has the ligand degraded?

Association binding data

How "on rate experiments" work

Association binding experiments are used to determine the association rate constant. You add radioligand and measure specific binding at various times thereafter.



The graph shows you the rate at which the binding approaches equilibrium. This is determined by four factors:

- The association rate constant, k_{on} or k_1. This is what you are trying to determine.
- The concentration of radioligand. If you use more radioligand, the system equilibrates faster.
- The dissociation rate constant, k_{off} or k_{-1}. Some people are surprised to see that the observed rate of association depends in part on the dissociation rate constant. During the incubation, radioligand is binding to the receptors and radioligand is dissociating from receptors. The system reaches equilibrium when the two rates are equal. So the observed rate of association measures how long it takes to reach equilibrium. If the radioligand dissociates quickly from the receptor, equilibrium will be reached faster.
- Temperature affects the values of k_{on} and k_{off}.

The next section explains how to calculate k_{on} from the rate of equilibration.

Analyzing on rate experiments with Prism

To analyze association (on-rate) data:

1. Enter the data with X equal to time and Y equal to specific binding. (If you enter total binding, you'll need to use a more complicated equation that accounts for the kinetics of nonspecific binding.)

2. Fit the specific binding data to the one phase exponential association equation.

$$Y = Ymax \cdot \left(1 - e^{-K \cdot X}\right)$$

The variable K in the exponential association equation is the *observed* rate constant, k_{ob}, expressed in units of inverse time. If you entered X values in minutes, then k_{ob} is expressed in min^{-1}. This is not the same as the association rate constant, k_{on}.

This equation assumes that a small fraction of the radioligand binds to receptors, so the concentration of free radioligand equals the amount you added and does not change over time.

3. To calculate the association rate constant (k_{on} or k_1) usually expressed in units of $Molar^{-1}$ min^{-1}, use this equation:

$$k_{on} = \frac{k_{ob} - k_{off}}{[radioligand]}$$

Variable	Units	Comment
K_{on}	$Molar^{-1}$ min^{-1}	What you want to know.
k_{ob}	min^{-1}	The value K determined by fitting an exponential association equation to your data.
k_{off}	min^{-1}	The dissociation rate constant. See previous section.
[radioligand]	Molar	Set by the experimenter. Assumed to be constant during the experiment (a small fraction binds).

To determine k_{on} if you don't know k_{off}:

1. Perform the association binding experiment at several different concentrations of radioligand.
2. Use nonlinear regression to find k_{ob} at each concentration.
3. Create a graph with [radioligand] on the X axis and k_{ob} on the Y axis.
4. Fit linear regression to this graph. The slope equals k_{on} and the Y-intercept is k_{off}.

Comparison of Prism with Ligand

Ligand is a special-purpose nonlinear regression program for analyzing equilibrium radioligand binding data written by Dr. Peter Munson and collaborators (*Methods in Enzymology* 92:543, 1983). In general, Prism and Ligand do the same thing and will calculate the same results. But there are a few differences:

- Prism is a more versatile program. It can fit kinetic data, and can create publication quality graphs. Ligand is designed to do one thing: fit equilibrium binding data.
- Although Prism analyzes many curves in parallel, each analysis is independent of the others. Ligand can analyze several curves at once, constraining some variables to be shared between curves.
- Prism reports EC_{50} and K_i values. Ligand reports K_a values, which are the reciprocal of K_i values.
- Ligand offers more choices for weighting.

Advanced radioligand binding analyses

Advanced radioligand binding template

This chapter explains some analysis techniques that are a bit more complicated than the basic ones explained in the previous chapter. Don't try to use these methods until you have become familiar with curve fitting with Prism.

All the methods described in this chapter fit data to user-defined equations. These equations are not built-in to Prism. Rather than type in the lengthy equations, you can use the template file radiolig.pzt.

The **radiolig.pzt** template is not designed for you to analyze data. Its purpose is a stealthy one. When you open this template file, Prism copies all its equations to your user-defined equation list. From then on, you can choose these equation from the user-defined list. You don't have to do anything special with the template. Just open and close it to transfer all the equations.

These equations are designed to be helpful starting places. Using these equations is different than using the built in ones. You will probably not be able to get satisfactory analyses by blindly applying these equations.

- In most cases, you must fix some of the variables to constant values.
- You may need to change the equation if your model differs from the one presented here.
- You may need to change the initial values of some of the variables.
- You may wish to rewrite the equation to use different units.
- You may decide to set different variables to constant values.

Homologous competitive binding curves

Advantages of using the same compound as hot and cold ligand

The standard way to determine receptor number and affinity is to perform a saturation binding experiment where you vary the concentration of radioligand. An alternative is to keep the radioligand concentration constant, and compete for binding with the same ligand, but not radioactively labeled. Since the hot and cold ligands are the same, this is called a homologous competitive binding experiment. The advantage of this approach is that you need less radioligand than you would need if you varied radioligand concentration.

The analyses depend on these assumptions:

- The receptor (or receptors) has identical affinity for the labeled and unlabeled ligand. Since iodination often changes the binding properties of ligands, you may have to iodinate the unlabeled compound as well (with nonradioactive iodine).
- There is no cooperativity.
- No ligand depletion. The methods in this section assume that only a small fraction of ligand binds. In other words the method assumes that free concentration equals the concentration you added. A later section in this chapter extends the model (for one class of receptors) to situations where this assumption is not valid.

Note that the value of B_{max} is determined, in part, from the amount of binding at the highest concentration of unlabeled compound. These tubes have the smallest number of cpm, and therefore are subject to the largest relative counting error. Reduce the counting error by leaving these tubes in the counter for a long time. See "The Poisson distribution" on page 308.

One class of receptors

The standard binding equation (rectangular hyperbola) predicts the amount of specific ligand binding as a function of the total amount added. The complication here is that only a variable fraction of the bound ligand is labeled. The equations below adapt the binding equation to the situation where you have a mixture of labeled and unlabeled ligands and only detect binding of the labeled ligand.

```
Cold=10^(X+9)   ;nM
L=HotnM +Cold      ;nM
F=HotnM/L         ;unitless
Kd=10^(logKd+9) ;nM
Y=NS + (((L*Bmax)/(L+Kd))*F)
```

Variable	Units	Comments
X	log(molar)	log molar concentration of the unlabeled compound
Y	CPM or fmol/mg or sites/cell	Total binding of the labeled compound
Bmax	Same as Y values	Initial value = 10*Ymax (This assumes that you used a concentration of ligand that binds to one tenth of these receptors)
LogKd	log(molar)	Initial value = 0.7*XMID
NS	Same as Y values	Initial value = 1.0*YMIN
HotnM	nM	Concentration of labeled ligand in every tube. Set this to a constant value that you know from experimental design.

The X values are the logarithm of the concentrations of the unlabeled ligand. The first line calculates the concentration of unlabeled (cold) ligand in nanomolar.

The second line calculates the total concentration of ligand in each tube. It is the sum of the concentration of radioactive labeled ligand (a constant value that you enter in nM) and the concentration of cold (that varies with X).

The third line computes the fraction of the total ligand that is hot. Assuming that the hot and cold ligands bind with the same affinity, it is also the fraction of the ligand bound to receptors that is radioactively labeled.

Since the experiment is performed with the concentrations of unlabeled ligand equally spaced on a log scale, the confidence intervals will be most accurate when the Kd is fit as the log(Kd). The fourth line converts the logKd (that will be fit) to the Kd (that goes into the next equation).

The final line computes the amount of radioligand binding (Y) as the sum of the nonspecific binding (NS) plus the amount of specific binding of the hot ligand. This is calculated as the total amount of ligand bound times F, the fraction of the total ligand that is labeled.

Make HotnM a constant equal to the concentration of labeled ligand in nM . Prism cannot fit this value, you must make it a constant and enter its value.

Two classes of receptors

When the labeled and unlabeled ligands are identical, use this equation to fit binding to two binding sites. This method assumes that a small fraction of the ligand binds to the receptors, so the concentration of free labeled ligand is the same in all tubes.

```
cold=10^(X+9)        ;nM
L=HotnM+Cold         ;nM
F=HotnM/L            ;unitless
Kd1=10^(logKd1+9)    ;nM
Kd2=10^(logKd2+9)    ;nM
Y=NS + F*((L*Bmax1)/(L+Kd1) +(L*Bmax2)/(L+Kd2))
```

The logic of this equation is similar to the one-site equation explained in detail above.

This equation (and this experimental method) will prove to be useful only when the amount of binding to the two receptor sites is approximately equal. If binding to the low affinity sites makes up only a small fraction of the counts, you won't have much information about the low affinity sites and the curve fitting is unlikely to be helpful. Since the ligand will bind to a higher fraction of high affinity sites than low affinity sites, the number of counts binding to the two sites will be approximately equal only when the low affinity sites are present in much larger numbers than the high affinity sites.

This equation is useful to simulate binding to two sites. Try a few simulations to see how difficult it is to see the second site. Fitting data to this equation is difficult, and requires very good initial values. You will almost certainly have to try many sets of initial values to get the fit to converge on a reasonable solution. It is especially important to adjust the initial values of the two Bmax values.

Variable	Units	Comments
X	log(molar)	Concentration of the unlabeled compound.
Y	CPM or fmol/mg or sites/cell.	Total binding of the labeled compound.
Bmax1	Same as Y values.	Initial value = 1*Ymax (this assumes that you used a concentration of ligand that binds to almost all of the first class of receptors).
Bmax2	Same as Y values.	Initial value = 20*YMAX (this assumes that you used a concentration of ligand that binds to five percent of the second class of receptors).
LogKd1	log(molar)	Initial value = 1.2*XMID.
LogKd2	log(molar)	Initial value = 0.8*XMID.
NS	Same as Y values.	Initial value = 1.0*YMIN.
HotnM	nM	Concentration of labeled ligand in every tube. Set this to a constant value that you know from experimental design.

Ligand depletion

The standard methods explained in the previous chapter are based on the assumption that a small fraction of the radioligand binds to receptors. This means that the concentration of radioligand added is very similar to the concentration of radioligand free in solution. This is sometimes called “zone A”.

In some experimental situations, the receptors are present in high concentration and have a high affinity for the ligand, and that assumption is not true. A large fraction of the radioligand binds to receptors so the total concentration added is quite a bit higher than the free concentration. The system is not in "zone A". The free ligand concentration is depleted by binding to receptors.

If possible you should avoid experimental situations where the free ligand concentration is far from the total concentration. You can do this by using less tissue in your assays. The problem is that this will also decrease the number of counts. An alternative is to increase the volume of the assay without changing the amount of tissue. The problem with this approach is that you'll need more radioligand. See "Total vs. free concentrations of ligand. Ligand depletion." on page 312.

If you can't avoid radioligand depletion, you need to account for the depletion in your analyses. The simplest way to do this is to subtract the number of

cpm (counts per minute) of total binding from the cpm of added ligand to calculate the number of cpm free in solution. This can then be converted to the free concentration in molar. There are four problems with this approach:

- If you used this method, experimental error in determining specific binding would also affect the calculated value of free ligand concentration. When you fit curves, both X and Y would include experimental error, and the errors will be related. This violates the assumptions of nonlinear regression. Using simulated data, Swillens (see reference below) has shown that this can be a substantial problem.
- Since the free concentration in the nonspecific tubes is not the same as the free concentration in the total tubes, it is difficult to deal with nonspecific binding using this approach. You can not calculate specific binding as the difference between the total binding and nonspecific binding.
- This method works only for saturation binding experiments, and cannot be extended to analysis of competition curves.
- You cannot implement this method with Prism, which does not let you subtract Y from X. (Since Prism allows for many Y data sets per table, but only one X column, subtracting Y from X would be ambiguous).

S. Swillens (Molecular Pharmacology, 47: 1197-1203, 1995) developed an alternative approach – fit total binding as a function of added ligand using an equation that accounts both for nonspecific binding and for ligand depletion. By analyzing simulated experiments, that paper shows that fitting total binding gives more reliable results than you would get by calculating free ligand by subtraction. The next three sections explain how to use this approach for different kinds of experiments. The equations shown here are not exactly the same as in Swillens' paper, but the ideas are the same.

Analyzing saturation binding experiments with ligand depletion

Theory

From the law of mass action, total binding follows this equation.

$$\text{Total Binding} = \text{Specific} + \text{Nonspecific}$$

$$\text{Total Binding} = \frac{B_{max} \cdot [\text{Free Ligand}]}{K_d + [\text{Free Ligand}]} + [\text{Free Ligand}] \cdot NS$$

The first term is the specific binding, which equals fractional occupancy times B_{max}, the total number of binding sites. The second term is nonspecific binding, which is assumed to be proportional to free ligand concentration.

The variable NS is the fraction of the free ligand that binds to nonspecific sites.

To write a more useful equation:

- The free concentration of ligand is the concentration you added minus the concentration that bound (specific and nonspecific).
- The amount of ligand added is the independent variable, X.
- Total binding is the variable you measure, Y.

$$Y = \frac{B_{max} \cdot [\text{Free Ligand}]}{K_d + [\text{Free Ligand}]} + [\text{Free Ligand}] \cdot NS$$

$$[\text{Free Ligand}] = X - Y$$

Combining the two equations,

$$Y = \frac{B_{max} \cdot (X - Y)}{K_d + (X - Y)} + (X - Y) \cdot NS$$

X, Y and B_{max} are expressed in units of cpm. To keep the equation consistent, therefore, Kd must also be converted to cpm units.

You cannot enter that equation into Prism for nonlinear regression because it is an implicit equation where Y appears on both sides of the equal sign. But simple algebra rearranges it into a quadratic equation. The solution is shown as a user defined Prism equation.

```
;X is total ligand added in cpm. Y is total binding in cpm

;SpecAct is specific radioactivity in cpm/fmol
;Vol is reaction volume in ml
;Both must be set to be CONSTANTS

;Calc KD in cpm from nM
KdCPM=KdnM * Vol * 1000 * SpecAct
; (nm/L * mL * 0.001 L/ml * 1000000 fmol/nmol * cpm/fmol)

a=-1-NS
b=KdCPM + NS*KdCPM + X + 2*X*NS + Bmax
c=-1*X*(NS*KdCPM + X*NS+Bmax)

Y=(-b+sqrt(b*b-4*a*c) )/(2*a) ;Y is in cpm
```

Variable	Units	Comments
X	CPM	Total ligand added.
Y	CPM	Total binding.
SpecAct	cpm/fmol	Specific radioactivity of ligand. Set to constant value which you know from experimental design.
Vol	ml	Reaction volume. Set to constant value which you know from experimental design.
NS	Unitless fraction	Initial value = 1.0*YMID/XMID
BMAX	CPM	Initial value = 0.5*YMAX
KdnM	nM	Initial value = 1.0 (no way to do this by rule, so pick a value in the right ball park).

Notes:

- Enter your data with X equal to the total ligand added in cpm and Y equal to the total binding in cpm. You can convert to more useful units after fitting.
- You do not need to experimentally measure nonspecific binding (but see below).
- Set SpecAct (specific radioactivity in cpm/fmol) and Vol (reaction volume in ml) to constant values.
- After fitting the curve, transform the data table to change X to nM and Y to fmol/mg or sites/cell or some other useful unit. Also transform the XY table of the curve fit results in the same way. Add both tables to a new graph.

Determining nonspecific binding experimentally

The method described above fits total binding data to an equation that includes both specific and nonspecific components. It does not require that you experimentally determine nonspecific binding. While this is convenient, many would feel uneasy trusting those results without determining nonspecific binding experimentally.

You can experimentally determine nonspecific binding by including a large concentration of an unlabeled ligand in your incubations. This will bind to virtually all the receptors, leaving only nonspecific sites free to bind radioligand. The conventional approach is to measure total and nonspecific binding at each ligand concentration, and to define specific binding as the difference. This approach cannot be used when a high fraction of ligand binds because

the free concentration of ligand in the total tubes is not the same as the free concentration of ligand in the nonspecific tubes.

We assume that nonspecific binding is a constant fraction of the concentration of free ligand.

$$\text{Nonspecific binding} = Y = NS \cdot [\text{Ligand}]$$

We also assume that the free concentration of ligand equals the added concentration (X) minus the amount that bound (Y).

$$[\text{Ligand}] = X - Y$$

Combining the two equations:

$$Y = (X - Y) \cdot NS$$

$$Y = X \cdot \frac{NS}{NS + 1}$$

To experimentally determine nonspecific binding:

1. Enter the data with X equal to the added concentration in cpm and Y equal to the nonspecific binding in cpm.
2. Fit to nonlinear regression to this user-defined equation: $Y = X*NS/(NS + 1)$
3. Since this is really a linear equation, you'll get the same fit no matter what initial value you enter. Set the initial value of NS equal to 0.01.
4. Compare the value of NS determined here with the NS determined from analysis of the total binding. The two values should be similar.
5. If you think the value of NS determined here is more accurate than the NS determined from analysis of the total binding, you can refit the total binding data holding NS constant equal to the value determined here.

Heterologous competitive binding with ligand depletion

The standard sigmoidal equations used to fit competitive binding data assume that a small fraction of the radioligand binds. This means that the free concentration of radioligand is almost equal to the concentration you added, and that the free concentration is the same in all tubes in the assay.

If a large (say greater than 10%) fraction of the radioligand binds to receptors, then the free concentration will not equal the added concentration of radioligand. High concentrations of unlabeled drug will compete for more binding sites, and thus increase the free concentration of radioligand. Since nonspe-

cific binding is proportional to the free concentration, nonspecific binding will also not be the same in all tubes.

This is the standard equation for competitive binding:

$$Y = \frac{[\text{Free Ligand}] \cdot B_{max}}{[\text{Free Ligand}] + K_d\left(1 + \frac{[\text{Cold Ligand}]}{K_i}\right)} + \text{Nonspecific}$$

When a large fraction of ligand binds, this equation needs two corrections:

- The free concentration of labeled ligand equals the amount you added minus the amount that bound.

$$[\text{Free ligand}] = [\text{Added ligand}] - Y$$

- The nonspecific binding is not the same for all tubes. Nonspecific binding is a constant fraction of the free concentration of labeled ligand. As you increase the concentration of cold ligand, less radioligand binds to receptors so the free concentration of radioligand increases. Since nonspecific binding is assumed to be proportional to the free concentration of radioligand, there will be more nonspecific binding in the tubes with higher concentrations of unlabeled drug.

$$\text{Nonspecific binding} = NS \cdot [\text{Free ligand}]$$

Y, [Free ligand], and [Added ligand] are expressed in units of cpm. To be consistent, therefore the Kd needs to also be expressed in cpm units. [Cold ligand] and Ki are expressed in the same units (molar), so the ratio is unitless.

Combine those equations, and you end up with a complicated quadratic equation whose solution is shown here:

```
KdCPM=KdnM*SpAct*vol*1000
; nmol/L *(cpm/fmol * ml * .001L/ml * 1000000fmol/nmol) = cpm
R=NS+1
S=[1+10^(X-LogKi)]*KdCPM+Hot
a=-1*R
b=R*S+NS*Hot + Bmax
c= -1*Hot*(S*NS + Bmax)
Y= (-1*b + sqrt(b*b-4*a*c))/(2*a)
```

You need to set four of the parameters to constant values. Hot is the number of cpm of labeled ligand added to each tube. SpAct is the specific activity of the radioligand in cpm/fmol, Vol is the incubation volume in ml, and Kd is the KdnM of the radioligand in nM (determined from other experiments). The program fits this equation to your data to determine the logKI. It also fits two

other variables which of are less interest: Bmax which is the maximum binding of radioligand (if present at a saturating concentration) in cpm, and NS which is the fraction of the free ligand that binds nonspecifically.

Variable	Units	Comments
X	log(Molar)	
Y	CPM	
Hot	CPM	Amount of labeled ligand added to each tube. Set to constant value.
SpAct	cpm/fmol	Specific radioactivity of labeled ligand. Set to constant value.
Vol	ml	Incubation volume. Set to constant value.
KdnM	nM	Kd of labeled ligand. Set to constant value.
LogKi	log(Molar)	Initial value = 1.0*XMID
Bmax	Units of Y axis, usually cpm	Initial value = 10*YMAX (This assumes that you've used a concentration of radioligand that binds to one tenth of the receptors. You may wish to adjust.)
NS	Unitless fraction	Initial value =0.01

Notes:

- This equation accounts for ligand depletion when a large fraction of the radioligand binds to receptors. It does not adjust for depletion of the unlabeled compound. It assumes that the concentration of unlabeled compound that you added (antilog of X) equals the free concentration. If your unlabeled compound binds with high affinity, this assumption may not be true.
- You may use this equation for any competitive binding curve, even if only a small fraction of the radioligand binds. The results will be identical to the results from the more conventional equations.
- This equation is not easily extended to a situation with two binding sites.

Homologous competitive binding with ligand depletion

In a homologous competitive binding curve, the labeled and unlabeled ligands are identical. As the unlabeled ligand competes for binding of the la-

beled ligand, the free concentration of labeled ligand rises. If a large fraction of the radioligand is bound to the receptors, this will affect the results substantially. It affects both specific binding and nonspecific. The tubes with higher concentrations of unlabeled ligand have higher concentrations of free radioligand (less is bound to receptors) and thus more nonspecific binding.

We can't simply adapt the equation of the previous section, because it assumes that the unlabeled ligand is not depleted. Here we assume that hot and cold ligands are depleted together (since they have the same affinity).

Here is the equation for homologous competition.

$$Y = \frac{B_{max} \cdot [\text{Free Radioligand, nM}]}{K_d + [\text{Free Radioligand, nM}] + [\text{Free Cold ligand, nM}]} + [\text{Free Radioligand, cpm}] \cdot NS$$

Y is the total binding of hot ligand that you measure in cpm. You vary the concentration of cold ligand. The logarithm of the concentration of unlabeled ligand is X. NS is the fraction of free radioligand that binds nonspecifically.

To keep units consistent, the radioligand concentration is expressed in nM in the left half of the equation (to be consistent with K_d and the concentration of cold ligand) and is expressed in cpm on the right half of the equation (to be consistent with Y).

Each tube contains the same number of cpm of hot ligand, which we'll call HotCPM concentration of free radioligand in cpm equals HotCPM-Y. The fraction of hot radioligand that is free is therefore (HotCPM - Y)/HotCPM. This fraction will be different in different tubes. Since the hot and cold ligands are chemically identical, that is also the fraction of cold ligand that is free in solution. Since X is the logarithm of the total concentration of cold ligand, the free concentration of cold ligand is 10^X(HotCPM - Y)/HotCPM.

Combining these equations leads to a lengthy quadratic equation whose solution is shown below as a user-defined Prism equation.

```
ColdnM=10^(X+9)
KDnM=10^(LogKD+9)
HotnM=HotCPM/(SpAct*vol*1000)
; cpm/(cpm/fmol * ml * .001L/ml * 1000000fmol/nmol)
TotalnM=HotnM+ColdnM
Q=HotCPM*(TotalnM + KDnM)
a=(Ns+1)*TotalnM*-1
b=Q*(NS+1)+TotalnM*HotCPM*NS + Bmax*HotnM
c=-1*Q*HotCPM*NS - HotCPM*Bmax*HotnM
Y= (-1*b + sqrt(b*b-4*a*c))/(2*a)
```

Variable	Units	Comments
X	log(Molar)	
Y	cpm	
HotCPM	cpm	Amount of labeled ligand added to each tube. Set to constant value.
SpAct	cpm/fmol	Specific radioactivity of labeled ligand. Set to a constant value.
Vol	ml	Incubation volume. Set to a constant value.
logKD	log(Molar)	Initial value = 1.0*XMID
Bmax	Units of Y axis, usually cpm	Initial value = 10*YMAX (This assumes that you've used a concentration of radioligand that binds to one tenth of the receptors. You may wish to adjust.)
NS	Unitless fraction	Initial value =0.01

When fitting data to this equation, you need to set three parameters to constant values. HotCPM is the number of cpm of hot ligand added to each tube. Vol is the incubation volume in ml. SpAct is the specific radioactivity in cpm/fmol. Prism fits Bmax in the units of the Y axis (usually cpm which you can convert to more useful units) and logKd as log molar.

Notes:

- Although you add the same amount of labeled ligand to each tube, the free concentration changes. High concentrations of unlabeled drug compete for the binding of the labeled ligand, and thus increase the free concentration of the labeled ligand.
- Since nonspecific binding is proportional to the free concentration of labeled ligand, the amount of nonspecific binding is not the same in all tubes. The tubes with the highest concentration of unlabeled drug have the highest nonspecific binding of the radioligand.
- Because the free concentration varies among tubes, as does the nonspecific binding, the IC_{50} is nearly meaningless.
- These equations can also be used when a small fraction of ligand binds to the receptors. The results will be the same as if you used the simpler equations that assume a small fraction of ligand binds.

- Unfortunately, this equation is not easily extended to fit data when there are two classes of receptors.

Competitive binding to two classes of receptor where each ligand has a different affinity for each sites

The standard equation for competitive binding to two sites assumes that the labeled ligand has equal affinity for both sites. It is easy to derive an equation for when the labeled ligand binds differently to the two sites.

This is the standard equation for competitive binding to one site:

$$Y = \frac{[\text{Hot Ligand}] \cdot B_{max}}{[\text{Hot Ligand}] + K_d\left(1 + \frac{[\text{Cold Ligand}]}{K_i}\right)} + \text{Nonspecific}$$

Binding is the sum of specific and nonspecific binding. To create an equation for two sites, you simply need to create an equation with two specific binding components with different values for B_{max}, K_d, and K_i.:

```
;Enter data with X=log[unlabeled] and Y=CPM
ColdnM=10^(X+9)
KI1nM = 10^(LogKI1+9)
KI2nM = 10^(LogKI2+9)
SITE1= HotnM*Bmax1/(HotnM + KD1*(1+coldnM/Ki1nM))
SITE2= HotnM*Bmax2/(HotnM + KD2*(1+coldnM/Ki2nM))
Y = SITE1 + SITE2 + NSCPM
```

Notes:

- This equation does not account for ligand depletion. It assumes that the free concentration equals the added concentration.
- When using this equation to fit data, you will need to assign constant values to KD1 and KD2, the KD of the hot ligand for the two sites. You will need to obtain these values from other experiments. Perhaps you can isolate tissue with only one of the receptor types and measure KD in that preparation.

Variable	Units	Comments
X	log(Molar)	Concentration of unlabeled drug.
Y	cpm	Total binding of radioligand.
HotnM	nM	Concentration of labeled ligand added to each tube. Set to a instant value.
KD1	nM	Kd of the labeled ligand for the first site.

		Set to a constant value based on other experiments.
KD2	nM	Kd of the labeled ligand for the second site. Set to a constant value .
logKI1	log(Molar)	Affinity of the unlabeled drug for the first site. Initial value = 1.2*XMID
logKI2	log(Molar)	Affinity of the unlabeled drug for the second site. Initial value = 0.8*XMID
Bmax1	Units of Y axis, usually cpm	Initial value = 2*YMAX (This assumes that you've used a concentration of radioligand that binds to half of the receptors. You may wish to adjust.)
Bmax2	Units of Y axis, usually cpm	Initial value = 10*YMAX (This assumes that you've used a concentration of radioligand that binds to one tenth of the receptors. .)
NSCPM	Units of Y axis, usually cpm.	Nonspecific binding. Initial value = 1.0 * YMIN.

Kinetics of competitive binding

The standard methods of analyzing competitive binding experiments assume that the incubation has reached equilibrium. These experiments are usually used to learn the affinity of the receptors for the unlabeled compound, the Ki. The law of mass action tells us that the Ki is the ratio of the off-rate to the on-rate of the unlabeled compound. You can determine these values in a kinetics experiment. Add labeled and unlabeled ligand together and measure the binding of the labeled ligand over time. This method was described by Motulsky and Mahan in Molecular Pharmacology 25:1-9, 1984.

```
KA = K1*L*1E-9 + k2
KB = K3*I*1e-9 + K4
S=SQRT((KA-KB)^2+4*K1*K3*L*I*1e-18)
KF = 0.5 * (Ka + KB + S)
KS = 0.5 * (KA + KB - S)
DIFF=KF - KS
Q=Bmax*K1*L*1e-9/DIFF
Y=Q*(k4*DIFF/(KF*KS)+((K4-Kf)/KF)*exp(-KF*X)-((K4-KS)/KS)*exp(-KS*X))
```

Variable	Units	Comments
X	Minutes	Time.
Y	cpm	Specific binding .
k1	M^{-1} min^{-1}	Association rate of radioligand. Set to a constant value known from other experiments.
k2	min^{-1}	Dissociation rate of radioligand. Set to a constant value known from other experiments.
k3	M^{-1} min^{-1}	Association rate of unlabeled ligand. Variable to be fit. Try 1e8 as an initial value.
k4	min^{-1}	Dissociation rate of unlabeled ligand. Variable to be fit. Try 0.01 as an initial value.
L	nM	Concentration of radioligand. Set to a constant value you know from experimental design.
Bmax	Units of Y axis. Usually cpm.	Total concentration of receptor. Either leave as a variable or set to a constant you know from other experiments. If a variable, set the initial value to 100*Ymax (assumes that it bind to 1% of receptors.
I	nM	Constant set experimentally. Concentration of unlabeled ligand.

Notes:

- This equation does not account for ligand depletion. It assumes that a small fraction of radioligand binds to receptors, so that the free concentration of radioligand is very close to the added concentration.
- This method will only give reliable results if you have plenty of data points at early time points.

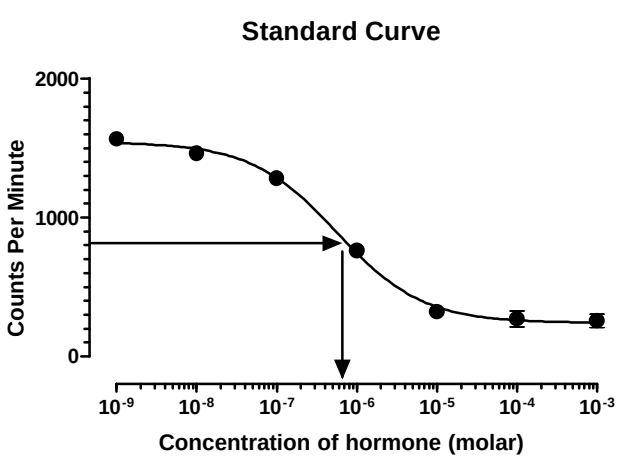
Reading unknowns from standard curves

Introduction

Before learning how to analyze standard curves with Prism, let's review how it is done manually.

1. Perform an assay with various known concentrations of a substance you are trying to measure.
2. Graph these data to make a standard curve – concentration on the X axis, and assay measurement on the Y axis.
3. Connect the points or fit a curve through them.
4. Perform the same assay with your unknown samples.
5. For each unknown, read across the graph from the spot on the Y axis that corresponds to the assay measurement of the unknown until you intersect the standard curve.
6. Read down the graph until you intersect the X axis.
7. The concentration of substance in the unknown sample is the value on the X axis.

In the example below, the unknown sample had 820 counts per minute, so the concentration of the hormone is $8x10^{-7}$ M.



Term	Meaning
Assay	An experimental procedure used to determine the concentration of a substance. The measurement (which might be optical density, radioactivity, luminescence, or something else) varies with the concentration of the substance you are measuring.
Standard curve	A graph of assay measurements (Y) as a function of known concentrations of the substance (X). This is used to calibrate the assay.
Unknowns	Samples in which you want to know the concentration of the substance you are assaying.
Interpolation	Reading the unknown concentration from the standard curve.

How to fit standard curves

Before you can read unknown values, you first must fit a line or curve through your standard points. Prism lets you fit a standard curve with one of these methods:

- Linear regression. The standard curve should start below your lowest unknown value and go a bit beyond your highest unknown value. There is no benefit to continuing the standard curve far beyond the range of your unknowns. Standard curves are often nearly linear within a narrow range, but become curved after that. If you restrict you stan-

dard curve values to that narrow range, linear regression might be useful even if the overall standard curve is not straight.

- Nonlinear regression. You will need to pick an equation. Many standard curves are generated by radioimmunoassay (RIA) or similar assays (i.e. ELIZA). These assays are based on competitive binding. The compound you are assaying competes for binding to an enzyme or antibody with a labeled compound. Therefore the standard curve is described by equations for competitive binding. Try the one-site competitive binding curve. If that doesn't fit your data well, try the sigmoidal dose response curve with variable slope. When fitting sigmoidal curves, the X values are the logarithm of concentration.
- Cubic spline (or lowess). These are general methods for curve fitting. The advantage of these methods is that they are simple. You don't need to pick an equation. The disadvantage is that the curves tend to wiggle too much.
- Polynomial regression. If you don't know what equation describes your data, you can try the second, third or fourth order polynomial equations. With Prism, you perform polynomial regression by choosing a polynomial equation from the nonlinear regression dialog. The higher order polynomial equations generate standard curves with more inflection points.

When you use nonlinear regression in other contexts, the choice of an equation is very important. If the equation does not describe a model that makes scientific sense, the results of nonlinear regression will be meaningless. With standard curve calculations, the choice of an equation is not so important because you are not interested in the best-fit values of the variables in the equation. All you have to do is assess visually that the curve nicely fits the standard points.

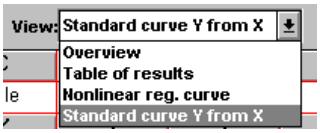
Determining unknown concentrations from standard curves

To read values off the standard curve:

1. Enter the unknown values on the same table as your standard curve. Just below the standard curve values, enter your unknowns as Y without X. This example shows X and Y values for five standards, and Y values for four unknowns.

	X	Y
1	0.0	0.00
2	2.0	0.12
3	4.0	0.21
4	6.0	0.29
5	8.0	0.57
6		0.14
7		0.48
8		0.09
9		0.36

2. Click on the Graphs button, and look at a graph of your standard curve.
3. Click the analyze button, and choose how to fit the standard curve. Choose either linear or nonlinear regression, or create a LOWESS, Spline, or Point-to-point curve.
4. On the Parameters dialog, check the standard curve option (X from Y). If you are using nonlinear regression, you need to press the Output options button to see the standard curve choices.
5. Click on the Results button to see the tabular results. Drop down the View list and pick the standard curve.



6. Look at the table of X and Y values. The Y column contains values you entered, and the X column shows the calculated concentrations in the same units as you used for the X axis.
7. If necessary, you can transform the results (i.e. convert to anitlogs). Click the New Analysis button. On the New Analysis dialog, choose Transforms and choose to analyze the data you are looking at.

Standard curves with replicate unknown values

Prism can interpolate from a standard curve, even if you have replicate unknown values.

Enter the data with all the replicates as shown below. The top part of the table is the standard curve. Below that are the unknown values. The standards and the unknowns do not need to have the same number of replicate determinations.

	X Values	A	
	X Title	Data Set-A	
	X	Y1	Y2
1	-9.0	1597	1531
2	-8.0	1453	1471
3	-7.0	1314	1245
4	-6.0	751	771
5	-5.0	336	306
6	-4.0	328	212
7	-3.0	207	307
8		1123	1085
9		1345	1298
10		1456	1421
11		987	

When you fit the standard curve, select the standard curve option (X from unpaired Y). If you fit the curve with nonlinear regression, this is on the Output options dialog. The example was fit using nonlinear regression with a sigmoidal dose-response curve with variable slope.

The standard curve results are shown on two output views.

View: Standard curve X from Y

	X Values	A
	X Title	Data Set-A
	X	Y
1	-6.614	1104.000
2	-7.103	1321.500
3	-7.565	1438.500
4	-6.409	987.000

The Y column shows the average of the replicate unknown Y value you entered. The X values are the concentrations that correspond to that mean Y value.

View: Unknown X Values

	A	
	Data Set-A	
	Y1	Y2
1	-6.649	-6.579
2	-7.175	-7.037
3	-7.673	-7.473
4	-6.409	

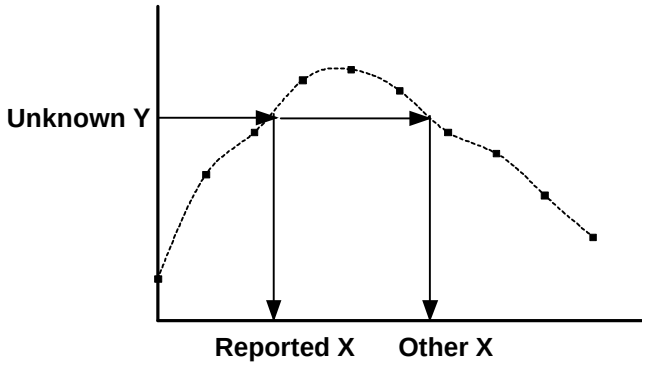
Each value is a concentration corresponding to one of the replicate values you entered, and is expressed in the same units as the X axis of your standard curve. Because Prism cannot deal with replicate X values, Prism places these unknown X values in a Y column on the results table. Think of them as X values on your standard curve. But think of them as Y values when you want to do further analyses.

To calculate the mean and SD (or SEM) of the replicate values, press Analyze and choose row statistics.

Problems with standard curves

Reading unknown values from a linear regression line is completely straightforward. Reading unknown values from a curve is subject to the following potential problems:

- Prism can only read unknowns off the standard curve within the range of the standard curve. If you enter an unknown value that is larger than the highest standard or smaller than the lowest standard, Prism will not try to determine the concentration unless you fit the standard curve with linear regression. You should interpret these extrapolations cautiously.
- If you calculate X from Y, beware of a possible ambiguity. It is possible that two or more points on the curve have identical Y values but different X values. In this situation, Prism will report the lowest of the X values within the range of the standard curve, and will not warn you that other answers exist.



- You will get more accurate results if you define the curve with more line segments. Prism defines a curve as a large number of points. To find unknown values, Prism linearly interpolates between the two points on either side of the unknown value. If you define the curve with more line segments, the interpolation will be more accurate. To learn how to increase the number of line segments, see "Method options for nonlinear regression" on page 263.

Menus

File menu

Menu choice	Explanation
New Project Ctrl-N	Create a new project, without closing the current project(s). You may open up to four projects at once.
Open Ctrl-O	Retrieve and open an existing Prism file or template.
Close Ctrl-W	Close the current project. If you haven't already saved the project, Prism will prompt you to do so.
Save Ctrl-S	Save the current project file.
Save As	Save with a new name or to a new location.
Print	Print the current sheet, an entire section, or the entire project.
Printer Setup	Select a printer, and choose a paper size.
Export	Export data as a text file (delimited by commas or tabs). Export a graph or page layout in WMF, BMP, PCX, TIF, or GIF format.
Import	Import a text file onto a data table or notes page. Import an image onto a graph or page layout.
Merge	Merge another Prism file into the current project.
Exit ALT-F4	Quit the program. Prism will prompt you to save the file before quitting.
File 1	The nine most recently listed files are listed. Select one to open it.

Edit menu

Menu Choice	Explanation
Undo/Redo Ctrl-Z	Reverse the last action, if possible. Immediately after selecting Undo, the menu choice changes to Redo.
Cut Ctrl-X	Delete the selected object(s) or the selected block of data and copy to the clipboard.
Copy Ctrl-C	Copy the selected object(s) or selected block of data to the clipboard without deleting.
Paste Ctrl-V	Paste objects onto a graph or page. Or paste a block of data onto the data table.
Clear DEL or Backspace	Delete the selected objects or the selected range of data. Do not copy to the clipboard.
Embed table	Embed a selected range of a data or results table onto a graph or layout.
Select Ctrl-A (all)	Select an entire row, column of data table. Select an entire graph or an entire page layout.
Delete	Delete a selected region from the data table and fill in the gap.
Insert	Insert rows or columns into the data table. First select row(s), column(s), or a block of values.
Number Format	Choose the number of decimal points to display in the selected columns.
Exclude	Exclude the selected values from analyses and graphs. The values will appear in italics on the data table.
Sort by X value	Sort all rows in the data table in ascending value by X value.
Options	Set global options or preferences for the program, for graphs, and for analyses.

View menu

Menu Choice	Explanation
Zoom Small	See the entire page at once.
Zoom 50%	View the graph or page layout at half actual size.
Zoom 75%	View the graph or page layout at 75% actual size.
Actual Size (100%)	View the graph or page layout at actual size.
Zoom 150%	View the graph or page layout at 1.5 times full size.
Zoom Large (200%)	View the graph or page layout at twice actual size.
Show Tool Bar	Show/hide buttons and tools at the top of the screen.
Show Rulers	Show/hide the ruler at the top and left of the screen.
Show Coordinates	Show/hide coordinates of the mouse pointer (only shown if rulers are showing.). The coordinates are the distance from the left and top edges of the page in inches (or cm). If you are working on a graph, the X, Y coordinates (in data units) are also shown on top of the tool bar.
Show Grid	Show/hide a grid on graphs and layouts to aid alignment.
Select Section	Choose to look at data, results, graph, layout, or notes.
Select Sheet	Select which sheet you will work on.
Select View	When looking at a results sheet, select which output view (page) you want to look at.

Sheet menu

Menu Choice	Explanation
Create	Add a new sheet to your project. Choose whether you wish to add a data table, analysis, graph, page layout or notes.
Duplicate	Duplicate the current sheet. After duplicating, you can change one or both of the copies.
Delete	Delete the current sheet. If you delete a data table, you will also delete the data from graphs. If you delete a graph, you will also delete that graph from page layouts.
Rename	Change the sheet name and its order in the section.
Freeze	Freeze results or graph so it will not update when data change.
Change	Change the current sheet. Don't confuse with Change menu.
Analysis Parameters Ctrl-T	Change the choices for the analysis you are looking at.
Axes	Change the scale, range or numbering of the axes.
Symbols and Lines	Change the appearance of symbols, lines and bars.
Background Color	Select the background color for graphs and pages.

Properties menu

Menu Choice	Explanation
Fill Pattern	The pattern inside of selected boxes and ovals.
Fill Color	The color inside of selected boxes and ovals.
Line & Border Thickness	The thickness of selected lines, arrows, arcs, axes, and the border of selected boxes and ovals.
Line & Border Pattern	The pattern (dotted, dashed, etc.) of lines, arrows, and borders.
Line & Border Color	The color of selected lines, arrows, axes and borders.
Arrowhead direction	Add or remove arrowheads from lines, arrows and axes.
Arrowhead size	Choose among three sizes of arrowheads.
Table Properties	Change the appearance of an embedded table placed on a graph or layout.

Text menu

Menu Choice	Explanation
Font	Change the font (or size) of selected text.
Insert Greek	Insert a Greek letter chosen from the secondary menu.
Insert International	Insert an international letter (i.e. ë or ö) from the secondary menu.
Insert Math	Insert a mathematical symbol (i.e. ± or ≥) from the secondary menu.
Insert Wingding	Insert an arrow, circled digit, or other strange symbols.
Color	Change the color of selected text.
Justify	Choose left, right, center or full justification from the secondary menu. Titles are justified with respect to the axis. Labels are justified with respect to their object boundaries.
Rotate	Rotate text vertically. Only rotates text.
Notes Format	Choose margins, tab spacing, and line spacing.
Border thickness	Place a border around a block of text and choose its thickness.
Background color	Choose the background color of text.
Underline Ctrl-U	Underline the selected text.
Bold Ctrl-B	Bold the selected text.
Italics Ctrl-I	Italicize the selected text.
Superscript Ctrl +	Superscript the selected text.
Subscript Ctrl -	Subscript the selected text.

Note: To superscript or subscript selected text, hold the Ctrl key and press the + or - key on your keypad. The + and - keys on top of the regular keys will not work.

Arrange menu

Menu Choice	Explanation
Align	Select two or more objects before selecting this command. From the secondary menu, choose whether to align the X-axes or Y-axes of graphs, or the left, right, center, top or bottom of objects.
Center on page	Center all objects on the page. Prism slides all objects up or down, and right or left, to equalize the margins.
Balance Graph Scaling	Resize some of the selected graphs to equalize the scaling (magnification) factors so fonts and symbols match.
Bring to Front	Move the selected object "in front" of other objects. When objects overlap, only the one in front is visible.
Send to Back	Move the selected object "in back" of other objects. When objects overlap, only the one in front is visible.
Group	All selected objects are considered to be one group. Objects in a group move together, and properties are assigned together.
Ungroup	Ungroup objects that you previously grouped.
Lock Object	Lock the selected object so you won't accidentally move it on the page.
Duplicate Object	Duplicate the selected object(s). The new object will appear offset from the original.
Position Object	Specify the distance of the object from the left and top of the page.
Use Snaps	When selected, objects always snap to an invisible grid. A check mark appears in front of this menu choice when it is selected. Choose the command again to deselect it.

Window menu

Menu Choice	Explanation
Cascade Windows	Show the open projects in overlapping windows.
Tile Windows	Show the open projects in nonoverlapping windows.
StatMate	Launch the StatMate program. This is a separate program that is often sold with Prism. If you installed Statmate, this command will launch StatMate without closing Prism. StatMate must be installed in the PRISM2 directory.
Project 1	All open projects (up to four) are listed. Choose a project to bring its window to the front.

Help menu

Menu Choice	Explanation
Contents	Look at the Table of Contents.
Search	Search by keyword.
Getting Started	Learn how to use Prism.
Commands	Information about the menu commands.
Tips	Ideas to use Prism more effectively.
Technical Support	How to contact GraphPad for technical support.
Using Help	Learn how to use the Help system.
About Prism	Find your serial number and version.

Troubleshooting

Windows resources

Windows refers to a special area of memory as "Windows resources". All programs share the same limited pool of resource memory. No matter how much memory you install in your computer, the pool of "resource" memory is constant and limited.

When the pool of available resources gets low, all Windows programs act erratically. When less than 25% of the resources are free, Prism is unable to update the computer screen properly. You may see a graph superimposed on a data table. Text may be displayed in the wrong place. Rotated text will appear horizontal. Graphs may be incomplete. Text will appear in the wrong font or size.

To find out the percent of free Windows (system) resources:

1. Pull down the Help menu from Program Manager (not Prism). If you are using another Windows shell such as PC Tools or Norton Desktop, pull down the Help menu from the desk top.
2. Choose About Program Manager. (If you use a nonstandard Windows shell, this may be worded differently).
3. The About dialog shows the percent of system resources free.

To free up Windows resources:

Resources can be depleted for two reasons.

- You have too many programs open. Close some programs to free resources.
- Some programs have exited without releasing resources. This can happen in several situations, but usually happens when programs crash or freeze. The only solution to this problem is to exit all programs, and then exit and restart Windows.

Depletion of Windows resources is a much larger problem with Windows 3.1 than with Windows 95. If you are having problems with too few resources, we recommend that you switch to Windows 95.

Printing problems

If you encounter problems printing, consider the following:

- Do you have the correct printer driver, and is it up to date? All Windows programs access printers through “printer driver” software. The Windows program comes with drivers for many printers, and printer manufacturers often provide Windows drivers. Many drivers are updated frequently. Microsoft freely distributes drivers for popular printers.

Method	How to access
Phone	US: 800-426-9400 Elsewhere: Look in your Windows manual
Modem	(USA) 206-936-6735
World wide web	http://www.microsoft.com
FTP	ftp.microsoft.com

- If you can’t obtain the driver from Microsoft, contact your printer manufacturer.
- Graphics programs such as Prism use the drivers more extensively than word processors. An out-of-date driver may work fine with some programs, but not work well with Prism. GraphPad does not provide printer drivers.
- How are fonts printed? Prism only uses TrueType fonts. Some printer drivers give you the choice of either down-loading fonts to the printer or using fonts built-in to the printer (or installed with a printer cartridge). When the second method works properly, printing is faster and perhaps sharper. When it doesn’t work properly, text will look ugly and Y axis title may appear horizontal instead of vertical. Some drivers work best when you select “Print TrueType text as graphics”. Others work best with "Download TrueType as bitmap soft fonts".
- What printer resolution have you selected? Many printer drivers give you a choice of resolution. Sometimes the resolution is expressed as dots per inch (dpi) — a higher number means higher resolution. Other times the resolution is expressed in words such as “draft quality” vs. “high quality”.
- What other options are set in the driver? Printer drivers commonly offer several options, although the details vary. For example, some drivers offer a choice about “dithering” (select none).

To experiment with different options in the printer driver:

1. Select Printer setup from Prism’s File menu.

2. Press the Options button, and make changes on the dialog. Some drivers offer additional dialogs accessed by pressing another button (i.e. "Advanced" or "More").
3. Print a page.

   The changes you make to the printer setup affect only the Prism project you are currently working with. In fact, the printer setup is saved with the file. Once you have decided how you want to configure the printer driver, you should make those changes in the control panel so they will apply to future projects.

To make global changes to a printer driver:

Changes you make from the Printer Setup command within Prism apply only to the project you are currently working on. This is useful for experimentation. To configure a printer driver for future projects (and other programs), you need to make the changes from the Windows control panel.

1. Quit Prism.
2. From the Windows program manager, find the icon for the control panel and double-click it.
3. Double-click on the Printers icon.
4. Select a printer, and press Setup. Some drivers offer additional dialogs accessed by pressing a button (i.e. "Advanced..." or "More...").

Note: Changes you make in the control panel will affect all programs, not just Prism.

TrueType fonts

Prism only uses TrueType fonts. TrueType fonts are installed when you installed Windows.

How to tell if you TrueType fonts are loaded on your system:

1. Make a graph.
2. Click on any title.
3. Pull down the Text menu and choose Font. The left side of the Font dialog lists all available TrueType fonts.

There are two reasons why no fonts are listed. Either no fonts are installed, or the TrueType system is not enabled.

To enable the TrueType system:

1. From the desktop, double-click on the icon for the Windows control panel.
2. Double-click on the Fonts icon.

3. From the control panel Fonts dialog, press the button labeled "TrueType".
4. Select "Enable TrueType fonts". Since Prism only uses TrueType fonts, it doesn't matter how you set "Show only TrueType fonts in applications".

TrueType fonts are always enabled with Windows 95.

To install TrueType fonts:

1. From the desktop, double-click on the icon for the Windows control panel.
2. Double-click on the Fonts icon.
3. From the control panel Fonts dialog, press the button labeled "Add". Follow the instructions.

Temporary files

Like many programs, Prism creates temporary files as it works. When you exit Prism, it deletes the temporary files. If you turn off or reboot the computer without exiting the program, the temporary files are not deleted. Over time, therefore, your hard drive can accumulate temporary files from Prism and other programs. Once your disk is littered with hundreds of temporary files, Windows and Prism can act strangely. You should delete temporary files periodically.

To erase your temporary files:

1. Close all applications.
2. Go to the directory that holds temporary files. Depending on how your system is configured, this may be C:\WINDOWS\TEMP. C:\WINDOWS, C:\TEMP. or C:\DOS (or elsewhere).
3. Delete *.tmp. You can identify Prism files because they begin with ~PZT followed by four digits. You should delete all temporary files, not just those created by Prism.

Appendices

File names

After you install Prism, the following files will be on your hard drive. The first three are essential. The others are required to display the help screens and the help navigator.

File for Win95	File for Win3.1	Function
PRISM.EXE	PRISM.EXE	Main program file. Essential.
VIC32.DLL	VIC.DLL	Essential overlay.
CTL3D32.DLL	CTL3D.DLL	Essential overlay in Windows system directory.
PRISM.HLP	PRISM.HLP	Needed for on-line help.
PRISM.DHN	PRISM.DHN	Help navigator index.
D2HLNK32.DLL	D2HLNK16.DLL	Needed for help navigator. In Windows System directory (Win 3.1) or Windows Help directory (Win 95)
D2HNAV32.EXE	D2HNAV.EXE	Needed for help navigator. Windows directory (Win 3.1) or Windows Help directory (Win 95)
MFC30.DLL	MSOUTLIN.VBX	Needed for help navigator. Windows system directory.
MSVCRT20.DLL		Needed for help navigator. Windows system directory.

Prism also comes with template files with the extension .pzt and sample data files with the extension .txt. See "Template files" on page 32.

Prism project files have the extension .PZM. Prism template files have the extension .PZT. These files contain your data, analyses and graphs. Several template files are included with Prism.

Prism places the files PRISM2.CNF (preferences and equations) and PRISM.CLP (clipboard information) in your Windows directory, usually C:\WINDOWS.

Keyboard shortcuts

Press	To
Ctrl A	Select all.
Ctrl B	Make text bold.
Ctrl C	Copy to clipboard.
Ctrl D	Go to data (same as pressing the Data tab).
Ctrl E	Exclude selected data.
Ctrl F	Change the font of the selected text.
Ctrl G	Go to graphs (same as pressing the Graphs tab).
Ctrl I	Italicize the selected text.
Ctrl J	Align objects.
Ctrl L	Go to layouts (same as pressing the Layout tab).
Ctrl M	Magnify. Zoom larger.
Ctrl N	Go to notes (same as pressing the Notes tab).
Ctrl O	Open a file.
Ctrl P	Print.
Ctrl R	Go to results (same as pressing the Results tab).
Ctrl S	Save.
Ctrl T	Change analysis parameters.
Ctrl U	Underline the selected text.
Ctrl V	Paste from clipboard.
Ctrl W	Close the window.
Ctrl X	Cut.
Ctrl Y	Unmagnify. Zoom smaller.
Ctrl Z	Undo.
Ctrl +	Superscript selected text. (Use the plus key next to the keypad.)
Ctrl -	Subscript selected text. (Use the minus key next to the keypad.)
Ctrl PgUp	Go to previous sheet.
Ctrl PgDn	Go to next sheet.
Ctrl Space	Select column in table.
Shift Space	Select row in table.
Ctrl	Constrain objects to only move horizontally. Constrain ovals to be circles, and rectangles to be squares.
Shift	Constrain objects to only move vertically.
Ctrl & Shift	Toggle between small magnification and ordinary magnification. Hold Ctrl and Shift, then click.
F1	Help.
Alt F4	Exit.

Control F6 or Control Tab	Go to next document. Only works when two or more projects are open, each in its own window.

Shortcut menus

Click the right mouse button to access shortcut menus from any part of the program.

Uninstalling GraphPad Prism

To remove all traces of Prism from your system, follow these steps.

Step 1: Delete the files

1. From Windows, pull down the File menu and select Run.
2. Type **C:\PRISM2\UNINSTAL**
3. Substitute another drive or directory if needed. This uninstall program removes all program files from the PRISM directory and configuration files from the WINDOWS directory. It does not erase your data or project files.
4. Delete your unneeded Prism project and template files.

Step 2. Delete the Prism icon

1. Click once to select the Prism icon.
2. Press DEL to delete it.

Step 3. Delete Prism from the registration data base

Windows maintains a registration database that associates the file extensions PZM and PZT with the Prism program. There is no harm in leaving Prism in this database, even after you have deleted the program. If you want to remove Prism, follow these steps.

From Windows 3.1:

1. From Program Manager, pull down the File menu and choose Run.
2. Type "REGEDIT" and press Return.
3. Highlight "GraphPad Prism".
4. Pull down the Edit menu (from the Registration Info Editor) and choose "Delete File Type".
5. When asked to confirm, answer Yes.

From Windows 95:

1. From the desktop, double-click the My Computer icon.
2. Pull down the View menu and choose Options.
3. Click the File Types tab.
4. Highlight GraphPad Prism Project.
5. Press the Remove button. When asked to confirm, answer Yes.

Note: Installing Prism did not make any changes in the system files (auto-exec.bat, config.sys, system.ini, or win.ini), so you don t have to worry about those files when you uninstall Prism.

Technical support

Support

If you encounter problems, contact us for help. We offer unlimited free support for one year.

- Please read the manual or help screens before calling. Don't expect the technical support staff to teach you how to use Prism. We especially recommend that you work through the Prism tutorial at the beginning of the manual.
- Our technical support staff cannot teach you how to use Windows.
- Our free support does not include free statistical consulting. We can sometimes provide limited statistical consulting for an hourly fee.
- If several people in your lab share one copy of Prism (on one computer), please route all technical support through one person whenever possible.
- If possible, please be near your computer when you call.
- If you encounter problems with printing, please read "Printing problems" on page 364 before calling. You may need to update your Windows printer driver.
- Before calling, be sure that you know the serial number and the version of your program.

Before you call for support

Know your serial number and full version number

1. While running Prism, pull down the Help menu.
2. Choose About.
3. The dialog shows the version number (i.e. "2.01"), serial number ("i.e. GPA-23456-789"), and release date.

We need to know the full version number (two decimal places). It is not enough to say you have version 2.0 or "the latest version". The full version number is not printed on the manual or disk label. To find it, you must run Prism, pull down the Help menu and choose About.

If possible, check the World Wide Web to see if a more recent version is available.

Go to http://www.graphpad.com on the World Wide Web and follow the self-explanatory prompts. If a more recent update is available, you will be able to download it free (this applies to minor updates reflected in a change in the second digit after the decimal, i.e. 2.01 to 2.02. It does not apply to major updates, which we will charge for.)

If possible, be at your computer when you call.

We can often provide better support when you are at your computer when you call.

Minor Updates

We periodically release minor upgrades to improve the program and fix minor problems. These upgrades will be indicated by a change in the third digit of the version number (i.e. from 2.01 to 2.02) or by a different program release date.

We announce the current version numbers in our newsletter. Minor upgrades will be freely available on the internet. If you are unable to access the internet, we can mail you update disks for a small handling fee.

Major Upgrades

Major upgrades will be indicated by a change in the first or second digit of the version number (i.e. from 2.0 to 2.1, or from 2.1 to 3.0). If you own the previous version, major upgrades will cost less than one-third the list price of the new version. If you have an older version, upgrades may cost more.

Contacting GraphPad Software

Email:	support@graphpad.com *or* sales@graphpad.com
World Wide Web:	http://www.graphpad.com
Fax:	(USA) 619-457-8141
Mail:	GraphPad Software, Inc. 10855 Sorrento Valley Road #203 San Diego, CA 92121 USA
Phone:	619-457-3909

GraphPad StatMate

StatMate is an ideal companion to Prism and other statistics programs. StatMate complements Prism by providing a collection of 8 statistical calculators to help you design statistically sound experiments and to help you interpret your own and published results.

Planning an experiment

- StatMate helps you pick the right statistical test to use. Tell it the kind of data you are collecting, and the kind of comparison you want to make. StatMate suggests a test, and tells you whether it can be performed with other GraphPad programs.
- How many subjects (data points) do you need? Naturally, the answer is "it depends". It depends on how large a difference you are looking for, how much your data vary, and on how willing you are to risk mistakenly finding a difference by chance or mistakenly missing a real difference. StatMate helps you see the tradeoffs, so you can pick an appropriate sample size.
- To avoid subtle biases, it is best to randomly assign subjects to treatments. StatMate will do this for you. Simply tell it how many treatment groups you need, and how many subjects you want in each group. StatMate assigns each subject to a group.

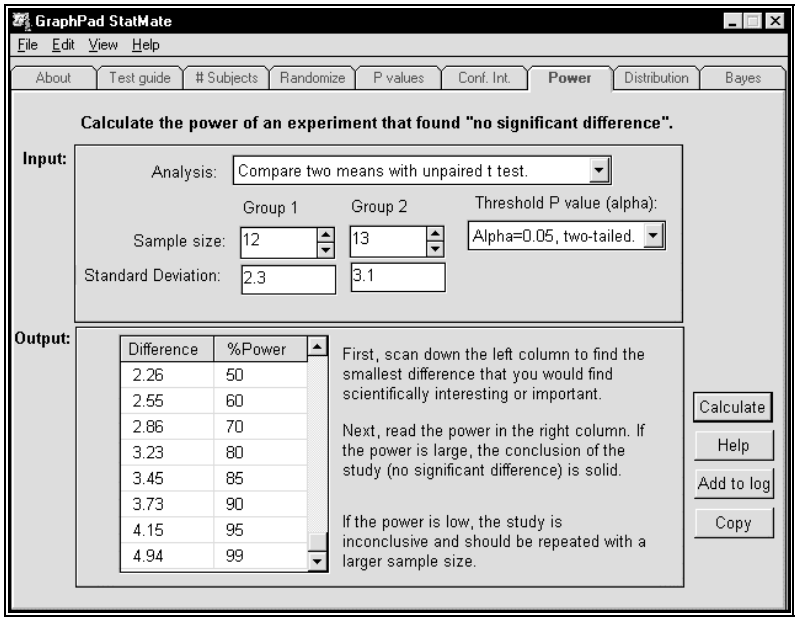
Interpreting an experiment

- Many publications (and some programs) just tell you that "$P<0.05$" or "$P>0.05$" but don't tell you the exact value of P. StatMate will calculate P (to four decimal places) from t, z, F, r, or chi-square. It will also calculate a P value from the results of the runs test, sign test, and McNemar's test.

- While many publications focus on P values, results are often easier to interpret when expressed as a confidence interval. StatMate calculates a confidence interval of a proportion, count, or mean. You specify whether you want 90%, 95% or 99% confidence.
- Just because a study reaches a conclusion that the results "are not statistically significant" doesn't mean that the treatment was ineffective. It is possible that the study missed a small effect due to small sample size and/or large scatter. StatMate calculates the power of a test to detect various hypothetical differences.
- When reviewing the results of an experiment (or a laboratory test) you need to interpret the statistical information in the context of what you knew before you did the experiment (based on theory and previous data). StatMate performs the Bayesian calculations required to do this.

The StatMate Screen

The StatMate looks like this:



The eight calculators are denoted by the tabs across the top of the Window. In this example, you pressed the Power tab to bring the power calculator to the front. You then completed the input portion of the calculator on the top of the Window and pressed Calculate. After reading the results on the bottom, you can press Copy (place results on the clipboard for transfer to a spreadsheet program or word processor), Add to log button (accumulate several sets of results in the results log, and then print or export the entire log), or Help (help interpret the results).

System requirements

StatMate runs under Windows 3.1 or Windows 95. It occupies less than a megabyte of space on your hard drive. GraphPad StatMate currently (mid 1995) lists for $95. We offer discounts to students and for group purchases.

The relationship between Prism, StatMate, and InStat

Prism and StatMate are companion programs, with no overlap in function. If you install StatMate (as suggested) into the Prism directory, you can launch StatMate directly from Prism by clicking a button on the toolbar.

All features of InStat are now included in either Prism or StatMate.

Intuitive Biostatistics (book)

If you like the style of this manual, you'll probably also like *Intuitive Biostatistics*, by Harvey Motulsky, president of GraphPad software and author of this manual. Here is the publisher's description:

> "I*ntuitive Biostatistics* provides a nonmathematical introduction to biostatistics for medical and health sciences students, graduate students in biological sciences, physicians and researchers. Using nontechnical language, this text focuses on explaining the proper scientific interpretation of statistical tests rather than on the mathematical logic of the tests themselves. *Intuitive Biostatistics* covers all the topics typically found in an introductory statistics text, but with the emphasis on confidence intervals rather than P values, making it easier for students to understand both. Additionally, it introduces a broad range of topics left out of most other introductory texts but used frequently in biomedical publications, including survival curves, multiple comparisons, sensitivity and specificity of lab tests, Bayesian thinking, lod scores, and logistic, proportional hazards and nonlinear regression. By emphasizing interpretation rather than calculation, *Intuitive Biostatistics* provides a clear and virtually painless introduction to statistical principles, enabling readers to understand statistical results published in biological and medical journals."

You may order the book from GraphPad software with software purchases only. To order from a bookstore or the publisher (Oxford University Press) , cite this number: ISBN 0-19-508607-4 (US $24.95).

Index

—A—

—B—

—C—

—D—

—E—

—F—

—G—

—H—

—I—

—J—

—K—

—L—

—M—

—N—

—O—

—P—

—Q—

—R—

—S—

—T—

—U—

—V—

—W—

—X—

—Y—

—Z—