

Analyzing Data with GraphPad Prism

A companion to GraphPad Prism version 3

Harvey Motulsky
President GraphPad Software Inc.
Hmotulsky@graphpad.com

GraphPad Software, Inc.

© 1999 GraphPad Software, Inc. All rights reserved.

All Rights Reserved.

GraphPad Prism, Prism and InStat are registered trademarks of GraphPad Software, Inc. GraphPad is a trademark of GraphPad Software, Inc.

Use of the software is subject to the restrictions contained in the accompanying software license agreement.

Citation: H.J. Motulsky, Analyzing Data with GraphPad Prism, 1999, GraphPad Software Inc., San Diego CA, www.graphpad.com.

Acknowledgements: The following individuals made substantial contributions towards development of this book: Arthur Christopoulos (Neuroscience Research in Psychiatry, Univ. Minnesota), Lee Limbird (Dept. Pharmacology, Vanderbilt University), Rick Neubig (Dept. Pharmacology, Univ. Michigan), Paige Searle (Vice-president GraphPad Software).

How to reach GraphPad:

Phone: 858-457-3909 (619-457-3909 before June 12, 1999)

Fax: 858-457-8141 (619-457-8141 before June 12, 1999)

Email: support@graphpad.com or sales@graphpad.com

Web: www.graphpad.com

Mail: GraphPad Software, Inc.
5755 Oberlin Drive #110
San Diego, CA 92121 USA

Contents

Preface	2
Introduction to statistical comparisons	3
Garbage in, garbage out.....	3
When do you need statistical calculations?	3
The key concept: Sampling from a population.....	4
Confidence intervals	8
P values	9
Hypothesis testing and statistical significance.....	11
Statistical power	13
A Bayesian perspective on interpreting statistical significance.....	15
Beware of multiple comparisons	17
Outliers.....	18
Analyzing one group	23
Entering data to analyze one group	23
Frequency distributions	24
Column statistics.....	25
Interpreting descriptive statistics.....	26
The results of normality tests.....	29
The results of a one-sample t test.....	30
The results of a Wilcoxon rank sum test.....	34
Row means and totals	37
t tests and nonparametric comparisons	39
Introduction to comparing of two groups	39
Entering data to compare two groups with a t test (or a nonparametric test)	39
Choosing an analysis to compare two groups.....	41
The results of an unpaired t test.....	45
The results of a paired t test	51
The results of a Mann-Whitney test	57
The results of a Wilcoxon matched pairs test.....	59
One-way ANOVA and nonparametric comparisons	65
Introduction to comparisons of three or more groups.....	65
Choosing one-way ANOVA and related analyses	67
The results of one-way ANOVA	72
The results of repeated measures one-way ANOVA	82
The results of a Kruskal-Wallis test	85
The results of a Friedman test.....	89
Two-way analysis of variance	93
Introduction to two-way ANOVA.....	93
Entering data for two-way ANOVA.....	93

Choosing the two-way ANOVA analysis	94	Interpreting the results of nonlinear regression.....	207
The results of two-way ANOVA	97	Approach to interpreting nonlinear regression results.....	207
Survival curves	109	Do the nonlinear regression results make sense?	208
Introduction to survival curves	109	How certain are the best-fit values?	209
Entering survival data	109	How good is the fit?.....	212
Choosing a survival analysis	112	Does the curve systematically deviate from the data?	214
Interpreting survival analysis.....	113	Could the fit be a local minimum?	216
Graphing survival curves.....	120	Have you violated an assumption of nonlinear regression?.....	217
Contingency tables	121	Have you made a common error when using nonlinear regression?.....	218
Introduction to contingency tables	121	Comparing two curves.....	221
Entering data into contingency tables	122	Comparing the fits of two models.....	221
Choosing how to analyze a contingency table.....	123	Comparing fits to two sets of data (same equation).....	224
Interpreting analyses of contingency tables	124	The distributions of best-fit values.....	229
The confidence interval of a proportion	131	Why care about the distribution of best-fit values	229
Correlation.....	135	Using simulations to determine the distribution of a parameters	229
Introduction to correlation.....	135	Example simulation 1. Dose-response curves.....	230
Entering data for correlation.....	135	Example simulation 2. Exponential decay.....	231
Choosing a correlation analysis.....	136	Detailed instructions for comparing parameter distributions.....	234
Results of correlation.....	137	Analyzing radioligand binding data	237
Linear regression	141	Introduction to radioligand binding.....	237
Introduction to linear regression.....	141	Law of mass action.....	238
Entering data for linear regression	143	Nonspecific binding.....	240
Choosing a linear regression analysis.....	143	Ligand depletion.....	241
Results of linear regression	146	Calculations with radioactivity	241
Graphing linear regression.....	152	Analyzing saturation radioligand binding data	249
Fitting linear data with nonlinear regression.....	153	Introduction to saturation binding experiments	249
Introducing curve fitting and nonlinear regression.....	157	Nonspecific binding.....	249
Introduction to models	157	Fitting a curve to determine B_{max} and K_d	250
How models are derived	158	Checklist for saturation binding.....	252
Why Prism can't pick a model for you.....	161	Scatchard plots	253
Introduction to nonlinear regression	162	Analyzing saturation binding with ligand depletion	257
Other kinds of regression.....	166	Analyzing competitive binding data	263
Fitting a curve without choosing a model.....	167	What is a competitive binding curve?	263
Choosing or entering an equation (model)	171	Competitive binding data with one class of receptors.....	266
Classic equations built-in to Prism	171	Shallow competitive binding curves.....	268
Importing equations and the equation library	178	Homologous competitive binding curves.....	273
User-defined equations	181	Competitive binding to two receptor types (different K_d for hot ligand)	282
Constraining variables in user-defined equations.....	187	Heterologous competitive binding with ligand depletion	284
How to fit different portions of the data to different equations	190	Analyzing kinetic binding data	287
How to simulate a theoretical curve	193	Dissociation ("off rate") experiments.....	287
Fitting curves with nonlinear regression	195	Association binding experiments.....	288
Fitting curves with nonlinear regression.....	195	Analysis checklist for kinetic binding experiments.....	291
Initial values for nonlinear regression	197	Using kinetic data to test the law of mass action	291
Constants for nonlinear regression.....	199	Kinetics of competitive binding.....	293
Method options for nonlinear regression.....	200	Analyzing dose-response curves	297
Output options for nonlinear regression.....	204	Introducing dose-response curves.....	297
Default options for nonlinear regression	205	Fitting sigmoid dose-response curves with Prism	300
		Why Prism fits the $logEC_{50}$ rather than EC_{50}	301

Other measures of potency: pEC50, EC80, EC90, etc.	302
Checklist. Interpreting a dose-response curve.	304
The operational model of agonist action	305
Dose-response curves in the presence of antagonists	311
Analyzing enzyme kinetic data.....	317
Introduction to enzyme kinetics	317
How to determine V _{max} and K _M	321
Comparison of enzyme kinetics with radioligand binding	322
Displaying enzyme kinetic data on a Lineweaver- Burke plot	323
Allosteric enzymes.....	325
Enzyme kinetics in the presence of an inhibitor.....	325
Reading unknowns from standard curves.....	329
Introduction to standard curves.....	329
How to fit standard curves.....	330
Determining unknown concentrations from standard curves	331
Standard curves with replicate unknown values	332
Problems with standard curves.....	333
Preprocessing data.....	335
General.....	335
Transforming data	335
Normalizing data	339
Pruning rows	340
Subtracting (or dividing by) baseline values	341
Transposing rows and columns.....	343
Analyzing continuous data.....	345
Smooth, differentiate or integrate a curve.....	345
Area under the curve.....	347
Appendix. Using GraphPad Prism to analyze data	351
The Prism User's Guide.....	351
Projects, sections and sheets	351
How to analyze data with Prism.....	352
Changing an analysis	353
Analyzing repeated experiments	353
Intuitive Biostatistics (book)	355
Technical support	355
Index.....	357

Preface

GraphPad Prism combines scientific graphics, statistics and curve fitting in an easy to use program. If you don't already have Prism, go to www.graphpad.com to learn about Prism and download a free demonstration version.

Please also check our web site for a list of any corrections to this book, and for additional information on data analysis. Tell your colleagues that this entire book is available on the web site as an Acrobat pdf file.

GraphPad Prism comes with two volumes, of which you are reading one. The other volume, *Prism User's Guide*, explains how to use Prism. This book, *Analyzing Data with GraphPad Prism*, explains how to pick an appropriate test and make sense of the results.

This book overlaps only a bit with my text, *Intuitive Biostatistics*, published by Oxford University Press (ISBN 0195086074). The goal of *Intuitive Biostatistics* is to help the reader make sense of statistics encountered in the scientific and clinical literature. It explains the ideas of statistics, without a lot of detail about analyzing data. Its sections on ANOVA and nonlinear regression are very short. The goal of this book (*Analyzing Data with GraphPad Prism*) is to help you analyze your own data, with an emphasis on nonlinear regression and its application to analyzing radioligand binding, dose-response, and enzyme kinetic data.

If you have any comments about this book, or suggestions for the future, please email me at hmotulsky@graphpad.com

Harvey Motulsky

March 1999

Introduction to statistical comparisons

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.

While this volume provides far more background information than most program manuals, it cannot entirely replace the need for statistics texts and consultants. If you pick an inappropriate analysis, the results won't be useful.

GraphPad provides free technical support when you encounter problems with the program. However, we can only provide limited free help with choosing statistics tests or interpreting the results (consulting can sometimes be arranged for a fee).

When do you 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 difficult 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 necessary when observed differences are small compared to experimental imprecision and biological variability. When you work with experimental systems with no biological variability and little experimental error, heed 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 can't avoid large amounts of variability, yet care about relatively small differences. Statistical methods are necessary to draw valid conclusions from such data.

The key concept: Sampling from a population

Sampling from a population

The basic idea of statistics is simple: you want to extrapolate from the data you have collected to make general conclusions about the larger population from which the data sample was derived.

To do this, statisticians have developed methods based on a simple model: Assume that all your data are randomly sampled from an infinitely large population. Analyze this sample, and use the results to make inferences about the population.

This model is an accurate description of some situations. For example, quality control samples really are randomly selected from a large population. Clinical trials do not enroll a randomly selected sample of patients, but it is usually reasonable to extrapolate from the sample you studied to the larger population of similar patients.

In a typical experiment, you don't really sample from a population. But you do want to extrapolate from your data to a more general conclusion. The concepts of sample and population can still be used if you define the sample to be the data you collected, and the population to be the data you would have collected if you had repeated the experiment an infinite number of times.

Note that the term *sample* has a specific definition in statistics that is very different than its usual definition. Learning new meanings for commonly used words is part of the challenge of learning statistics.

The need for independent samples

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 concept of independence can be difficult to grasp. Consider the following three situations.

- 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 biochemical experiment three times, each time in triplicate. You do not have nine independent values, as an error in preparing the reagents for one experiment could affect all three triplicates. 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.

Data are independent when any random factor that causes a value to be too high or too low affects only that one value. If a random factor (that you didn't account for in the analysis of the data) can affect more than one value, but not all of the values, then the data are not independent.

How you can use statistics to extrapolate from sample to population

Statisticians have devised three basic approaches to make conclusions about populations from samples of data:

The first method is to assume that the populations follow a special distribution, known as the Gaussian (bell shaped) distribution. Once you assume that a population is distributed in that manner, statistical tests let you make inferences about the mean (and other properties) of the population. Most commonly used statistical tests assume that the population is Gaussian.

The second method is to rank all values from low to high, and then compare the distribution of ranks. This is the principle behind most

commonly used nonparametric tests, which are used to analyze data from non-Gaussian distributions.

The third method is known as resampling. With this method, you create a population of sorts, by repeatedly sampling values from your sample. This is best understood by an example. Assume you have a single sample of five values, and want to know how close that sample mean is likely to be from the true population mean. Write each value on a card and place the cards in a hat. Create many pseudo samples by drawing a card from the hat, and returning it. Generate many samples of $N=5$ this way. Since you can draw the same value more than once, the samples won't all be the same. When randomly selecting cards gets tedious, use a computer program instead. The distribution of the means of these computer-generated samples gives you information about how accurately you know the mean of the entire population. The idea of resampling can be difficult to grasp. To learn about this approach to statistics, read the instructional material available at www.resample.com. Prism does not perform any tests based on resampling. Resampling methods are closely linked to bootstrapping methods.

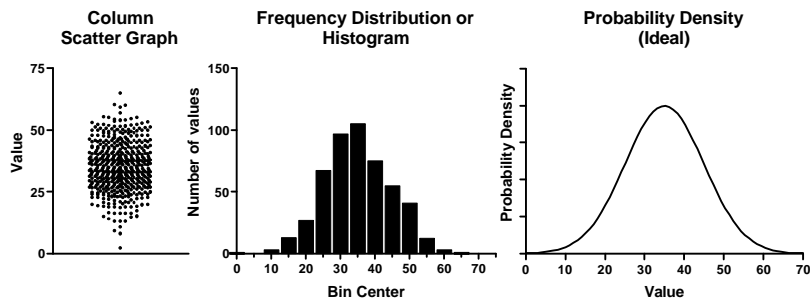
Limitations of statistics

The statistical model is simple: Extrapolate from the sample you collected to a more general situation, assuming that your sample was randomly selected from a large population. The problem is that the statistical inferences can only apply to the population from which your samples were obtained, but you often want to make conclusions that extrapolate even beyond that large population. For example, 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. Unfortunately, statistical calculations can't help with this further extrapolation. You must use scientific judgment and common sense to make inferences that go beyond the limitations of statistics. Thus, statistical logic is only part of data interpretation.

The Gaussian distribution

When many independent random factors act in an additive manner to create variability, data will follow a bell-shaped distribution called the Gaussian distribution, illustrated in the figure below. The left panel shows the distribution of a large sample of data. Each value is shown as a dot,

with the points moved horizontally to avoid too much overlap. This is called a column scatter graph. The frequency distribution, or histogram, of the values is shown in the middle panel. It shows the exact distribution of values in this particular sample. The right panel shows an ideal Gaussian distribution.



The Gaussian distribution has some special mathematical properties that form the basis of many statistical tests. Although no data follow that mathematical ideal exactly, many kinds of data follow a distribution that is approximately Gaussian.

The Gaussian distribution is also called a Normal distribution. Don't confuse this use of the word "normal" with its usual meaning.

The Gaussian distribution plays a central role in statistics because of a mathematical relationship known as the Central Limit Theorem. To understand this theorem, follow this imaginary experiment:

1. Create a population with a known distribution (which does not have to be Gaussian).
2. Randomly pick many samples from that population. Tabulate the means of these samples.
3. Draw a histogram of the frequency distribution of the means.

The central limit theorem says that if your samples are large enough, the distribution of means will follow a Gaussian distribution even if the population is not Gaussian. Since most statistical tests (such as the t test and ANOVA) are concerned only with differences between means, the Central Limit Theorem lets these tests work well even when the populations are not Gaussian. For this to be valid, the samples have to be reasonably large. How large is that? It depends on how far the population distribution differs from a Gaussian distribution. Assuming the population doesn't have a really weird distribution, a sample size of ten or so is generally enough to invoke the Central Limit Theorem.

To learn more about why the ideal Gaussian distribution is so useful, read about the *Central Limit Theorem* in any statistics text.

Confidence intervals

Confidence interval of a mean

The mean you calculate from a sample is not likely to be exactly equal to the population mean. The size of the discrepancy depends on the size and variability of the sample. If your sample is small and variable, the sample mean may be quite far from the population mean. If your sample is large with little scatter, the sample mean will probably be very close to the population mean. Statistical calculations combine sample size and variability (standard deviation) to generate a confidence interval (CI) for the population mean. You can calculate intervals for any desired degree of confidence, but 95% confidence intervals are most common. If you assume that your sample is randomly selected from some population (that follows a Gaussian distribution), 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 usually don't know the population mean, you'll never know when this happens.

Confidence intervals of other parameters

Statisticians have derived methods to generate confidence intervals for almost any situation. For example when comparing groups, you can calculate the 95% confidence interval for the difference between the group means. 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.

Similarly, methods exist to compute a 95% confidence interval for a proportion, the ratio of proportions, the best-fit slope of linear regression, and almost any other statistical parameter.

Nothing special about 95%

There is nothing magical about 95%. You could calculate confidence intervals for any desired degree of confidence. If you want more confidence that an interval contains the true parameter, then the intervals have to be wider. So 99% confidence intervals are wider than 95% confidence intervals, and 90% confidence intervals are narrower than 95% confidence intervals.

P values

What is a P value?

Assume that you've collected data from two samples of animals treated with different drugs. You've measured an enzyme in each animal's plasma, and the means are different. You want to know whether that difference is due to an effect of the drug – whether the two populations have different means. Observing different sample means is not enough to persuade you to conclude that the populations have different means. It is possible that the populations have the same mean (the drugs have no effect on the enzyme you are measuring), and that the difference you observed is simply a coincidence. There is no way you can ever be sure if the difference you observed reflects a true difference or if it is just a coincidence of random sampling. All you can do is calculate probabilities.

Statistical calculations can answer this question: If the populations really have the same mean, what is the probability of observing such a large difference (or larger) between sample means in an experiment of this size? The answer to this question is called the *P value*.

The P value is a probability, with a value ranging from zero to one. If the P value is small, you'll conclude that the difference between sample means is unlikely to be a coincidence. Instead, you'll conclude that the populations have different means.

What is a null hypothesis?

When statisticians discuss P values, they use the term *null hypothesis*. The null hypothesis simply states that there is no difference between the groups. Using that term, you can define the P value to be the probability of observing a difference as large or larger than you observed if the null hypothesis were true.

Common misinterpretation of a P value

Many people misunderstand P values. 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 (the null hypothesis is true). 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. However, this would be an incorrect conclusion. 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. This distinction may be more clear after you read "A Bayesian perspective on interpreting statistical significance" on page 15.

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.

Note: This example is for an unpaired t test that compares the means of two groups. The same ideas can be applied to other statistical tests.

The two-tail P value answers this question: Assuming the null hypothesis is true, what is the chance that randomly selected samples would have means as far apart as (or further than) you 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 is true, what is the chance that randomly selected samples would have means as far apart as (or further than) 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 in the same manner.

You should only choose a one-tail P value when two things are true.

- You must predict which group will have the larger mean (or proportion) before you collect any data.
- If the other group ends up with the larger mean – even if it is quite a bit larger – you must be willing to attribute that difference to chance.

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

- The relationship between P values and confidence intervals is easier to understand 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 to use a one-tail P value, observed a large difference between means, but the "wrong" group had the larger mean? In other

words, the observed difference was in the opposite direction to your experimental hypothesis. To be rigorous, you must conclude that the difference is due to chance, even if the difference is huge. While tempting, it is not fair 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.

Hypothesis testing and statistical significance

Statistical hypothesis testing

Much of statistical reasoning was developed in the context of quality control where you need a definite yes or no answer from every analysis. Do you accept or reject the batch? The logic used to obtain the answer is called *hypothesis testing*.

First, define 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 practice, the threshold value (called α) is almost always set to 0.05 (an arbitrary value that has been widely adopted).

Next, define the *null hypothesis*. If you are comparing two means, the null hypothesis is that the two populations have the same mean. In most circumstances, the null hypothesis is the opposite of the experimental hypothesis that the means come from different populations.

Now, perform the appropriate statistical test to compute the P 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". You cannot conclude that the null hypothesis is true. All you can do is conclude that you don't have sufficient evidence to reject the null hypothesis.

Statistical significance in science

The term *significant* is seductive, and easy to misinterpret. Using the conventional definition with $\alpha = 0.05$, a result is said to be *statistically significant* when the result would occur less than 5% of the time 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 biologically or clinically important or interesting. Moreover, 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. By chance, you obtained larger values in one group and smaller values in the other. Finding a statistically significant result when the populations are identical is called making a Type I error. If you define statistically significant to mean " $P < 0.05$ ", then you'll make a Type I error in 5% of experiments where there really is no difference.
- The populations really are different, so your conclusion is correct. The difference may be large enough to be scientifically interesting. Or it may be tiny and trivial.

"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. Degrees of statistical significance are not distinguished. Some statisticians feel very strongly about this. Many scientists are not so rigid, and refer to results as being "barely significant", "very significant" or "extremely significant".

Prism summarizes the P value using the words in the middle column of this table. Many scientists label graphs with the symbols of the third column. These definitions are not entirely standard. If you report the results in this way, you should define the symbols in your figure legend.

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

Report the actual P value

The concept of statistical hypothesis testing works well for quality control, where you must decide to accept or reject an item or batch based on the results of a single analysis. Experimental science is more complicated than that, because you often integrate many kinds of experiments before reaching conclusions. You don't need to make a significant/not significant decision for every P value. Instead, report exact P values, so you and your readers can interpret them as part of a bigger picture.

The tradition of reporting only “ $P < 0.05$ ” or “ $P > 0.05$ ” began in the days before computers were readily available. Statistical tables were used to determine whether the P value was less than or greater than 0.05, and it would have been very difficult to determine the P value exactly. Today, with most statistical tests, it is very easy to compute exact P values, and you shouldn’t feel constrained to only report whether the P value is less than or greater than some threshold value.

Statistical power

Type II errors and power

When a study reaches a conclusion of “no statistically significant difference”, you should not necessarily conclude that the treatment was ineffective. It is possible that the study missed a real effect because you used a small sample or your data were quite variable. In this case you made a Type II error — obtaining a “not significant” result when in fact there is a difference.

When interpreting the results of an experiment that found no significant difference, you need to ask yourself how much power the study had to find various hypothetical differences if they existed. The power depends on the sample size and amount of variation within the groups, where variation is quantified by the standard deviation (SD).

Here is a precise definition of power: Start with the assumption that the two population means differ by a certain amount and that the SD of the populations has a particular value. Now assume that you perform many experiments with the sample size you used, and calculate a P value for each experiment. Power is the fraction of these experiments that would have a P value less than α (the largest P value you deem “significant”, usually set to 0.05). In other words, power equals the fraction of experiments that would lead to statistically significant results. Prism does not compute power, but the companion program GraphPad StatMate does.

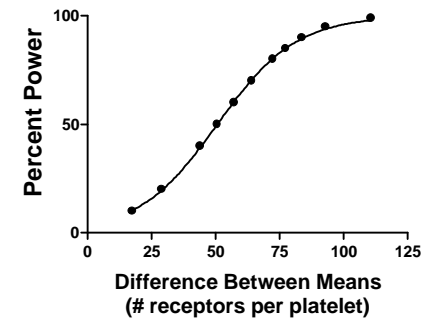
Example of power calculations

Motulsky et al. asked whether people with hypertension (high blood pressure) had altered numbers of α_2 -adrenergic receptors on their platelets (Clinical Science 64:265-272, 1983). There are many reasons to think that autonomic receptor numbers may be altered in hypertensives. They studied platelets because they are easily accessible from a blood sample. The results are shown here:

Variable	Hypertensives	Controls
Number of subjects	18	17
Mean receptor number (receptors per cell)	257	263
Standard Deviation	59.4	86.6

The two means were almost identical, and a t test gave a very high P value. The authors concluded that the platelets of hypertensives do not have an altered number of α_2 receptors.

What was the power of this study to find a difference if there was one? The answer depends on how large the difference really is. Prism does not compute power, but the companion program GraphPad StatMate does. Here are the results shown as a graph.



If the true difference between means was 50.58, then this study had only 50% power to find a statistically significant difference. In other words, if hypertensives really averaged 51 more receptors per cell, you’d find a statistically significant difference in about half of studies of this size, but would not find a statistically significant difference in the other half of the studies. This is about a 20% change (51/257), large enough that it could possibly have a physiological impact.

If the true difference between means was 84 receptors/cell, then this study had 90% power to find a statistically significant difference. If hypertensives really had such a large difference, you’d find a statistically significant difference in 90% of studies this size and would find a not significant difference in the other 10% of studies.

All studies have low power to find small differences and high power to find large differences. However, it is up to you to define “low” and “high” in the context of the experiment, and to decide whether the power was high enough for you to believe the negative results. If the power is too low, you shouldn’t reach a firm conclusion until the study has been repeated with

more subjects. Most investigators aim for 80% or 90% power to detect a difference.

Since this study had only a 50% power to detect a difference of 20% in receptor number (50 sites per platelet, a large enough difference to possibly explain some aspects of hypertension physiology), the negative conclusion is not solid.

A Bayesian perspective on interpreting statistical significance

Imagine that you are screening drugs to see if they lower blood pressure. Based on the amount of scatter you expect to see, and the minimum change you would care about, you've chosen the sample size for each experiment to have 80% power to detect the difference you are looking for with a P value less than 0.05. What happens when you repeat the experiment many times?

The answer is "it depends". It depends on the context of your experiment. Let's look at the same experiment performed in three contexts. First, we'll assume that you know a bit about the pharmacology of the drugs, and expect 10% of the drugs to be active. In this case, the *prior probability* is 10%. Second, we'll assume you know a lot about the pharmacology of the drugs, and expect 80% to be active. Third, we'll assume that the drugs were selected at random, and you expect only 1% to be active in lowering blood pressure.

What happens when you perform 1000 experiments in each of these contexts. The details of the calculations are shown on pages 143-145 of *Intuitive Biostatistics*, by Harvey Motulsky (Oxford University Press, 1995). Since the power is 80%, you expect 80% of truly effective drugs to yield a P value less than 0.05 in your experiment. Since you set the definition of statistical significance to 0.05, you expect 5% of ineffective drugs to yield a P value less than 0.05. Putting these calculations together creates these tables.

A. Prior probability=10%	Drug really works	Drug really doesn't work	Total
P<0.05 "significant"	80	45	125
P>0.05, "not significant"	20	855	875
Total	100	900	1000

B. Prior probability=80%	Drug really works	Drug really doesn't work	Total
P<0.05 "significant"	640	10	650
P>0.05, "not significant"	160	190	350
Total	800	200	1000

C. Prior probability=1%	Drug really works	Drug really doesn't work	Total
P<0.05 "significant"	8	50	58
P>0.05, "not significant"	2	940	942
Total	10	990	1000

The total for each column is determined by the prior probability – the context of your experiment. The prior probability equals the fraction of the experiments that are in the leftmost column. To compute the number of experiments in each row, use the definition of power and alpha. Of the drugs that really work, you won't obtain a P value less than 0.05 in every case. You chose a sample size to obtain a power of 80%, so 80% of the truly effective drugs yield "significant" P values, and 20% yield "not significant" P values. Of the drugs that really don't work (middle column), you won't get "not significant" results in every case. Since you defined statistical significance to be "P<0.05" (alpha=0.05), you will see a "significant" result in 5% of experiments performed with drugs that are really inactive and a "not significant" result in the other 95%.

If the P value is less than 0.05, so the results are "statistically significant", what is the chance that the drug is really active? The answer is different for each experiment.

Prior probability	Experiments with P<0.05 and...		Fraction of experiments with P<0.05 where drug really works
	...drug really works	...drug really doesn't work	
A. Prior probability=10%	80	45	80/125 = 64%
B. Prior probability=80%	640	10	640/650 = 98%
C. Prior probability=1%	8	50	8/58 = 14%

For experiment A, the chance that the drug is really active is 80/125 or 64%. If you observe a statistically significant result, there is a 64% chance that the difference is real and a 36% chance that the difference was caused by random sampling. For experiment B, there is a 98.5% chance that the difference is real. In contrast, if you observe a significant result in

experiment C, there is only a 14% chance that the result is real and a 86% chance that it is a coincidence of random sampling. For experiment C, the vast majority of “significant” results are due to chance.

Your interpretation of a “statistically significant” result depends on the context of the experiment. You can’t interpret a P value in a vacuum. Your interpretation depends on the context of the experiment. Interpreting results requires common sense, intuition, and judgment.

Beware of multiple comparisons

Interpreting an individual P value is easy. If the null hypothesis is true, the P value is the chance that random selection of subjects would result in a difference (or correlation or association...) as large (or larger) than observed in your study. If the null hypothesis is true, there is a 5% chance of randomly selecting subjects such that the trend is statistically significant.

However, many scientific studies generate more than one P value. Some studies in fact generate hundreds of P values. Interpreting multiple P values can be difficult.

If you test several independent null hypotheses, and leave the threshold at 0.05 for each comparison, there is greater than a 5% chance of obtaining at least one “statistically significant” result by chance. The second column in the table below shows you how much greater.

Number of Independent Null Hypotheses	Probability of obtaining one or more P values less than 0.05 by chance	Threshold to keep overall risk of type I error equal to 0.05
1	5%	0.0050
2	10%	0.0253
3	14%	0.0170
4	19%	0.0127
5	23%	0.0102
6	26%	0.0085
7	30%	0.0073
8	34%	0.0064
9	37%	0.0057
10	40%	0.0051
20	64%	0.0026
50	92%	0.0010
100	99%	0.0005
N	$100(1.00 - 0.95^N)$	$1.00 - 0.95^{(1/N)}$

To maintain the chance of randomly obtaining at least one statistically significant result at 5%, you need to set a stricter (lower) threshold for each individual comparison. This is tabulated in the third column of the table. If you only conclude that a difference is statistically significant when a P value is less than this value, then you’ll have only a 5% chance of finding any “significant” difference by chance among all the comparisons.

For example, if you test three null hypotheses and use the traditional cutoff of $\alpha = 0.05$ for declaring each P value to be significant, there would be a 14% chance of observing one or more significant P values, even if all three null hypotheses were true. To keep the overall chance at 5%, you need to lower the threshold for significance to 0.0170.

If you compare three or more groups, account for multiple comparisons using post tests following one-way ANOVA. These methods account both for multiple comparisons and the fact that the comparisons are not independent. See “How post tests work” on page 74.

You can only account for multiple comparisons when you know about all the comparisons made by the investigators. If you report only “significant” differences, without reporting the total number of comparisons, others will not be able to properly evaluate your results. Ideally, you should plan all your analyses before collecting data, and then report all the results.

Distinguish between studies that test a hypothesis from studies that generate a hypothesis. Exploratory analyses of large databases can generate hundreds of P values, and scanning these can generate intriguing research hypotheses. To test these hypotheses, you’ll need a different set of data.

Outliers

What is an outlier?

When analyzing data, you’ll sometimes find that one value is far from the others. Such a value is called an *outlier*, a term that is usually not defined rigorously. When you encounter an outlier, you may be tempted to delete it from the analyses. First, ask yourself these questions:

- Was the value entered into the computer correctly? If there was an error in data entry, fix it.
- Were there any experimental problems with that value? For example, if you noted that one tube looked funny, you have justification to exclude the value resulting from that tube without needing to perform any calculations.
- Is the outlier caused by biological diversity? If each value comes from a different person or animal, the outlier may be a correct value. It is an outlier not because of an experimental mistake, but rather because

that individual may be different from the others. This may be the most exciting finding in your data!

After answering no to those three questions, you have to decide what to do with the outlier. There are two possibilities.

- One possibility is that the outlier was due to chance. In this case, you should keep the value in your analyses. The value came from the same distribution as the other values, so should be included.
- The other possibility is that the outlier was due to a mistake - bad pipetting, voltage spike, holes in filters, etc. Since including an erroneous value in your analyses will give invalid results, you should remove it. In other words, the value comes from a different population than the other and is misleading.

The problem, of course, is that you can never be sure which of these possibilities is correct.

No mathematical calculation can tell you for sure whether the outlier came from the same or different population than the others. But statistical calculations can answer this question: If the values really were all sampled from a Gaussian distribution, what is the chance that you'd find one value as far from the others as you observed? If this probability is small, then you will conclude that the outlier is likely to be an erroneous value, and you have justification to exclude it from your analyses.

Statisticians have devised several methods for detecting outliers. All the methods first quantify how far the outlier is from the other values. This can be the difference between the outlier and the mean of all points, the difference between the outlier and the mean of the remaining values, or the difference between the outlier and the next closest value. Next, standardize this value by dividing by some measure of scatter, such as the SD of all values, the SD of the remaining values, or the range of the data. Finally, compute a P value answering this question: If all the values were really sampled from a Gaussian population, what is the chance of randomly obtaining an outlier so far from the other values? If the P value is small, you conclude that the deviation of the outlier from the other values is statistically significant.

Prism does not perform any sort of outlier detection. If you want to perform an outlier test by hand, you can calculate Grubb's test, described below.

Detecting outliers with Grubb's test

The Grubbs' method for assessing outliers is particularly easy to understand. This method is also called the ESD method (extreme studentized deviate).

The first step is to quantify how far the outlier is from the others. Calculate the ratio Z as the difference between the outlier and the mean divided by the SD. Calculate the mean and SD from all the values, including the outlier. Calculate Z for all the values, but only perform the Grubb's test with the most extreme outlier, the value that leads to the largest value of Z.

$$z = \frac{|\text{mean} - \text{value}|}{\text{SD}}$$

Since 5% of the values in a Gaussian population are more than 1.96 standard deviations from the mean, your first thought might be to conclude that the outlier comes from a different population if Z is greater than 1.96. This approach would only work when you know the *population* mean and SD from other data. Although this is rarely the case in experimental science, it is often the case in quality control. You know the overall mean and SD from historical data, and want to know whether the latest value matches the others. This is the basis for quality control charts.

When analyzing experimental data, you don't know the SD of the population. Instead, you calculate the SD from the data. The presence of an outlier increases the calculated SD. Since the presence of an outlier increases both the numerator (difference between the value and the mean) and denominator (SD of all values), Z does not get very large. In fact, no matter how the data are distributed, Z can not get larger than $(N-1)/\sqrt{N}$ where N is the number of values. For example, if N = 3, Z cannot be larger than 1.555 for any set of values.

Grubbs and others have tabulated critical values for Z, which are tabulated below for P=0.05 (two-tail). The critical value increases with sample size, as expected.

If your calculated value of Z is greater than the critical value in the table, then the P value is less than 0.05. This means that there is less than a 5% chance that you'd encounter an outlier so far from the others (in either direction) by chance alone, if all the data were really sampled from a single Gaussian distribution. Note that the method only works for testing the most extreme value in the sample (if in doubt, calculate Z for all values, but only calculate a P value for Grubbs' test from the largest value of Z).

N	Critical Z 5%	N	Critical Z 5%
3	1.15	27	2.86
4	1.48	28	2.88
5	1.71	29	2.89
6	1.89	30	2.91
7	2.02	31	2.92
8	2.13	32	2.94
9	2.21	33	2.95
10	2.29	34	2.97
11	2.34	35	2.98
12	2.41	36	2.99
13	2.46	37	3.00
14	2.51	38	3.01
15	2.55	39	3.03
16	2.59	40	3.04
17	2.62	50	3.13
18	2.65	60	3.20
19	2.68	70	3.26
20	2.71	80	3.31
21	2.73	90	3.35
22	2.76	100	3.38
23	2.78	110	3.42
24	2.80	120	3.44
25	2.82	130	3.47
26	2.84	140	3.49

Consult the references below for larger tables. You can also calculate an approximate P value as follows.

1. Calculate a t ratio from N (number of values in the sample) and Z (calculated for the suspected outlier as shown above).

$$t = \sqrt{\frac{N(N-2)Z^2}{(N-1)^2 - NZ^2}}$$

2. Determine the P value corresponding with that value of t with N-2 degrees of freedom. Use the Excel formula =TDIST(t,df,2), substituting values for t and df (the third parameter is 2, because you want a two-tailed P value).
3. Multiply the P value you obtain in step 2 by N. The result is an approximate P value for the outlier test. This P value is the chance of observing one point so far from the others if the data were all

sampled from a Gaussian distribution. If Z is large, this P value will be very accurate. With smaller values of Z, the calculated P value may be too large

The most that Grubbs' test (or any outlier test) can do is tell you that a value is unlikely to have come from the same Gaussian population as the other values in the group. You then need to decide what to do with that value. I would recommend removing significant outliers from your calculations in situations where experimental mistakes are common and biological variability is not a possibility. When removing outliers, be sure to document your decision. Others feel that you should never remove an outlier unless you noticed an experimental problem. Beware of a natural inclination to remove outliers that get in the way of the result you hope for, but to keep outliers that enhance the result you hope for.

If you use nonparametric tests, outliers will affect the results very little so do not need to be removed.

If you decide to remove the outlier, you then may be tempted to run Grubbs' test again to see if there is a second outlier in your data. If you do this, you cannot use the table shown above. Rosner has extended the method to detecting several outliers in one sample. See the first reference below for details.

Here are two references:

- B Iglewicz and DC Hoaglin. [How to Detect and Handle Outliers](#) (Asqc Basic References in Quality Control, Vol 16) Amer Society for Quality Control, 1993.
- V Barnett, T Lewis, V Rothamsted. [Outliers in Statistical Data](#) (Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics) John Wiley & Sons, 1994.

Analyzing one group

Entering data to analyze one group

Prism can calculate descriptive statistics of a column of values, test whether the group mean or median differs significantly from a hypothetical value, and test whether the distribution of values differs significantly from a Gaussian distribution. Prism can also create a frequency distribution histogram.

When you choose a format for the data table, choose a single column of Y values. Choose to have no X column (or choose an X column for labeling only). Enter the values for each group into a separate Y column. If your data table includes an X column, be sure to use it only for labels. Values you enter into the X column will not be analyzed by the analyses discussed here.

If you format the Y columns for entry of replicate values (for example, triplicates), Prism first averages the replicates in each row. It then calculates the column statistics on the means, without considering the SD or SEM of each row. If you enter ten rows of triplicate data, the column statistics are calculated from the ten row means, not from the 30 individual values. If you format the Y columns for entry of mean and SD (or SEM) values, Prism calculates column statistics for the means and ignores the SD or SEM values you entered.

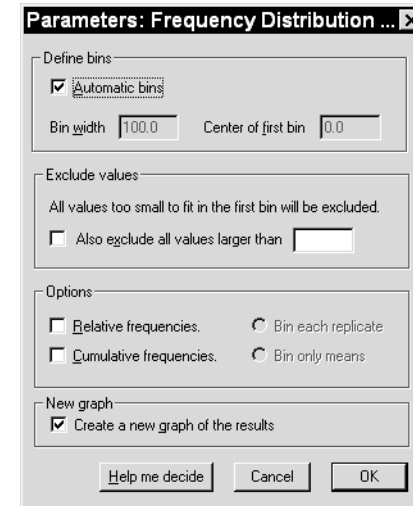
Prism provides two analyses for analyzing columns of numbers.

- Use the **frequency distribution** analysis to tabulate the distribution of values in the column (and create a histogram).
- Use the **column statistics** analysis to compute descriptive statistics, to test for normality, and to compare the mean (or median) of the group against a theoretical value using a one-sample t test or the Wilcoxon rank sum test.

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.

Click Analyze and select built-in analyses. Then choose Frequency Distribution from the list of statistical analyses to bring up the parameters dialog.



Histograms generally look best when the bin width is a round number and there are 10-20 bins. Either use the default settings, or clear the “auto” option to enter the center of the first bin and the width of all bins.

If you entered replicate values, Prism can either place each replicate into its appropriate bin, or average the replicates and only place the mean into a bin.

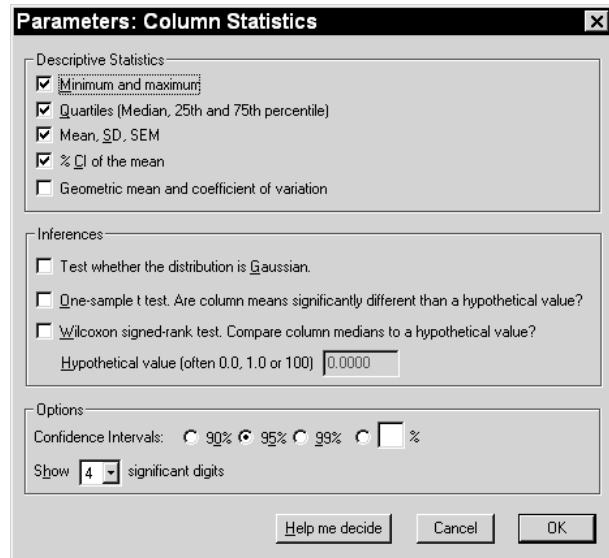
All values too small to fit in the first bin are omitted from the analysis. Enter an upper limit to omit larger values from the analysis.

Select **Relative frequencies** to determine the fraction of values in each bin, rather than the number of values in each bin. For example, if you have 50 data points of which 15 fall into the third bin, the results for the third bin will be 0.30 (15/50) rather than 15.

Select **Cumulative binning** to see the cumulative distribution. Each bin contains the number of values that fall within or below that bin. By definition, the last bin contains the total number of values.

Column statistics

From your data table, click Analyze and choose built-in analyses. Then choose Column Statistics from the list of statistical analyses to bring up the parameters dialog.



Choose the descriptive statistics you want to determine. Note that CI means confidence interval.

Check "test whether the distribution is Gaussian" to perform a normality test.

To determine whether the mean of each column differs from some hypothetical value, choose the one-sample t test, and enter the hypothetical value, which must come from outside this experiment. The hypothetical value could be a value that indicates no change such as 0.0, 1.0 or 100, or it could come from theory or previous experiments.

The one-sample t test assumes that the data are sampled from a Gaussian distribution. The Wilcoxon rank sum test compares whether the median of each column differs from a hypothetical value. It does not assume that the distribution is Gaussian, but generally has less power than the one-sample t test. Choosing between a one-sample t test and a Wilcoxon rank sum test involves similar considerations to choosing between an unpaired t test and the Mann-Whitney nonparametric test. See "t test or nonparametric test?" on page 42.

To interpret the results, see "Interpreting descriptive statistics" on page 26, "The results of normality tests" on page 29, "The results of a one-sample t test" on page 30, or "The results of a Wilcoxon rank sum test" on page 34.

Interpreting descriptive statistics

SD and CV

The **standard deviation** (SD) quantifies variability. 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.

Prism calculates the SD using the equation below. (Each y_i is a value, y_{mean} is the average, and N is sample size).

$$SD = \sqrt{\frac{\sum (y_i - y_{\text{mean}})^2}{N-1}}$$

The standard deviation computed this way (with a denominator of $N-1$) is called the *sample SD*, in contrast to the *population SD* which would have a denominator of N . Why is the denominator $N-1$ rather than N ? In the numerator, you compute the difference between each value and the mean of those values. You don't know the true mean of the population; all you know is the mean of your sample. Except for the rare cases where the sample mean happens to equal the population mean, the data will be closer to the sample mean than it will be to the population mean. This means that the numerator will be too small. So the denominator is reduced as well. It is reduced to $N-1$ because that is the number of *degrees of freedom* in your data. There are $N-1$ degrees of freedom, because you could calculate the remaining value from $N-1$ of the values and the sample mean.

The **coefficient of variation** (CV) equals 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. It only makes sense to report CV for variables where zero really means zero, such as mass or enzyme activity. Don't calculate CV for variables, such as temperature, where the definition of zero is somewhat arbitrary.

SEM

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 by this equation:

$$\text{SEM} = \frac{\text{SD}}{\sqrt{N}}$$

With large samples, 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.

95% Confidence interval

The **confidence interval** quantifies the precision of the mean. 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% CIs 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% of the cases. Since you don't know the population mean, you'll never know when this happens.

The confidence interval extends in each direction by a distance calculated from the standard error of the mean multiplied by a critical value from the t distribution. This value depends on the degree of confidence you want (traditionally 95%, but it is possible to calculate intervals for any degree of confidence) and on the number of degrees of freedom in this experiment (N-1). With large samples, this multiplier equals 1.96. With smaller samples, the multiplier is larger.

Quartiles and range

Quartiles divide the data into four groups, each containing an equal number of values. Quartiles divided by the 25th percentile, 50th percentile, 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.

Prism computes percentile values by first computing $R = P \cdot (N + 1) / 100$, where P is 25, 50 or 75 and N is the number of values in the data set. The result is the rank that corresponds to the percentile value. If there are 68 values, the 25th percentile corresponds to a rank equal to $.25 \cdot 69 = 17.25$. So the 25th percentile lies between the value of the 17th and 18th value (when ranked from low to high). Prism computes the 25th percentile as the average of those two values (some programs report the value one quarter of the way between the two).

Geometric mean

The geometric mean is the antilog of the mean of the logarithm of the values. It is less affected by outliers than the mean.

Which error bar should you choose?

It is easy to be confused about the difference between the standard deviation (SD) and standard error of the mean (SEM).

The SD quantifies scatter — how much the values vary from one another.

The SEM quantifies how accurately you know the true mean of the population. The SEM gets smaller as your samples get larger. This makes sense, because the mean of a large sample is likely to be closer to the true population mean than is the mean of a small sample.

The SD does not change predictably as you acquire more data. The SD quantifies the scatter of the data, and increasing the size of the sample does not increase the scatter. The SD might go up or it might go down. You can't predict. On average, the SD will stay the same as sample size gets larger.

If the scatter is caused by biological variability, your probably will want to show the variation. In this case, graph the SD rather than the SEM. You could also instruct Prism to graph the range, with error bars extending from the smallest to largest value. Also consider graphing every value, rather than using error bars.

If you are using an *in vitro* system with no biological variability, the scatter can only result from experimental imprecision. In this case, you may not want to show the scatter, but instead show how well you have assessed the mean. Graph the mean and SEM or the mean with 95% confidence intervals.

Ideally, the choice of which error bar to show depends on the source of the variability and the point of the experiment. In fact, many scientists always show the mean and SEM, to make the error bars as small as possible.

Other results

For other results, see "The results of normality tests" on page 29, "The results of a one-sample t test" on page 30, or "The results of a Wilcoxon rank sum test" on page 34.

The results of normality tests

How the normality test works

Prism tests for deviations from Gaussian distribution using the Kolmogorov-Smirnov (KS) test. Since the Gaussian distribution is also called the Normal

distribution, the test is called a normality test. The KS statistic (which some other programs call D) quantifies the discrepancy between the distribution of your data and an ideal Gaussian distribution – larger values denoting larger discrepancies. It is not informative by itself, but is used to compute a P value.

Prism calculates KS using the method of Kolmogorov and Smirnov. However, the method originally published by those investigators cannot be used to calculate the P value because their method assumes that you know the mean and SD of the overall population (perhaps from prior work). When analyzing data, you rarely know the overall population mean and SD. You only know the mean and SD of your sample. To compute the P value, therefore, Prism uses the Dallal and Wilkinson approximation to Lilliefors' method (*Am. Statistician*, 40:294-296, 1986). Since that method is only accurate with small P values, Prism simply reports "P>0.10" for large P values.

How to think about results from a normality test

The P value from the normality test answers this question: If you randomly sample from a Gaussian population, what is the probability of obtaining a sample that deviates from a Gaussian distribution as much (or more so) as this sample does? More precisely, the P value answers this question: If the population is really Gaussian, what is the chance that a randomly selected sample of this size would have a KS value as large as, or larger than, observed?

By looking at the distribution of a small sample of data, it is hard to tell whether or not the values came from a Gaussian distribution. Running a formal test does not make it easier. The tests simply have little power to discriminate between Gaussian and non-Gaussian populations with small sample sizes. How small? If you have fewer than five values, Prism doesn't even attempt to test for normality. But the test doesn't really have much power to detect deviations from Gaussian distribution unless you have several dozen values.

Your interpretation of the results of a normality test depends on the P value calculated by the test and on the sample size.

P value	Sample size	Conclusion
Small (>0.05)	Any	The data failed the normality test. You can conclude that the population is unlikely to be Gaussian.
Large	Large	The data passed the normality test. You can conclude that the population is likely to be Gaussian, or nearly so. How large does the sample have to be? There is no firm answer, but one rule-of-thumb is that the normality tests are only useful when your sample size is a few dozen or more.
Large	Small	You will be tempted to conclude that the population is Gaussian. But that conclusion may be incorrect. A large P value just means that the data are not inconsistent with a Gaussian population. That doesn't exclude the possibility of a non-Gaussian population. Small sample sizes simply don't provide enough data to discriminate between Gaussian and non-Gaussian distributions. You can't conclude much about the distribution of a population if your sample contains fewer than a dozen values.

The results of a one-sample t test

How a one-sample t test works

A one-sample t test compares the mean of a single column of numbers against a hypothetical mean you entered. Prism calculates the t ratio from this equation:

$$t = \frac{\text{Sample Mean} - \text{Hypothetical Mean}}{\text{Standard Error of Mean}}$$

A P value is computed from the t ratio and the numbers of degrees of freedom (which equals sample size minus 1).

How to think about results from the one-sample t test

Look first at the P value, which answers this question: If the data were sampled from a Gaussian population with a mean equal to the hypothetical value you entered, what is the chance of randomly selecting N data points and finding a mean as far from the hypothetical value (or further) as observed here?

“Statistically significant” is not the same as “scientifically important”. Before interpreting the P value or confidence interval, you should think about the size of the difference you are seeking. How large a difference (between the population mean and the hypothetical mean) would you consider to be scientifically important? How small a difference would you consider to be scientifically trivial? You need to use scientific judgment and common sense to answer these questions. Statistical calculations cannot help, as the answers depend on the context of the experiment.

You will interpret the results differently depending on whether the P value is small or large.

If the P value is small (one-sample t test)

If the P value is small (usually defined to mean less than 0.05), then it is unlikely that the discrepancy you observed between sample mean and hypothetical mean is due to a coincidence arising from random sampling. You can reject the idea that the difference is a coincidence, and conclude instead that the population has a mean different than the hypothetical value you entered. The difference is statistically significant. But is the difference scientifically significant? The confidence interval helps you decide.

The true difference between population mean and hypothetical mean is probably not the same as the difference observed in this experiment. There is no way to know the true difference between the population mean and the hypothetical mean. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true difference between the overall (population) mean and the hypothetical value you entered.

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a discrepancy that is scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial	Trivial	Although the true difference is not zero (since the P value is low), the difference is tiny and uninteresting. The data have a mean distinct from the hypothetical value, but the discrepancy is too small to be scientifically interesting.

Trivial	Important	Since the confidence interval ranges from a difference that you think is biologically trivial to one you think would be important, you can't reach a strong conclusion from your data. You can conclude that the data has a mean distinct from the hypothetical value you entered, but don't know whether that difference is scientifically trivial or important. You'll need more data to obtain a clear conclusion.
Important	Important	Since even the low end of the confidence interval represents a difference large enough to be considered biologically important, you can conclude that the data have a mean distinct from the hypothetical value, and the discrepancy is large enough to be scientifically relevant.

If the P value is large

If the P value is large, the data do not give you any reason to conclude that the overall mean differs from the hypothetical value you entered. This is not the same as saying that the true mean equals the hypothetical value. You just don't have evidence of a difference.

How large could the true difference really be? Because of random variation, the difference between the hypothetical mean and the sample mean in this experiment is unlikely to be equal to the true difference between population mean and hypothetical mean. There is no way to know the true difference between the population mean and the hypothetical mean. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true difference between the overall (population) mean of the data and the hypothetical mean you entered. When the P value is larger than 0.05, the 95% confidence interval will start with a negative number (the hypothetical mean is larger than the actual mean) and go up to a positive number (the actual mean is larger than the hypothetical mean).

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent differences that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial	Trivial	You can reach a crisp conclusion. Either the data has a mean equal to the hypothetical mean or they differ by a trivial amount.
Trivial	Large	You can't reach a strong conclusion. The data are consistent with a mean slightly smaller than the hypothetical mean, equal to the hypothetical mean, or larger than the hypothetical mean, perhaps large enough to be scientifically important. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Large	Trivial	You can't reach a strong conclusion. The data are consistent with a mean smaller than the hypothetical mean (perhaps small enough to be scientifically important), equal to the hypothetical mean, or slightly larger than the hypothetical mean. You can't make a clear conclusion without repeating the experiment with more subjects.
Large	Large	You can't reach a strong conclusion. The data are consistent with a mean smaller than the hypothetical mean (perhaps small enough to be scientifically important), equal to the hypothetical mean, or larger than the hypothetical mean (perhaps large enough to be scientifically important). In other words, you can't draw any conclusion at all. You need to repeat the experiment with more subjects.

Checklist. Is a one-sample t test the right test for these data?

Before accepting the results of any statistical test, first think carefully about whether you chose an appropriate test. Before accepting results from a one-sample t test, ask yourself these questions:

Question	Discussion
Is the population distributed according to a Gaussian distribution?	The one sample t test assumes that you have sampled your data from a population that follows a Gaussian distribution. While this assumption is not too important with large samples, it is important with small sample sizes, especially when N is less than 10. Prism tests for violations of this assumption, but normality tests have limited utility. See "The results of normality tests" on page 29. If your data do not come from a Gaussian distribution, you have three options. Your best option is to transform the values to make the distribution more Gaussian, perhaps by transforming all values to their reciprocals or logarithms. Another choice is to use the Wilcoxon rank sum nonparametric test instead of the t test. A final option is to use the t test anyway, knowing that the t test is fairly robust to departures from a Gaussian distribution with large samples.
Are the "errors" independent?	The term "error" refers to the difference between each value and the group mean. The results of a t test only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. See "The need for independent samples" on page 5.
If you chose a one-tail P value, did you predict correctly?	If you chose a one-tail P value, you should have predicted whether the mean of your data would be larger than or smaller than the hypothetical mean. 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 "One- vs. two-tail P values" on page 10.

The results of a Wilcoxon rank sum test

How the Wilcoxon rank sum test works

A Wilcoxon rank sum test compares the median of a single column of numbers against a hypothetical median you entered.

Prism follows these steps:

1. Calculate how far each value is from the hypothetical median.
2. Ignore values that exactly equal the hypothetical value. Call the number of remaining values N.
3. Rank these distances, paying no attention to whether the values are higher or lower than the hypothetical value.
4. For each value that is lower than the hypothetical value, multiply the rank by negative 1.
5. Sum the positive ranks. Prism reports this value.
6. Sum the negative ranks. Prism also reports this value.
7. Add the two sums together. This is the sum of signed ranks, which Prism reports as W.

If the data really were sampled from a population with the hypothetical mean, you'd expect W to be near zero. If W (the sum of signed ranks) is far from zero, the P value will be small. The P value answers this question: Assuming that you randomly sample N values from a population with the hypothetical median, what is the chance that W will be as far from zero (or further) than you observed?

Don't confuse the Wilcoxon rank sum test (compare one group with hypothetical median) with the Wilcoxon matched pairs test (compare medians of two paired groups). See "The results of a Wilcoxon matched pairs test" on page 59.

With small samples, Prism computes an exact P value. With larger samples, Prism uses an approximation that is quite accurate.

How to think about the results of a Wilcoxon rank sum test

The Wilcoxon signed rank test is a nonparametric test that compares the median of one column of numbers to a theoretical median.

Look first at the P value, which answers this question: If the data were sampled from a population with a median equal to the hypothetical value you entered, what is the chance of randomly selecting N data points and finding a median as far (or further) from the hypothetical value as observed here?

If the P value is small, you can reject the idea that the difference is a coincidence, and conclude instead that the population has a median distinct from the hypothetical value you entered.

If the P value is large, the data do not give you any reason to conclude that the overall median differs from the hypothetical median. This is not the same as saying that the medians are the same. You just have no compelling evidence that they differ. If you have small samples, the Wilcoxon test has

little power. In fact, if you have five or fewer values, the Wilcoxon test will always give a P value greater than 0.05 no matter how far the sample median is from the hypothetical median.

Checklist. Is the Wilcoxon test right for these data?

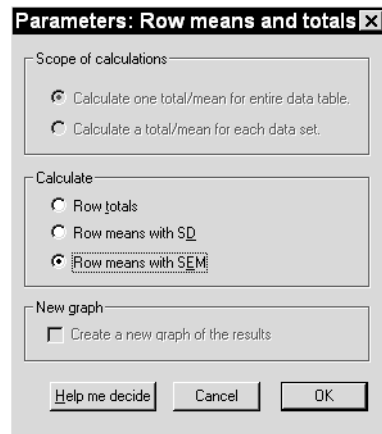
Before interpreting the results of any statistical test, first think carefully about whether you have chosen an appropriate test. Before accepting results from a Wilcoxon test, ask yourself these questions (Prism cannot help you answer them):

Question	Discussion
Are the "errors" independent?	The term "error" refers to the difference between each value and the group median. The results of a Wilcoxon test only make sense when the scatter is random – that any factor that causes a value to be too high or too low affects only that one value. Prism cannot test this assumption. See "The need for independent samples" on page 5.
Are the data clearly sampled from a non-Gaussian population?	By selecting a nonparametric test, you have avoided assuming that the data were sampled from a Gaussian distribution. But there are drawbacks to using a nonparametric test. If the populations really are Gaussian, the nonparametric tests have less power (are less likely to give you a small P value), especially with small sample sizes. Furthermore, Prism (along with most other programs) does not calculate confidence intervals when calculating nonparametric tests. If the distribution is clearly not bell-shaped, consider transforming the values (perhaps with logs or reciprocals) to create a Gaussian distribution and then using a one-sample t test.
Are the data distributed symmetrically?	The Wilcoxon test does not assume that the data are sampled from a Gaussian distribution. However it does assume that the data are distributed symmetrically around their median.
If you chose a one-tail P value, did you predict correctly?	If you chose a one-tail P value, you should have predicted which group has the larger median before collecting data. Prism does not ask you to record this prediction, but assumes it is correct. If your prediction was wrong, ignore the P value reported by Prism and state that $P > 0.50$. See "One- vs. two-tail P values" on page 10.

Row means and totals

If you enter data with replicate Y values, Prism will automatically graph mean and SD (or SEM). You don't have to choose any analyses. Use settings on the Symbols dialog (double-click on any symbol to see it) to plot individual points or to choose SD, SEM, 95%CI or range error bars.

To view a table of mean and SD (or SEM) values, use the Row Means or Totals analysis. Click Analyze and choose built-in analyses. Then choose Row means and totals from the list of data manipulations to bring up this dialog.



First, decide on the **scope of the calculations**. If you have entered more than one data set in the table, you have two choices. Most often, you'll calculate a row total/mean for each data set. The results table will have the same number of data sets as the input table. The other choice is to calculate one row total/mean for the entire table. The results table will then have a single data set containing the grand totals or means.

Then decide what to **calculate**: row totals, row means with SD or row means with SEM. To review the difference between SD and SEM see "Interpreting descriptive statistics" on page 26.

t tests and nonparametric comparisons

	A			B		
	Control			Treated		
	Y	SEM	N	Y	SEM	N
1	38.667000	2.6030	3	48.00000	2.0820	3

Introduction to comparing of two groups

Prism can compare two groups with a paired or unpaired t test, or with the nonparametric Mann-Whitney or Wilcoxon matched pairs test. These tests compare measurements (continuous variables) such as weight, enzyme activity, or receptor number. To compare two proportions see "Contingency tables" on page 121. To compare survival curves, see "Survival curves" on page 109.

Entering data to compare two groups with a t test (or a nonparametric test)

Enter the data into two Y columns (data sets). Format the data table for single columns of Y values and no X column (since X values are ignored by the t test analysis). The two groups do not have to be the same size (its ok to leave some cells empty).

	A	B
	Control	Treated
	Y	Y
1	34.0	45.0
2	43.0	47.0
3	39.0	52.0

If you have already averaged your 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. Enter data on only one row.

Indexed data

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. However Prism can import indexed or stacked data. Read the chapter on importing data in the Prism User's Guide.

Group	Value
1	34
1	43
1	39
2	45
2	47
2	52

Advantages of transforming the data first

Before comparing two data sets, consider whether you should first transform the values. The t test depends on the assumption that your data are sampled from populations that follow 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:

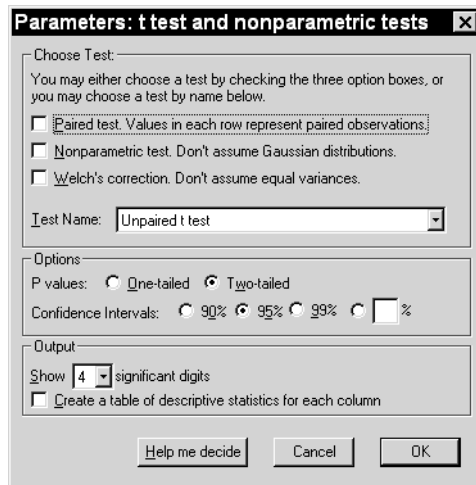
Type of data and distribution	Normalizing transform
Count (C comes from a Poisson distribution)	\sqrt{C}
Proportion (P comes from a binomial distribution)	$\text{Arcsine}\sqrt{P}$
Measurement (M comes from a lognormal distribution)	$\text{Log}(M)$
Time or duration (D)	$\frac{1}{D}$

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 "Paired or unpaired test" on page 41.

Choosing an analysis to compare two groups

Go to the data or results table you wish to analyze (see "Entering data to compare two groups with a t test (or a nonparametric test)" on page 39). Press the Analyze button, and choose built-in analyses. Then select t tests from the Statistical Analyses section. If your table has more than two columns, select the two columns you wish to compare. (If you want to compare three or more columns, see "One-way ANOVA and nonparametric comparisons" on page 65. Press ok from the Analyze dialog to bring up the parameters dialog for t tests and related nonparametric tests:



Paired or unpaired test?

When choosing a test, you need to decide whether to use a paired test. Choose a paired test when the two columns of data are matched. Here are some examples:

- You measure a variable (perhaps, weight) before an intervention, and then measure it in the same subjects after the 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 run a laboratory experiment several times, each time with a control and treated preparation handled in parallel.
- You measure a variable in twins, or child/parent pairs.

More generally, you should select a paired 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 paired 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.

t test or nonparametric test?

The t test, like many statistical tests, assumes that you have sampled data from populations that follow a Gaussian bell-shaped distribution. Biological data never follow a Gaussian distribution precisely, because a Gaussian distribution extends infinitely in both directions, and so it includes both infinitely low negative numbers and infinitely high positive numbers! But many kinds of biological data follow a bell-shaped distribution that is approximately Gaussian. Because ANOVA, t tests and other statistical tests work well even if the distribution is only approximately Gaussian (especially with large samples), these tests are used routinely in many fields of science.

An alternative approach does not assume that data follow a Gaussian distribution. In this approach, values are ranked from low to high and the analyses are based on the distribution of ranks. These tests, called *nonparametric* tests, are appealing because they make fewer assumptions about the distribution of the data. But there is a drawback. Nonparametric tests are less powerful than the parametric tests that assume Gaussian distributions. This means that P values tend to be higher, making it harder to detect real differences as being statistically significant. With large samples, the difference in power is minor. With small samples, nonparametric tests have little power to detect differences.

You may find it difficult to decide when to select nonparametric tests. You should definitely choose a nonparametric test in these situations:

- The outcome variable is a rank or score with only a few categories. Clearly the population is far from Gaussian in these cases.
- One, or 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. Using a nonparametric test with these data is easy. Assign an arbitrary low value to values that are too low to measure, and an arbitrary high value to values too high to measure. Since the nonparametric tests only consider the relative ranks of the values, it won't matter that you didn't know one (or a few) of the values exactly.

- You are sure that the population is far from Gaussian. Before choosing a nonparametric test, consider transforming the data (i.e. to logarithms, reciprocals). Sometimes a simple transformation will convert non-Gaussian data to a Gaussian distribution. See “Advantages of transforming the data first” on page 40.

In many situations, perhaps most, you will find it difficult to decide whether to select nonparametric tests. Remember that the Gaussian assumption is about the distribution of the overall population of values, not just the sample you have obtained in this particular experiment. Look at the scatter of data from previous experiments that measured the same variable. Also 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.

Prism performs normality testing in an attempt to determine whether data were sampled from a Gaussian distribution, but normality testing is less useful than you might hope (see “The results of normality” on page 29). Normality testing doesn’t help if you have fewer than a few dozen (or so) values.

Your decision to choose a parametric or nonparametric test matters the most when samples are small for reasons summarized here:

	Large samples (> 100 or so)	Small samples (<12 or so)
Parametric tests	Robust. P value will be nearly correct even if population is fairly far from Gaussian.	Not robust. If the population is not Gaussian, the P value may be misleading.
Nonparametric test	Powerful. If the population is Gaussian, the P value will be nearly identical to the P value you would have obtained from a parametric test. With large sample sizes, nonparametric tests are almost as powerful as parametric tests.	Not powerful. If the population is Gaussian, the P value will be higher than the P value obtained from a t test. With very small samples, it may be impossible for the P value to ever be less than 0.05, no matter how the values differ.

Normality test	Useful. Use a normality test to determine whether the data are sampled from a Gaussian population.	Not very useful. Little power to discriminate between Gaussian and non-Gaussian populations. Small samples simply don't contain enough information to let you make inferences about the shape of the distribution in the entire population.
----------------	--	---

Assume equal variances?

The unpaired t test assumes that the two populations have the same variances. Since the variance equals the standard deviation squared, this means that the populations have 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.

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 10.

You should only choose a one-tail P value when:

- You predicted which group would have the larger mean (if the means are in fact different) before collecting any data.
- You will attribute a difference in the wrong direction (the other group ends up with the larger mean), to chance, no matter how large the difference.

Since those conditions are rarely met, two-tail P values are usually more appropriate.

Confirm test selection

Based on your choices, Prism will show you the name of the test you selected.

Test	Paired	Nonparametric	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

Results

For help interpreting the results, see "The results of an unpaired t test" on page 45, "The results of a paired t test" on page 51, "The results of a Mann-Whitney test" on page 57, or "The results of a Wilcoxon matched pairs test" on page 59.

The results of an unpaired t test

How the unpaired t test works

To calculate a P value for an unpaired t test, Prism first computes a t ratio. The t ratio is the difference between sample means divided by the standard error of the difference, calculated by pooling the SEMs of the two groups. If the difference is large compared to the SE of the difference, then the t ratio will be large (or a large negative number), and the P value is small. The sign of the t ratio tells you only which group had the larger mean. The P value is derived from the absolute value of t.

For the standard t test, the number of degrees of freedom (df) equals the total sample size minus 2. Welch's t test (a rarely used test which doesn't assume equal variances) calculates df from a complicated equation. Prism calculates the P value from t and df.

A standard t test assumes the two groups have equal variances. To test this assumption, Prism calculates the variance of each group (the variance equals the standard deviation squared) and then calculates F, which equals the larger variance divided by the smaller variance. The degrees of freedom for the numerator and denominator equal the sample sizes minus 1. From F and the two df values, Prism computes a 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 and end up with F as large (or larger) as observed in your experiment. If the P value is small, conclude that the variances (and thus the standard deviations) are significantly different.

Don't base your conclusion just on this one F test. Also consider data from other experiments in the series. If you conclude that the two populations really do have different variances, you have three choices:

- Conclude that the two populations are different – that the treatment had an effect. In many 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. You may find that converting values to their reciprocals or logarithms will equalize the variances and also make the distributions more Gaussian. See "Advantages of transforming the data first" on page 40.
- Rerun the t test without assuming equal variances using Welch's modified t test.

How to think about results from an unpaired t test

The unpaired t test compares the means of two groups, assuming that data are sampled from Gaussian populations. The most important results are the P value and the confidence interval.

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?

"Statistically significant" is not the same as "scientifically important". Before interpreting the P value or confidence interval, you should think about the size of the difference you are looking for. How large a difference would you consider to be scientifically important? How small a difference would you consider to be scientifically trivial? Use scientific judgment and common sense to answer these questions. Statistical calculations cannot help, as the answers depend on the context of the experiment.

You will interpret the results differently depending on whether the P value is small or large.

If the P value is small

If the P value is small, then it is unlikely that the difference you observed is due to a coincidence of random sampling. You can reject the idea that the difference is a coincidence, and conclude instead that the populations have different means. The difference is statistically significant, but is it scientifically important? The confidence interval helps you decide.

Because of random variation, the difference between the group means in this experiment is unlikely to equal the true difference between population

means. There is no way to know what that true difference is. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true difference between the two means.

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a difference between means that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial difference	Trivial difference	Although the true difference is not zero (since the P value is low) the true difference between means is tiny and uninteresting. The treatment had an effect, but a small one.
Trivial difference	Important difference	Since the confidence interval ranges from a difference that you think would be biologically trivial to one you think would be important, you can't reach a strong conclusion. You can conclude that the means are different, but you don't know whether the size of that difference is scientifically trivial or important. You'll need more data to obtain a clear conclusion.
Important difference	Important difference	Since even the low end of the confidence interval represents a difference large enough to be considered biologically important, you can conclude that there is a difference between treatment means and that the difference is large enough to be scientifically relevant.

If the P value is large

If the P value is large, the data do not give you any reason to conclude that the overall means differ. Even if the true means were equal, you would not be surprised to find means this far apart just by coincidence. This is not the same as saying that the true means are the same. You just don't have convincing evidence that they differ.

How large could the true difference really be? Because of random variation, the difference between the group means in this experiment is

unlikely to be equal to the true difference between population means. There is no way to know what that true difference is. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true difference between the two means. When the P value is larger than 0.05, the 95% confidence interval will start with a negative number (representing a decrease) and go up to a positive number (representing an increase).

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a difference between means that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial decrease	Trivial increase	You can reach a crisp conclusion. Either the means really are the same or they differ by a trivial amount. At most, the true difference between means is tiny and uninteresting.
Trivial decrease	Large increase	You can't reach a strong conclusion. The data are consistent with the treatment causing a trivial decrease, no change, or an increase that might be large enough to be important. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Large decrease	Trivial increase	You can't reach a strong conclusion. The data are consistent with a trivial increase, no change, or a decrease that may be large enough to be important. You can't make a clear conclusion without repeating the experiment with more subjects.
Large decrease	Large increase	You can't conclude anything until you repeat the experiment with more subjects.

Checklist. Is an unpaired t test the right test for these data?

Before accepting the results of any statistical test, first think carefully about whether you chose an appropriate test. Before accepting results from an unpaired t test, ask yourself the questions below. Prism can help you answer the first two questions. You'll have to answer the others based on experimental design.

Question	Discussion
Are the populations distributed according to a Gaussian distribution?	The unpaired t test assumes that you have sampled your data from populations that follow a Gaussian distribution. While this assumption is not too important with large samples, it is important with small sample sizes (especially with unequal sample sizes). Prism tests for violations of this assumption, but normality tests have limited utility. See "The results of normality" on page 29. If your data do not come from Gaussian distributions, you have three options. Your best option is to transform the values to make the distributions more Gaussian (See "Advantages of transforming the data first" on page 40). Another choice is to use the Mann-Whitney nonparametric test instead of the t test. A final option is to use the t test anyway, knowing that the t test is fairly robust to violations of a Gaussian distribution with large samples.
Do the two populations have the same variances?	<p>The unpaired t test assumes that the two populations have the same variances (and thus the same standard deviation). Prism tests for equality of variance with an F test. The P value from this test answers this question: If the two populations really have the same variance, what is the chance that you'd randomly select samples whose ratio of variances is as far from 1.0 (or further) as observed in your experiment. A small P value suggests that the variances are different.</p> <p>Don't base your conclusion solely on the F test. Also think about data from other similar experiments. If you have plenty of previous data that convinces you that the variances are really equal, ignore the F test (unless the P value is really tiny) and interpret the t test results as usual.</p> <p>In some contexts, finding that populations have different variances may be as important as finding different means.</p>
Are the data unpaired?	The unpaired t test works by comparing the difference between means with the pooled standard deviations of the two groups. If the data are paired or matched, then you should choose a paired t test instead. If the pairing is effective in controlling for experimental variability, the paired t test will be more powerful than the unpaired test.

Are the "errors" independent?	The term "error" refers to the difference between each value and the group mean. The results of a t test only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six values in each group, but these were obtained from two animals in each group (in triplicate). In this case, some factor may cause all triplicates from one animal to be high or low. See "The need for independent samples" on page 5.
Are you comparing exactly two groups?	Use the t test only to compare two groups. To compare three or more groups, use one-way ANOVA followed by post tests. It is not appropriate to perform several t tests, comparing two groups at a time. Making multiple comparisons increases the chance of finding a statistically significant difference by chance and makes it difficult to interpret P values and statements of statistical significance.
Do both columns contain data?	If you want to compare a single set of experimental data with a theoretical value (perhaps 100%) don't fill a column with that theoretical value and perform an unpaired t test. Instead, use a one-sample t test. See "The results of a one-sample t test" on page 30.
Do you really want to compare means?	The unpaired t test compares the means of two groups. It is possible to have a tiny P value – clear evidence that the population means are different – even if the two distributions overlap considerably. In some situations – for example, assessing the usefulness of a diagnostic test – you may be more interested in the overlap of the distributions than in differences between means.
If you chose a one-tail P value, did you predict correctly?	If you chose a one-tail P value, you should 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 "One- vs. two-tail P values" on page 10.

The results of a paired t test

How a paired t test works

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.

First Prism 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 for a paired t test 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. The number of degrees of freedom equals the number of pairs minus 1. Prism calculates the P value from the t ratio and the number of degrees of freedom.

Test for adequate pairing

The whole point of using a paired experimental design and a paired test is to control for experimental variability. Some factors you don't control in the experiment will affect the before and the after measurements equally, so they will not affect the difference between before and after. By analyzing only the differences, therefore, a paired test corrects for those sources of scatter.

If pairing is effective, you expect the before and after measurements to vary together. Prism quantifies this by calculating the Pearson correlation coefficient, r . (See "Correlation coefficient" on page 137.) From r , Prism calculates 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) as observed in your experiment? The P value has one-tail, as you are not interested in the possibility of observing a strong negative correlation.

If the pairing was effective, r will be positive and the P value will be small. This means that the two groups are significantly correlated, so it made sense to choose a paired test.

If the P value is large (say larger than 0.05), you should question whether it made sense to use a paired test. Your choice of whether to use a paired test or not should not be based on this one P value, but also on the experimental design and the results you have seen in other similar experiments.

If r is negative, it means that the pairing was counterproductive! 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. Most likely this is just a matter of chance. If r is close to -1, you should review your experimental design, as this is a very unusual result.

How to think about results of a paired t test

The paired t test compares two paired groups so you can make inferences about the size of the average treatment effect (average difference between the paired measurements). The most important results are the P value and the confidence interval.

The P value answers this question: If the treatment really had no effect, what is the chance that random sampling would result in an average effect as far from zero (or more so) as observed in this experiment?

"Statistically significant" is not the same as "scientifically important". Before interpreting the P value or confidence interval, you should think about the size of the treatment effect you are looking for. How large a difference would you consider to be scientifically important? How small a difference would you consider to be scientifically trivial? Use scientific judgment and common sense to answer these questions. Statistical calculations cannot help, as the answers depend on the context of the experiment.

You will interpret the results differently depending on whether the P value is small or large.

If the P value is small

If the P value is small, then it is unlikely that the treatment effect you observed is due to a coincidence of random sampling. You can reject the idea that the treatment does nothing, and conclude instead that the treatment had an effect. The treatment effect is statistically significant. But is it scientifically significant? The confidence interval helps you decide.

Random scatter affects your data, so the true average treatment effect is probably not the same as the average of the differences observed in this experiment. There is no way to know what that true effect is. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true treatment effect (the true mean of the differences between paired values).

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a difference between means that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial difference	Trivial difference	Although the true effect is not zero (since the P value is low) it is tiny and uninteresting. The treatment had an effect, but a small one.
Trivial difference	Important difference	Since the confidence interval ranges from a difference that you think are biologically trivial to one you think would be important, you can't reach a strong conclusion from your data. You can conclude that the treatment had an effect, but you don't know whether it is scientifically trivial or important. You'll need more data to obtain a clear conclusion.
Important difference	Important difference	Since even the low end of the confidence interval represents a treatment effect large enough to be considered biologically important, you can conclude that the treatment had an effect large enough to be scientifically relevant.

If the P value is large

If the P value is large, the data do not give you any reason to conclude that the treatment had an effect. This is not the same as saying that the treatment had no effect. You just don't have evidence of an effect.

How large could the true treatment effect really be? The average difference between pairs in this experiment is unlikely to equal the true average difference between pairs (because of random variability). There is no way to know what that true difference is. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true treatment effect. When the P value is larger than 0.05, the 95% confidence interval will start with a negative number (representing a decrease) and go up to a positive number (representing an increase).

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a difference between means that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial decrease	Trivial increase	You can reach a crisp conclusion. Either the treatment has no effect or a tiny one.
Trivial decrease	Large increase	You can't reach a strong conclusion. The data are consistent with the treatment causing a trivial decrease, no change, or an increase that may be large enough to be important. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Large decrease	Trivial increase	You can't reach a strong conclusion. The data are consistent with a trivial increase, no change, or a decrease that may be large enough to be important. You can't make a clear conclusion without repeating the experiment with more subjects.
Large decrease	Large increase	You can't reach any conclusion.

Checklist. Is the paired t test the right test for these data?

Before accepting the results of any statistical test, first think carefully about whether you chose an appropriate test. Before accepting results from a paired t test, ask yourself these questions. Prism can help you answer the first two questions listed below. You'll have to answer the others based on experimental design.

Question	Discussion
Are the differences distributed according to a Gaussian distribution?	The paired t test assumes that you have sampled your pairs of values from a population of pairs where the difference between pairs follows a Gaussian distribution. While this assumption is not too important with large samples, it is important with small sample sizes. Prism tests for violations of this assumption, but normality tests have limited utility. If your data do not come from Gaussian distributions, you have two options. Your best option is to transform the values to make the distributions more Gaussian (see "Advantages of transforming the data first" on page 40). Another choice is to use the Wilcoxon matched pairs nonparametric test instead of the t test.

Was the pairing effective? The pairing should be part of the experimental design and not something you do after collecting data. Prism tests the effectiveness of pairing by calculating the Pearson correlation coefficient, r , and a corresponding P value. See "Correlation coefficient" on page 137. If r is positive and P is small, the two groups are significantly correlated. This justifies the use of a paired test.

If this P value is large (say larger than 0.05), you should question whether it made sense to use a paired test. Your choice of whether to use a paired test or not should not be based on this one P value, but also on the experimental design and the results you have seen in other similar experiments.

Are the pairs independent? The results of a paired t test only make sense when the pairs are independent – that whatever factor caused a difference (between paired values) to be too high or too low affects only that one pair. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six pairs of values, but these were obtained from three animals, with duplicate measurements in each animal. In this case, some factor may cause the after-before differences from one animal to be high or low. This factor would affect two of the pairs, so they are not independent. See "The need for independent samples" on page 5.

Are you comparing exactly two groups? Use the t test only to compare two groups. To compare three or more matched groups, use repeated measures one-way ANOVA followed by post tests. It is not appropriate to perform several t tests, comparing two groups at a time.

If you chose a one-tail P value, did you predict correctly? If you chose a one-tail P value, you should have predicted which group would have the larger mean before collecting data. Prism does not ask you to record this prediction, but assumes that it is correct. If your prediction was wrong, then ignore the reported P value and state that $P > 0.50$. See "One- vs. two-tail P values" on page 10.

Do you care about differences or ratios? The paired t test analyzes the differences between pairs. With some experiments, you may observe a very large variability among the differences. The differences are larger when the control value is larger. With these data, you'll get more consistent results if you look at the ratio (treated/control) rather than the difference (treated – control). See below.

Ratio t tests

The paired t test analyzes the differences between pairs. With some kinds of data, you may observe a very large variability among the differences. The differences are larger when the control value is larger. With these data, you'll get more consistent results if you look at the ratio (treated/control) rather than the difference (treated – control).

Analyzing ratios can lead to problems because ratios are intrinsically asymmetric – all decreases are expressed as ratios between zero and one; all increases are expressed as ratios greater than 1.0. Instead it makes more sense to look at the logarithm of ratios. Then no change is zero (the logarithm of 1.0), increases are positive and decreases are negative. Note that the logarithm of the ratio Treated/Control equals $\log(\text{treated})$ minus $\log(\text{control})$. This makes it easy to perform a ratio t test with Prism.

If you have paired data and think that it makes more sense to look at ratios rather than differences, follow these steps.

1. Transform both columns to logarithms.
2. Click analyze and perform a paired t test.
3. Interpret the P value: If there really were no differences between control and treated values, what is the chance of obtaining a ratio as far from 1.0 as was observed?
4. Prism also reports the confidence interval of the difference, which equals the confidence interval of the $\log(\text{ratio})$. Take the antilog of each end of the interval (with a calculator) to compute the 95% confidence interval of the ratio.

The results of a Mann-Whitney test

How the Mann-Whitney test works

The Mann-Whitney test, also called the rank sum test, is a nonparametric test that compares two unpaired groups. 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. 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.

The P value answers this question: If the populations really have the same median, what is the chance that random sampling would result in a sum of ranks as far apart (or more so) as observed in this experiment?

If your samples are small, and there are no ties, Prism calculates an exact P value. If your samples are large, or if there are ties, it approximates the P value from a Gaussian approximation. Here, 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. The approximation is quite accurate with large samples, and is standard (used by all statistics programs).

How to think about the results of a Mann-Whitney test

The Mann-Whitney test is a nonparametric test to compare two unpaired groups. The key result is a P value that 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) as observed in this experiment?

If the P value is small, you can reject the idea that the difference is a coincidence, and conclude instead that the populations have different medians.

If the P value is large, the data do not give you any reason to conclude that the overall medians differ. This is not the same as saying that the medians are the same. You just have no compelling evidence that they differ. If you have small samples, the Mann-Whitney test has little power. In fact, if the total sample size is seven or less, the Mann-Whitney test will always give a P value greater than 0.05 no matter how much the groups differ.

Checklist. Is the Mann-Whitney test the right test for these data?

Before interpreting the results of any statistical test, first think carefully about whether you have chosen an appropriate test. Before accepting results from a Mann-Whitney test, ask yourself these questions (Prism cannot help you answer them):

Question	Discussion
Are the "errors" independent?	The term "error" refers to the difference between each value and the group median. The results of a Mann-Whitney test only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six values in each group, but these were obtained from two animals in each group (in triplicate). In this case, some factor may cause all triplicates from one animal to be high or low. See "The need for independent samples" on page 5.
Are the data unpaired?	The Mann-Whitney test works by ranking all the values from low to high, and comparing the mean rank in the two groups. If the data are paired or matched, then you should choose a Wilcoxon matched pairs test instead.
Are you comparing exactly two groups?	Use the Mann-Whitney test only to compare two groups. To compare three or more groups, use the Kruskal-Wallis test followed by post tests. It is not appropriate to perform several Mann-Whitney (or t) tests, comparing two groups at a time.
Are the shapes of the two distributions identical?	The Mann-Whitney test does not assume that the populations follow Gaussian distributions. But it does assume that the shape of the two distributions is identical. The medians may differ – that is what you are testing for – but the test assumes that the shape of the two distributions is identical. If two groups have very different distributions, transforming the data may make the distributions more similar.
Do you really want to compare medians?	The Mann-Whitney test compares the medians of two groups. It is possible to have a tiny P value – clear evidence that the population medians are different – even if the two distributions overlap considerably.
If you chose a one-tail P value, did you predict correctly?	If you chose a one-tail P value, you should 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 "One- vs. two-tail P values" on page 10.

Are the data sampled from non-Gaussian populations?	By selecting a nonparametric test, you have avoided assuming that the data were sampled from Gaussian distributions. But there are drawbacks to using a nonparametric test. If the populations really are Gaussian, the nonparametric tests have less power (are less likely to give you a small P value), especially with small sample sizes. Furthermore, Prism (along with most other programs) does not calculate confidence intervals when calculating nonparametric tests. If the distribution is clearly not bell-shaped, consider transforming the values to create a Gaussian distribution and then using a t test.
---	--

The results of a Wilcoxon matched pairs test

How the Wilcoxon matched pairs test works

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?

In calculating the Wilcoxon test, Prism first computes the differences between each set of pairs, and ranks the absolute values of the differences from low to high. Prism then sums the ranks of the differences where column A was higher (positive ranks), sums the ranks where column B was higher (it calls these negative ranks), and reports the two sums. If the two sums of ranks are very different, the P value will be small. The P value answers this question: If the treatment really had no effect overall, what is the chance that random sampling would lead to a sum of ranks as far apart (or more so) as observed here?

If your samples are small and there are no tied ranks, Prism calculates an exact P value. If your samples are large or there are tied ranks, it calculates the P value from a Gaussian approximation. The term Gaussian, as used here, has to do with the distribution of sum of ranks, and does not imply that your data need to follow a Gaussian distribution.

Test for effective pairing

The whole point of using a paired test is to control for experimental variability. Some factors you don't control in the experiment will affect the before and the after measurements equally, so they will not affect the difference between before and after. By analyzing only the differences, therefore, a paired test corrects for these sources of scatter.

If pairing is effective, you expect the before and after measurements to vary together. Prism quantifies this by calculating the nonparametric Spearman correlation coefficient, r_s . From r_s , Prism calculates 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) as observed in your experiment. Here the P value is one-tail, as you are not interested in the possibility of observing a strong negative correlation.

If the pairing was effective, r_s will be positive and the P value will be small. This means that the two groups are significantly correlated, so it made sense to choose a paired test.

If the P value is large (say larger than 0.05), you should question whether it made sense to use a paired test. Your choice of whether to use a paired test or not should not be based on this one P value, but also on the experimental design and the results you have seen in other similar experiments (assuming you have repeated the experiments several times).

If r_s is negative, it means that 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. Most likely this is just a matter of chance. If r_s is close to -1, you should review your procedures, as the data are unusual.

How to think about the results of a Wilcoxon matched pairs test

The Wilcoxon matched pairs test is a nonparametric test to compare two paired groups. It is also called the Wilcoxon matched pairs signed ranks test.

The Wilcoxon test analyzes only the differences between the paired measurements for each subject. The P value answers this question: If the median difference really is zero overall, what is the chance that random sampling would result in a median difference as far from zero (or more so) as observed in this experiment?

If the P value is small, you can reject the idea that the difference is a coincidence, and conclude instead that the populations have different medians.

If the P value is large, the data do not give you any reason to conclude that the overall medians differ. This is not the same as saying that the means are the same. You just have no compelling evidence that they differ. If you have small samples, the Wilcoxon test has little power to detect small differences.

Checklist. Is the Wilcoxon test the right test for these data?

Before interpreting the results of any statistical test, first think carefully about whether you have chosen an appropriate test. Before accepting results from a Wilcoxon matched pairs test, ask yourself these questions:

Question	Discussion
Are the pairs independent?	The results of a Wilcoxon test only make sense when the pairs are independent – that whatever factor caused a difference (between paired values) to be too high or too low affects only that one pair. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six pairs of values, but these were obtained from three animals, with duplicate measurements in each animal. In this case, some factor may cause the after-before differences from one animal to be high or low. This factor would affect two of the pairs (but not the other four), so these two are not independent. See "The need for independent samples" on page 5.
Is the pairing effective?	<p>The whole point of using a paired test is to control for experimental variability, and thereby increase power. Some factors you don't control in the experiment will affect the before and the after measurements equally, so will not affect the difference between before and after. By analyzing only the differences, therefore, a paired test controls for some of the sources of scatter.</p> <p>The pairing should be part of the experimental design and not something you do after collecting data. Prism tests the effectiveness of pairing by calculating the Spearman correlation coefficient, r_s, and a corresponding P value. See "Results of correlation" on page 137. If r_s is positive and P is small, the two groups are significantly correlated. This justifies the use of a paired test.</p> <p>If the P value is large (say larger than 0.05), you should question whether it made sense to use a paired test. Your choice of whether to use a paired test or not should not be based solely on this one P value, but also on the experimental design and the results you have seen in other similar experiments.</p>

Are you comparing exactly two groups?	Use the Wilcoxon test only to compare two groups. To compare three or more matched groups, use the Friedman test followed by post tests. It is not appropriate to perform several Wilcoxon tests, comparing two groups at a time.
If you chose a one-tail P value, did you predict correctly?	If you chose a one-tail P value, you should 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 "One-vs. two-tail P values" on page 10.
Are the data clearly sampled from non-Gaussian populations?	By selecting a nonparametric test, you have avoided assuming that the data were sampled from Gaussian distributions. But there are drawbacks to using a nonparametric test. If the populations really are Gaussian, the nonparametric tests have less power (are less likely to give you a small P value), especially with small sample sizes. Furthermore, Prism (along with most other programs) does not calculate confidence intervals when calculating nonparametric tests. If the distribution is clearly not bell-shaped, consider transforming the values (perhaps to logs or reciprocals) to create a Gaussian distribution and then using a t test. See "Advantages of transforming the data first" on page 40.
Are the differences distributed symmetrically?	The Wilcoxon test first computes the difference between the two values in each row, and analyzes only the list of differences. The Wilcoxon test does not assume that those differences are sampled from a Gaussian distribution. However it does assume that the differences are distributed symmetrically around their median.

One-way ANOVA and nonparametric comparisons

Introduction to comparisons of three or more groups

Prism can compare three or more groups with ordinary or repeated measures ANOVA, or with the nonparametric Kruskal-Wallis or Friedman tests. Following ANOVA, Prism can perform the Bonferroni, Tukey, Newman-Keuls, or Dunnett's post test. Following nonparametric ANOVA, Prism can calculate the Dunn's post test.

These tests compare measurements (continuous variables) such as weight, enzyme activity, and receptor number. To compare proportions see "Contingency tables" on page 121. To compare survival curves, see "Survival curves" on page 109.

One-way ANOVA (and related nonparametric tests) compare three or more groups when the data are categorized in one way. For example, you might compare a control group with two treated groups. If your data are categorized in two ways (for example you want to compare control with two treated groups in both males and females) see "Two-way analysis of variance" on page 93.

Entering data for ANOVA (and nonparametric tests)

Enter the data into three or more Y columns (data sets), formatted as a single column of values. ANOVA calculations ignore the X column, so you'll usually want to format the table with no X column. However, it is ok to format the table with a text or numerical X column.

The groups do not have to be the same size (it's ok to leave some cells empty).

If you have already averaged your 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. Enter data on only the first row.

If you format the table for replicate Y values, Prism averages the replicates and bases the one-way ANOVA analysis only on the means.

Indexed data

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. However Prism can import indexed or stacked data (see the chapter on importing in the Prism User's Guide).

Group	Value
1	34
1	43
1	39
2	45
2	47
2	52
3	76
3	99
3	82

Advantages of transforming the data first

Before comparing columns, consider whether you should first transform the values. ANOVA depends on the assumption that your data are sampled from populations that follow 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:

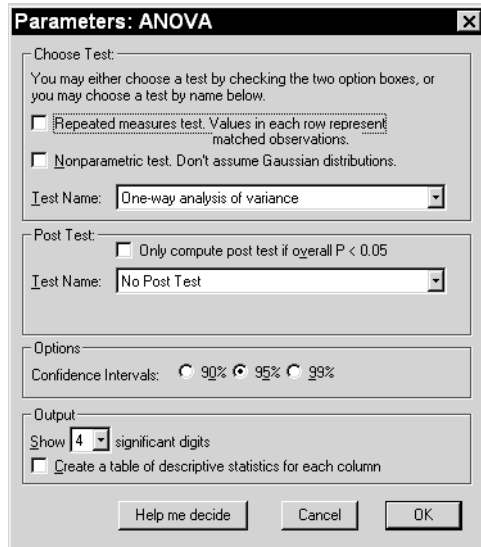
Type of data and distribution	Normalizing transform
Count (C comes from a Poisson distribution)	\sqrt{C}
Proportion (P comes from a binomial distribution)	$\text{Arcsine}\sqrt{P}$
Measurement (M comes from a lognormal distribution)	$\text{Log}(M)$
Time or duration (D)	$\frac{1}{D}$

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 "ANOVA or nonparametric test?" on page 68.

Choosing one-way ANOVA and related analyses

Start from the data or results table you wish to analyze (see "Entering data for ANOVA (and nonparametric tests)" on page 65. Press the Analyze button and choose built-in analyses. Then select One-way ANOVA from the list of statistical analyses. If you don't wish to analyze all columns in the table, select the columns you wish to compare. Press ok to bring up the Parameters dialog.



Repeated measures test?

You should choose repeated measures test when the experiment uses matched subjects. Here are some examples:

- You measure a variable in each subject before, during and after an intervention.
- 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 several treated preparations handled in parallel.
- You measure a variable in triplets, or grandparent/parent/child groups.

More generally, you should select a repeated measures test whenever you expect a value in one group to be closer to a *particular* value in the other groups than to a *randomly selected* value in the another 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.

The term *repeated measures* applies strictly when you give treatments repeatedly to one subject. The other examples are called *randomized block* experiments (each set of subjects is called a *block* and you randomly assign treatments within each block). The analyses are identical for repeated measures and randomized block experiments, and Prism always uses the term repeated measures.

ANOVA or nonparametric test?

ANOVA, as well as other statistical tests, assumes that you have sampled data from populations that follow a Gaussian bell-shaped distribution. Biological data never follow a Gaussian distribution precisely, because a Gaussian distribution extends infinitely in both directions, so it includes both infinitely low negative numbers and infinitely high positive numbers! But many kinds of biological data follow a bell-shaped distribution that is approximately Gaussian. Because ANOVA works well even if the distribution is only approximately Gaussian (especially with large samples), these tests are used routinely in many fields of science.

An alternative approach does not assume that data follow a Gaussian distribution. In this approach, values are ranked from low to high and the analyses are based on the distribution of ranks. These tests, called *nonparametric* tests, are appealing because they make fewer assumptions about the distribution of the data. But there is a drawback. Nonparametric tests are less powerful than the parametric tests that assume Gaussian distributions. This means that P values tend to be higher, making it harder to detect real differences as being statistically significant. If the samples are large the difference in power is minor. With small samples, nonparametric tests have little power to detect differences. With very small groups, nonparametric tests have zero power – the P value will always be greater than 0.05.

You may find it difficult to decide when to select nonparametric tests. You should definitely choose a nonparametric test in these situations:

- The outcome variable is a rank or score with only a few categories. Clearly the population is far from Gaussian in these cases.

- One, or 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 an arbitrary low value to values too low to measure, and an arbitrary high value to values too high to measure. Since the nonparametric tests only consider the relative ranks of the values, it won't matter that you didn't know one (or a few) of the values exactly.
- You are sure that the population is far from Gaussian. Before choosing a nonparametric test, consider transforming the data (perhaps to logarithms or reciprocals). Sometimes a simple transformation will convert non-Gaussian data to a Gaussian distribution.

In many situations, perhaps most, you will find it difficult to decide whether to select nonparametric tests. Remember that the Gaussian assumption is about the distribution of the overall population of values, not just the sample you have obtained in this particular experiment. Look at the scatter of data from previous experiments that measured the same variable. Also 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.

Prism performs normality testing in an attempt to determine whether data were sampled from a Gaussian distribution, but normality testing is less useful than you might hope (see "The results of normality tests" on page 29). Normality testing doesn't help if you have fewer than a few dozen (or so) values.

Your decision to choose a parametric or nonparametric test matters the most when samples are small for reasons summarized here:

	Large samples (> 100 or so)	Small samples (<12 or so)
Parametric tests	Robust. P value will be nearly correct even if population is fairly far from Gaussian.	Not robust. If the population is not Gaussian, the P value may be misleading.
Nonparametric test	Powerful. If the population is Gaussian, the P value will be nearly identical to the P value you would have obtained from a parametric test. With large sample sizes, nonparametric tests are almost as powerful as parametric tests.	Not powerful. If the population is Gaussian, the P value will be higher than the P value obtained from ANOVA. With very small samples, it may be impossible for the P value to ever be less than 0.05, no matter how the values differ.
Normality test	Useful. Use a normality test to determine whether the data are sampled from a Gaussian population.	Not very useful. Little power to discriminate between Gaussian and non-Gaussian populations. Small samples simply don't contain enough information to let you make inferences about the shape of the distribution in the entire population.

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:

No Post Test
Tukey: Compare all pairs of columns.
Newman-Keuls: Compare all pairs of columns
Bonferroni: Compare all pairs of columns.
Bonferroni: Compare selected pairs of columns.
Dunnnett: Compare all columns vs. control column
Test for linear trend between mean & col. number

Choosing an appropriate post test is not straightforward, and different statistics texts make different recommendations.

Select **Dunnnett s** test if one column represents control data, and you wish to compare all other columns to that control column but not to each other.

Select the **test for linear trend**, if the columns are arranged in a natural order (i.e. dose or time) and you want to test whether there is a trend so that values increase (or decrease) as you move from left to right across columns.

Select the **Bonferroni test for selected pairs of columns** when you only wish to compare certain column pairs. You must select those pairs based on experimental design, and ideally should specify the pairs of interest before collecting any data. If you base your decision on the results (i.e. compare the smallest with the largest mean), then you have effectively compared all columns, and it is not appropriate to use the test for selected pairs.

Select the Bonferroni, Tukey, or Newman-Keuls test (also known as the Student-Newman, Keuls or SNK test) if you want to compare all pairs of columns.

The only advantage of the **Bonferroni** method is that it is 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 many 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 among statisticians. 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 Newman-Keuls P values test. For that reason, and because the Newman-Keuls test does not generate confidence intervals, we suggest selecting Tukey's test. (If you select the Tukey test you are actually selecting the Tukey-Kramer test, which includes the extension by Kramer to allow for unequal sample sizes.)

Confirm test selection

Based on the option boxes you selected, Prism will choose a test for you:

Test	Matched	Nonparametric
Ordinary one-way ANOVA	No	No
Repeated measures one-way ANOVA	Yes	No
Kruskal-Wallis test	No	Yes
Friedman test	Yes	Yes

Results

For help interpreting the results, see "The results of one-way ANOVA" on page 72, "The results of a Wilcoxon rank sum test" on page 34, "The results of repeated measures one-way ANOVA" on page 82, or "The results of a Friedman test" on page 89.

The results of one-way ANOVA

How one-way ANOVA works

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) as observed in this experiment?

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 among group means and variability within the groups. Variability within groups is quantified as the sum of 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 (the mean of all values in all groups). 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 text for details. 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. The P value answers this question: If the populations all have the same mean, what is the chance that randomly selected groups would lead to an F ratio as big (or bigger) as the one obtained in your experiment?

R² value

R² is the fraction of the overall variance (of all the data, pooling all the groups) attributable to differences among the group means. It compares the variability among 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 R² value is calculated from the ANOVA table and equals the between group sum-of-squares divided by the total sum-of-squares. Some programs (and books) don't bother reporting

this value. Others refer to it as η^2 (eta squared) rather than R^2 . It is a descriptive statistic that quantifies the strength of the relationship between group membership and the variable you measured.

Bartlett's test for equal variances

ANOVA is based on the assumption that the populations all have the same variance. If your samples have four or more values, Prism tests this assumption with Bartlett's test. It reports the value of Bartlett's statistic with a P value that answers this question: If the populations really have the same variance, what is the chance that you'd randomly select samples whose variances are as different (or more different) as observed in your experiment. (Since the variance is the standard deviation squared, testing for equal variances is the same as testing for equal standard deviations).

Bartlett's test is very sensitive to deviations from a Gaussian distribution – more sensitive than the ANOVA calculations are. A low P value from Bartlett's test may be due to data that are not Gaussian, rather than due to unequal variances. Since ANOVA is fairly robust to non-Gaussian data (at least when sample sizes are equal), the Bartlett's test can be misleading. Some statisticians suggest ignoring the Bartlett's test, especially when the sample sizes are equal (or nearly so).

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. Think about data from other similar experiments before making a conclusion.

If you conclude that the populations have different variances, you have three choices:

- Conclude that the populations are different – the treatments had an effect. In many 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 among the means. This may be the most important conclusion from the experiment.
- Transform the data to equalize the variances, then rerun the ANOVA. Often you'll find that converting values to their reciprocals or logarithms will equalize the variances and make the distributions more Gaussian.
- Use a modified ANOVA that does not assume equal variances. Prism does not provide such a test.

How post tests work

Post test for a linear trend

If the columns represent ordered and equally spaced (or nearly so) groups, the post test for a linear trend determines whether the column means increase (or decrease) systematically as the columns go from left to right.

The post test for a linear trend works by calculating linear regression on column number vs. group mean. Prism reports the slope and r^2 , as well as the P value for the linear trend. This P value answers this question: If there really is no linear trend between column number and column mean, what is the chance that random sampling would result in a slope as far from zero (or further) than you obtained here? Equivalently, it is the chance of observing a value of r^2 that high or higher, just by coincidence of random sampling.

Prism also reports a second P value testing for nonlinear variation. After correcting for the linear trend, this P value tests whether the remaining variability among column means is greater than expected by chance. It is the chance of seeing that much variability due to random sampling.

Finally, Prism shows an ANOVA table which partitions total variability into three components: linear variation, nonlinear variation, and random or residual variation. It is used to compute the two F ratios, which lead to the two P values. The ANOVA table is included to be complete, but will not be of use to most scientists

For more information about the post test for a linear trend, see the excellent text, [Practical Statistics for Medical Research](#) by DG Altman, published in 1991 by Chapman and Hall.

Other post tests

The Bonferroni, Tukey, Newman-Keuls and Dunnett's post tests are all modifications of t tests. They account for multiple comparisons, as well as for the fact that the comparisons are interrelated.

Recall that n unpaired t test computes the t ratio as the difference between two group means divided by the standard error of the difference (computed from the standard errors of the two group means, and the two sample sizes). The P value is then derived from t. The post tests work in a similar way. Instead of dividing by the standard error of the difference, they divide by a value computed from the residual mean square (shown on the ANOVA table). Each test uses a different method to derive a P value from this ratio,

For the difference between each pair of means, Prism reports the P value as >0.05 , <0.05 , <0.01 or <0.001 . These P values account for multiple comparisons. Prism reports a P value for the difference between each pair of means, but the probability values apply to the entire family of

comparisons, not for each individual comparison. If the null hypothesis is true (all the values are sampled from populations with the same mean), then there is only a 5% chance that any one or more comparisons will have a P value less than 0.05.

Prism also reports a 95% confidence interval for the difference between each pair of means (except for the Newman-Keuls post test, which cannot be used for confidence intervals). These intervals account for multiple comparisons. There is a 95% chance that all of these intervals contain the true differences between population, and only a 5% chance that any one or more of these intervals misses the true difference. A 95% confidence interval is computed for the difference between each pair of means, but the 95% probability applies to the entire family of comparisons, not to each individual comparison.

How to think about results from one-way ANOVA

One-way ANOVA compares the means of three or more groups, assuming that data are sampled from Gaussian populations. The most important results are the P value and the post tests.

The overall 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 from one another (or more so) as you observed in this experiment?

If the overall P value is large, the data do not give you any reason to conclude that the means differ. Even if the true means were equal, you would not be surprised to find means this far apart just by coincidence. This is not the same as saying that the true means are the same. You just don't have compelling evidence that they differ.

If the overall P value is small, then it is unlikely that the differences you observed are due to a coincidence of random sampling. You can reject the idea that all the populations have identical means. This doesn't mean that every mean differs from every other mean, only that at least one differs from the rest. Look at the results of post tests to identify where the differences are.

How to think about the results of post tests

If the columns are organized in a natural order, the post test for linear trend tells you whether the column means have a systematic trend, increasing (or decreasing) as you go from left to right in the data table. See "Post test for a linear trend" on page 74.

With other post tests, look at which differences between column means are statistically significant. For each pair of means, Prism reports whether the P value is less than 0.05, 0.01 or 0.001.

"Statistically significant" is not the same as "scientifically important". Before interpreting the P value or confidence interval, you should think about the size of the difference you are looking for. How large a difference would you consider to be scientifically important? How small a difference would you consider to be scientifically trivial? Use scientific judgment and common sense to answer these questions. Statistical calculations cannot help, as the answers depend on the context of the experiment.

As discussed below, you will interpret the post test results differently depending on whether the difference is statistically significant or not.

If the difference is statistically significant

If the P value for a post test is small, then it is unlikely that the difference you observed is due to a coincidence of random sampling. You can reject the idea that those two populations have identical means.

Because of random variation, the difference between the group means in this experiment is unlikely to equal the true difference between population means. There is no way to know what that true difference is. With most post tests (but not the Newman-Keuls test), Prism presents the uncertainty as a 95% confidence interval for the difference between all (or selected) pairs of means. You can be 95% sure that this interval contains the true difference between the two means.

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a difference between means that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial difference	Trivial difference	Although the true difference is not zero (since the P value is low) the true difference between means is tiny and uninteresting. The treatment had an effect, but a small one.
Trivial difference	Important difference	Since the confidence interval ranges from a difference that you think is biologically trivial to one you think would be important, you can't reach a strong conclusion from your data. You can conclude that the means are different, but you don't know whether the size of that difference is scientifically trivial or important. You'll need more data to reach a clear conclusion.
Important difference	Important difference	Since even the low end of the confidence interval represents a difference large enough to be considered biologically important, you can conclude that there is a difference between treatment means and that the difference is large enough to be scientifically relevant.

If the difference is not statistically significant

If the P value from a post test is large, the data do not give you any reason to conclude that the means of these two groups differ. Even if the true means were equal, you would not be surprised to find means this far apart just by coincidence. This is not the same as saying that the true means are the same. You just don't have compelling evidence that they differ.

How large could the true difference really be? Because of random variation, the difference between the group means in this experiment is unlikely to equal the true difference between population means. There is no way to know what that true difference is. Prism presents the uncertainty as a 95% confidence interval (except with the Newman-Keuls test). You can be 95% sure that this interval contains the true difference between the two means. When the P value is larger than 0.05, the 95% confidence interval will start with a negative number (representing a decrease) and go up to a positive number (representing an increase).

To interpret the results in a scientific context, look at both ends of the confidence interval for each pair of means, and ask whether those differences would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial decrease	Trivial increase	You can reach a crisp conclusion. Either the means really are the same or they are different by a trivial amount. At most, the true difference between means is tiny and uninteresting.
Trivial decrease	Large increase	You can't reach a strong conclusion. The data are consistent with the treatment causing a trivial decrease, no change, or an increase that might be large enough to be important. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Large decrease	Trivial increase	You can't reach a strong conclusion. The data are consistent with a trivial increase, no change, or a decrease that may be large enough to be important. You can't make a clear conclusion without repeating the experiment with more subjects.
Large decrease	Large increase	You can't reach any conclusion. Repeat the experiment with a much larger sample size.

Checklist. Is one-way ANOVA the right test for these data?

Before accepting the results of any statistical test, first think carefully about whether you chose an appropriate test. Before accepting results from a one-way ANOVA, ask yourself the questions below. Prism can help answer the first two questions. You'll need to answer the others based on experimental design.

Question	Discussion
Are the populations distributed according to a Gaussian distribution?	One-way ANOVA assumes that you have sampled your data from populations that follow a Gaussian distribution. While this assumption is not too important with large samples, it is important with small sample sizes (especially with unequal sample sizes). Prism can test for violations of this assumption, but normality tests have limited utility. See "The results of normality" on page 29. If your data do not come from Gaussian distributions, you have three options. Your best option is to transform the values (perhaps to logs or reciprocals) to make the distributions more Gaussian (see "Advantages of transforming the data first" on page 66. Another choice is to use the Kruskal-Wallis nonparametric test instead of ANOVA. A final option is to use ANOVA anyway, knowing that it is fairly robust to violations of a Gaussian distribution with large samples.
Do the populations have the same standard deviation?	<p>One-way ANOVA assumes that all the populations have the same standard deviation (and thus the same variance). This assumption is not very important when all the groups have the same (or almost the same) number of subjects, but is very important when sample sizes differ.</p> <p>Prism tests for equality of variance with Bartlett's test. The P value from this test answers this question: If the populations really have the same variance, what is the chance that you'd randomly select samples whose variances are as different as those observed in your experiment. A small P value suggests that the variances are different.</p> <p>Don't base your conclusion solely on Bartlett's test. Also think about data from other similar experiments. If you have plenty of previous data that convinces you that the variances are really equal, ignore Bartlett's test (unless the P value is really tiny) and interpret the ANOVA results as usual. Some statisticians recommend ignoring Bartlett's test altogether if the sample sizes are equal (or nearly so).</p> <p>In some experimental contexts, finding different variances may be as important as finding different means. If the variances are different, then the populations are different – regardless of what ANOVA concludes about differences between the means.</p> <p>See "Bartlett's test for equal variances" on page 73.</p>

Are the data unmatched?	One-way ANOVA works by comparing the differences among group means with the pooled standard deviations of the groups. If the data are matched, then you should choose repeated measures ANOVA instead. If the matching is effective in controlling for experimental variability, repeated measures ANOVA will be more powerful than regular ANOVA.
Are the "errors" independent?	The term "error" refers to the difference between each value and the group mean. The results of one-way ANOVA only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six values in each group, but these were obtained from two animals in each group (in triplicate). In this case, some factor may cause all triplicates from one animal to be high or low. See "The need for independent samples" on page 5.
Do you really want to compare means?	One-way ANOVA compares the means of three or more groups. It is possible to have a tiny P value – clear evidence that the population means are different – even if the distributions overlap considerably. In some situations – for example, assessing the usefulness of a diagnostic test – you may be more interested in the overlap of the distributions than in differences between means.
Is there only one factor?	<p>One-way ANOVA compares three or more groups defined by one factor. For example, you might compare a control group, with a drug treatment group and a group treated with drug plus antagonist. Or you might compare a control group with five different drug treatments.</p> <p>Some experiments involve more than one factor. For example, you might compare three different drugs in men and women. There are two factors in that experiment: drug treatment and gender. These data need to be analyzed by two-way ANOVA, also called two factor ANOVA. See "Two-way analysis of variance" on page 93.</p>

Is the factor “fixed” rather than “random”?

Prism performs Type I ANOVA, also known as fixed-effect ANOVA. This tests for differences among the means of the particular groups you have collected data from. Type II ANOVA, also known as random-effect ANOVA, assumes that you have randomly selected groups from an infinite (or at least large) number of possible groups, and that you want to reach conclusions about differences among ALL the groups, even the ones you didn't include in this experiment. Type II random-effects ANOVA is rarely used, and Prism does not perform it. If you need to perform ANOVA with random effects variables, consider using the program NCSS from www.ncss.com.

The results of repeated measures one-way ANOVA

How repeated measures ANOVA works

Repeated measures one-way ANOVA compares three or more matched groups, based on the assumption that the differences between matched values are Gaussian. For example, one-way ANOVA may compare measurements made before, during and after an intervention, when each subject was assessed three times. 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) as observed in this experiment?

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 (residual). 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. The main F ratio tests the null hypothesis that the column means are identical. The other F ratio tests the null hypothesis that the row means are identical (this is the test for effective matching). In each case, the F ratio is expected to be near 1.0 if the null hypothesis is true. If F is large, the P value will be small.

The circularity assumption

Repeated measures ANOVA assumes that the random error truly is random. A random factor that causes a measurement in one subject to be a bit high (or low) should have no affect on the next measurement in the same

subject. This assumption is called *circularity* or *sphericity*. It is closely related to another term you may encounter, *compound symmetry*.

Repeated measures ANOVA is quite sensitive to violations of the assumption of circularity. If the assumption is violated, the P value will be too low. You'll violate this assumption when the repeated measurements are made too close together so that random factors that cause a particular value to be high (or low) don't wash away or dissipate before the next measurement. To avoid violating the assumption, wait long enough between treatments so the subject is essentially the same as before the treatment. When possible, also randomize the order of treatments.

You only have to worry about the assumption of circularity when you perform a repeated measures experiment, where each row of data represents repeated measurements from a single subject. It is impossible to violate the assumption with randomized block experiments, where each row of data represents data from a matched set of subjects. See "Repeated measures test?" on page 67

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 an 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 powerful 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.

How to think about results from repeated measures one-way ANOVA

Repeated measures ANOVA compares the means of three or more matched groups. The term *repeated measures* strictly applies only when you give treatments repeatedly to each subject, and the term *randomized block* is used when you randomly assign treatments within each group (block) of matched subjects. The analyses are identical for repeated measures and randomized block experiments, and Prism always uses the term repeated measures.

Your approach to interpreting repeated measures ANOVA results will be the same as interpreting the results of ordinary one-way ANOVA. See "How to think about results from one-way ANOVA" on page 75.

Checklist. Is repeated measures one way ANOVA the right test for these data?

Before accepting the results of any statistical test, first think carefully about whether you chose an appropriate test. Before accepting results from repeated measures one-way ANOVA, ask yourself the questions listed below. Prism can help you answer the first question. You must answer the remaining questions based on experimental design.

Question	Discussion
Was the matching effective?	<p>The whole point of using a repeated measures test is to control for experimental variability. Some factors you don't control in the experiment will affect all the measurements from one subject equally, so will not affect the difference between the measurements in that subject. By analyzing only the differences, therefore, a matched test controls for some of the sources of scatter.</p> <p>The matching should be part of the experimental design and not something you do after collecting data. Prism tests the effectiveness of matching with an F test (distinct from the main F test of differences between columns). If the P value for matching is large (say larger than 0.05), you should question whether it made sense to use a repeated measures test. Ideally, your choice of whether to use a repeated measures test should be based not only on this one P value, but also on the experimental design and the results you have seen in other similar experiments.</p>
Are the subjects independent?	<p>The results of repeated measures ANOVA only make sense when the subjects are independent. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six rows of data, but these were obtained from three animals, with duplicate measurements in each animal. In this case, some factor may affect the measurements from one animal. Since this factor would affect data in two (but not all) rows, the rows (subjects) are not independent. See "The need for independent samples" on page 5.</p>

Is the random variability distributed according to a Gaussian distribution?	<p>Repeated measures ANOVA assumes that each measurement is the sum of an overall mean, a treatment effect (the average difference between subjects given a particular treatment and the overall mean), an individual effect (the average difference between measurements made in a certain subject and the overall mean) and a random component. Furthermore, it assumes that the random component follows a Gaussian distribution and that the standard deviation does not vary between individuals (rows) or treatments (columns). While this assumption is not too important with large samples, it can be important with small sample sizes. Prism does not test for violations of this assumption.</p>
---	--

Is there only one factor?	<p>One-way ANOVA compares three or more groups defined by one factor. For example, you might compare a control group, with a drug treatment group and a group treated with drug plus antagonist. Or you might compare a control group with five different drug treatments.</p> <p>Some experiments involve more than one factor. For example, you might compare three different drugs in men and women. There are two factors in that experiment: drug treatment and gender. Similarly, there are two factors if you wish to compare the effect of drug treatment at several time points. These data need to be analyzed by two-way ANOVA, also called two factor ANOVA.</p>
---------------------------	--

Is the factor "fixed" rather than "random"?	<p>Prism performs Type I ANOVA, also known as fixed-effect ANOVA. This tests for differences among the means of the particular groups you have collected data from. Type II ANOVA, also known as random-effect ANOVA, assumes that you have randomly selected groups from an infinite (or at least large) number of possible groups, and that you want to reach conclusions about differences among ALL the groups, even the ones you didn't include in this experiment. Type II random-effects ANOVA is rarely used, and Prism does not perform it.</p>
---	--

The results of a Kruskal-Wallis test

How the Kruskal-Wallis test works

The Kruskal-Wallis test is a nonparametric test that compares three or more unpaired groups. To perform the Kruskal-Wallis test, Prism first ranks all the values from low to high, disregarding which group each value belongs. 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 all the groups. Prism then sums the ranks in each group, and reports the sums. If the sums of the ranks are very different, the P value will be small.

The discrepancies among the rank sums are combined to create a single value called the Kruskal-Wallis statistic (some books refer to this value as H). A larger Kruskal-Wallis statistic corresponds to a larger discrepancy among rank sums.

The P value answers this question: If the populations really have the same median, what is the chance that random sampling would result in sums of ranks as far apart (or more so) as observed in this experiment? More precisely, if the null hypothesis is true then what is the chance of obtaining a Kruskal-Wallis statistic as high (or higher) as observed in this experiment.

If your samples are small and no two values are identical (no ties), Prism calculates an exact P value. If your samples are large or if there are ties, it approximates the P value from the chi-square distribution. The approximation is quite accurate with large samples. With medium size samples, Prism can take a long time to calculate the exact P value. While it does the calculations, Prism displays a progress dialog and you can press Cancel to interrupt the calculations if an approximate P value is good enough for your purposes.

How Dunn's post test works

Dunn's post test compares the difference in the sum of ranks between two columns with the expected average difference (based on the number of groups and their size).

For each pair of columns, Prism reports the P value as >0.05 , <0.05 , <0.01 or <0.001 . The calculation of the P value takes into account the number of comparisons you are making. If the null hypothesis is true (all data are sampled from populations with identical distributions, so all differences between groups are due to random sampling), then there is a 5% chance that at least one of the post tests will have $P < 0.05$. The 5% chance does not apply to each comparison but rather to the entire family of comparisons.

For more information on the post test, see [Applied Nonparametric Statistics](#) by WW Daniel, published by PWS-Kent publishing company in 1990 or [Nonparametric Statistics for Behavioral Sciences](#) by S Siegel and NJ Castellan, 1988. The original reference is O.J. Dunn, *Technometrics*, 5:241-252, 1964.

Prism refers to the post test as the Dunn's post test. Some books and programs simply refer to this test as the post test following a Kruskal-Wallis test, and don't give it an exact name.

How to think about a Kruskal-Wallis test

The Kruskal-Wallis test is a nonparametric test to compare three or more unpaired groups. It is also called Kruskal-Wallis one-way analysis of variance by ranks. The key result is a P value that 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) as you observed in this experiment?

If the P value is small, you can reject the idea that the differences are all a coincidence. This doesn't mean that every group differs from every other group, only that at least one group differs from one of the others. Look at the post test results to see which groups differ from which other groups.

If the overall Kruskal-Wallis P value is large, the data do not give you any reason to conclude that the overall medians differ. This is not the same as saying that the medians are the same. You just have no compelling evidence that they differ. If you have small samples, the Kruskal-Wallis test has little power. In fact, if the total sample size is seven or less, the Kruskal-Wallis test will always give a P value greater than 0.05 no matter how the groups differ.

How to think about post tests following the Kruskal-Wallis test

Dunn's post test calculates a P value for each pair of columns. These P values answer this question: If the data were sampled from populations with the same median, what is the chance that one or more pairs of columns would have medians as far apart as observed here? If the P value is low, you'll conclude that the difference is statistically significant. The calculation of the P value takes into account the number of comparisons you are making. If the null hypothesis is true (all data are sampled from populations with identical distributions, so all differences between groups are due to random sampling), then there is a 5% chance that at least one of the post tests will have $P < 0.05$. The 5% chance does not apply separately to each individual comparison but rather to the entire family of comparisons.

Checklist. Is the Kruskal-Wallis test the right test for these data?

Before interpreting the results of any statistical test, first think carefully about whether you have chosen an appropriate test. Before accepting

results from a Kruskal-Wallis test, ask yourself these questions about your experimental design:

Question	Discussion
Are the “errors” independent?	The term “error” refers to the difference between each value and the group median. The results of a Kruskal-Wallis test only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have nine values in each of three groups, but these were obtained from two animals in each group (in triplicate). In this case, some factor may cause all three values from one animal to be high or low. See “The need for independent samples” on page 5.
Are the data unpaired?	If the data are paired or matched, then you should consider choosing the Friedman test instead. If the pairing is effective in controlling for experimental variability, the Friedman test will be more powerful than the Kruskal-Wallis test.
Are the data sampled from non-Gaussian populations?	By selecting a nonparametric test, you have avoided assuming that the data were sampled from Gaussian distributions. But there are drawbacks to using a nonparametric test. If the populations really are Gaussian, the nonparametric tests have less power (are less likely to detect a true difference), especially with small sample sizes. Furthermore, Prism (along with most other programs) does not calculate confidence intervals when calculating nonparametric tests. If the distribution is clearly not bell-shaped, consider transforming the values (perhaps to logs or reciprocals) to create a Gaussian distribution and then using ANOVA. See “Advantages of transforming the data first” on page 66.
Do you really want to compare medians?	The Kruskal-Wallis test compares the medians of three or more groups. It is possible to have a tiny P value – clear evidence that the population medians are different – even if the distributions overlap considerably.

Are the shapes of the distributions identical?	The Kruskal-Wallis test does not assume that the populations follow Gaussian distributions. But it does assume that the shapes of the distributions are identical. The medians may differ – that is what you are testing for – but the test assumes that the shapes of the distributions are identical. If two groups have very different distributions, consider transforming the data to make the distributions more similar.
--	---

The results of a Friedman test

How the Friedman test works

The Friedman test is a nonparametric test that compares three or more paired groups. 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.

The whole point of using a matched test is to control for experimental variability between subjects, thus increasing the power of the test. Some factors you don’t control in the experiment will increase (or decrease) all the measurements in a subject. Since the Friedman test ranks the values in each row, it is not affected by sources of variability that equally affect all values in a row (since that factor won’t change the ranks within the row).

The P value answers this question: If the different treatments (columns) really are identical, what is the chance that random sampling would result in sums of ranks as far apart (or more so) as observed in this experiment?

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 meets your needs.

If two or more values (in the same row) have the same value, it is impossible to calculate the exact P value, so Prism computes the approximate P value.

Following Friedman's test, Prism can perform Dunn's post test. For details, see [Applied Nonparametric Statistics](#) by WW Daniel, published by PWS-Kent publishing company in 1990 or [Nonparametric Statistics for Behavioral Sciences](#) by S Siegel and NJ Castellan, 1988. The original reference is O.J. Dunn, *Technometrics*, 5:241-252, 1964. Note that some

books and programs simply refer to this test as the post test following a Friedman test, and don't give it an exact name.

How to think about a Friedman test

The Friedman test is a nonparametric test to compare three or more matched groups. It is also called Friedman two-way analysis of variance by ranks (because repeated measures one-way ANOVA is the same as two-way ANOVA without any replicates.)

The P value answers this question: If the median difference really is zero, what is the chance that random sampling would result in a median difference as far from zero (or more so) as observed in this experiment?

If the P value is small, you can reject the idea that all of the differences between columns are coincidences of random sampling, and conclude instead that at least one of the treatments (columns) differs from the rest. Then look at post test results to see which groups differ from which other groups.

If the P value is large, the data do not give you any reason to conclude that the overall medians differ. This is not the same as saying that the medians are the same. You just have no compelling evidence that they differ. If you have small samples, Friedman's test has little power.

How to think about post tests following the Friedman test

Dunn's post test compares the difference in the sum of ranks between two columns with the expected average difference (based on the number of groups and their size). For each pair of columns, Prism reports the P value as >0.05 , <0.05 , <0.01 or <0.001 . The calculation of the P value takes into account the number of comparisons you are making. If the null hypothesis is true (all data are sampled from populations with identical distributions, so all differences between groups are due to random sampling), then there is a 5% chance that at least one of the post tests will have $P < 0.05$. The 5% chance does not apply to EACH comparison but rather to the ENTIRE family of comparisons.

Checklist. Is the Friedman test the right test for these data?

Before interpreting the results of any statistical test, first think carefully about whether you have chosen an appropriate test. Before accepting results from a Friedman test, ask yourself these questions:

Question	Discussion
Was the matching effective?	The whole point of using a repeated measures test is to control for experimental variability. Some factors you don't control in the experiment will affect all the measurements from one subject equally, so they will not affect the difference between the measurements in that subject. By analyzing only the differences, therefore, a matched test controls for some of the sources of scatter. The matching should be part of the experimental design and not something you do after collecting data. Prism does not test the adequacy of matching with the Friedman test.
Are the subjects (rows) independent?	The results of a Friedman test only make sense when the subjects (rows) are independent – that no random factor has effected values in more than one row. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six rows of data obtained from three animals in duplicate. In this case, some random factor may cause all the values from one animal to be high or low. Since this factor would affect two of the rows (but not the other four), the rows are not independent.
Are the data clearly sampled from non-Gaussian populations?	By selecting a nonparametric test, you have avoided assuming that the data were sampled from Gaussian distributions. But there are drawbacks to using a nonparametric test. If the populations really are Gaussian, the nonparametric tests have less power (are less likely to give you a small P value), especially with small sample sizes. Furthermore, Prism (along with most other programs) does not calculate confidence intervals when calculating nonparametric tests. If the distribution is clearly not bell-shaped, consider transforming the values (perhaps to logs or reciprocals) to create a Gaussian distribution and then using repeated measures ANOVA.

Two-way analysis of variance

	X Labels	A		B		C	
		Before		During		After	
	X	Y1	Y2	Y1	Y2	Y1	Y2
1	Men	123	132	143	154	162	156
2	Women	143	145	141	156	175	164

The ANOVA calculations ignore any X values. You may wish to format the X column as text in order to label your rows. Or you may omit the X column altogether, or enter numbers.

You may leave some replicates blank, and still perform ordinary two-way ANOVA (so long as you enter at least one value in each row for each data set). You cannot perform repeated measures ANOVA if there are any missing values.

If you have averaged your data elsewhere, you may format the data table to enter mean, SD/SEM and N. The N values do not have to all be the same, but you cannot leave N blank. You cannot perform repeated measures two-way ANOVA if you enter averaged data.

Some programs expect you to enter two-way ANOVA data in an indexed format. All the data are in one column, and two other columns are used to denote different levels of the two factors. You cannot enter data in this way into Prism. Prism can import indexed data with a single index variable, but cannot import data with two index variables, as would be required for two-way ANOVA.

As with one-way ANOVA, it often makes sense to transform data before analysis, to make the distributions more Gaussian. See "Advantages of transforming the data first" on page 66.

Introduction to two-way ANOVA

Two-way ANOVA, also called two factor ANOVA, determines how a response is affected by two factors. 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 factor systematically affect the results? In our example: Are the mean responses the same for all three drugs?
2. Does the second factor systematically affect the results? In our example: Are the mean responses the same for men and women?
3. Do the two factors interact? In our example: Are the difference between drugs the same for men and women? Or equivalently, is the difference between men and women the same for all drugs?

Although the outcome measure (dependent variable) is a continuous variable, each factor must be categorical, for example: male or female; low, medium or high dose; wild type or mutant. ANOVA is not an appropriate test for assessing the effects of a continuous variable, such as blood pressure or hormone level (use a regression technique instead).

Prism can perform ordinary two-way ANOVA accounting for repeated measures when there is matching on one of the factors (but not both). Prism cannot perform any kind of nonparametric two-way ANOVA.

Entering data for two-way ANOVA

Arrange your data so the data sets (columns) represent different levels of one factor, and different rows represent different levels of the other factor. For example, to compare three time points in men and women, enter your data like this:

Choosing the two-way ANOVA analysis

Start from the data or results table you wish to analyze (see "Entering data for two-way ANOVA" on page 93). Click Analyze and choose built-in analyses. Then choose two-way ANOVA from the list of statistical analyses

Parameters: Two-way Analysis of Variance

Variable names

Name of the variable that defines the columns:

(i.e. "Drug" or "Treatment")

Name of the variable that defines the rows:

(i.e. "Time", "Concentration" or "Gender")

Repeated Measures

No matching. Perform ordinary two-way ANOVA.

Matched values by row. Column A, Y1 matches column B, Y1.

Matched values by column. Row 1, Y1 matches row 2, Y1.

Post tests

Bonferroni post-tests to compare replicate means by row.

For each row, compare the mean of one column with the mean of the other column.

Output

Show significant figures

Variable names

Label the two factors to make the output more clear. If you don't enter names, Prism will use the generic names "Column factor" and "Row factor".

Repeated measures

You should choose a repeated measures analysis when the experiment used paired or matched subjects. See "Repeated measures test?" on page 67. Prism can calculate repeated measures two-way ANOVA with matching by either row or column, but not both. This is sometimes called a *mixed model*.

	X Labels	A		B	
		Control		Treated	
	X	Y1	Y2	Y1	Y2
1	one	23	24	28	31
2	two	34	41	41	54
3	three	43	47	56	60

The table above shows example data testing the effects of three doses of drugs in control and treated animals. The decision to use repeated measures depends on the experimental design.

Here is an experimental design that would require analysis using repeated measures by row: The experiment was done with six animals, two for each

dose. The control values were measured first in all six animals. Then you applied a treatment to all the animals, and made the measurement again. In the table above, the value at row 1, column A, Y1 (23) came from the same animal as the value at row 1, column B, Y1 (28). The matching is by row.

Here is an experimental design that would require analysis using repeated measures by column: The experiment was done with four animals. First each animal was exposed to a treatment (or placebo). After measuring the baseline data (dose = zero), you inject the first dose and make the measurement again. Then inject the second dose and measure again. The values in the first Y1 column (23, 34, 43) were repeated measurements from the same animal. The other three columns came from three other animals. The matching was by column.

The term *repeated measures* is appropriate for those examples, because you made repeated measurements from each animal. Some experiments involve matching but no repeated measurements. The term *randomized block* experiments describes this kind of experiments. For example, imagine that the three rows were three different cell lines. All the Y1 data came from one experiment, and all the Y2 data came from another experiment performed a month later. The value at row 1, column A, Y1 (23.0) and the value at row 1, column B, Y1 (28.0) came from the same experiment (same cell passage, same reagents). The matching is by row. Randomized block data are analyzed identically to repeated measures data. Prism only uses the term *repeated measures*.

It is also possible to design experiments with repeated measures in both directions. Here is an example: The experiment was done with two animals. First you measured the baseline (control, zero dose). Then you injected dose 1 and made the next measurement, then dose 2 and measured again. Then you gave the animal the experimental treatment, waited an appropriate period of time, and made the three measurements again. Finally, you repeated the experiment with another animal (Y2). So a single animal provided data from both Y1 columns (23, 34, 43 and 28, 41, 56). Prism cannot perform two-way ANOVA with repeated measures in both directions, and so cannot analyze this experiment.

Beware of matching within treatment groups – that does not count as repeated measures. Here is an example: The experiment was done with six animals. Each animal was given one of two treatments at one of three doses. The measurement was then made in duplicate. The value at row 1, column A, Y1 (23) came from the same animal as the value at row 1, column A, Y2 (24). Since the matching is within a treatment group, it is a replicate, not a repeated measure. Analyze these data with ordinary two-way ANOVA, not repeated measures ANOVA.

Post tests following two-way ANOVA

If you have two data sets (columns), Prism can perform post tests to compare the two means at each row. In the example above, data set A was control and data set B was treated. Each row represents a different dose. Prism can perform post tests to compare the control value and the treated value at each dose. Although other kinds of post tests are possible after two-way ANOVA, Prism only perform this one kind of post test (which biologists use most frequently).

The results of two-way ANOVA

How two-way ANOVA works

Two-way ANOVA, determines how a response is affected by two factors. For example, you might measure a response to three different drugs in both men and women.

The ANOVA table breaks down the overall variability between measurements (expressed as the sum of squares) into four components:

- Interactions between row and column. This is differences between rows that is not the same at each column, equivalent to variation between columns that is not the same at each row.
- Variability among columns.
- Variability among rows.
- Residual or error. Variation among replicates not related to systematic differences between rows and columns.

With repeated measures ANOVA there is a fifth component: variation between subjects.

The ANOVA table shows how the sum of squares is partitioned into the four (or five) components. For each component, the table shows sum-of-squares, degrees of freedom, mean square, and the F ratio. Each F ratio is the ratio of the mean-square value for that source of variation to the residual mean-square. (with repeated measures ANOVA, the denominator of one F ratio is the mean square for matching rather than residual). 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. The F ratios are not very informative by themselves, but are used to determine P values.

How Prism computes two-way ANOVA

Model I (fixed effects) vs. Model II (random effects) ANOVA

To understand the difference between fixed and random factors, consider an example of comparing responses in three species at three times. If you were interested in those three particular species, then species is considered to be a fixed factor. It would be a random factor if you were interested in differences between species in general, and randomly selected those three species. Time is considered to be a fixed factor if you chose 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 almost always considered to be a fixed factor.

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 calculates only Model I two-way ANOVA. Since most experiments deal with fixed-factor variables, this is rarely a limitation.

ANOVA from data entered as mean, SD (or SEM) and N

If your data are balanced (same sample size for each condition), you'll get the same results if you enter raw data, or mean, SD (or SEM) and N. If your data are unbalanced, it is impossible to calculate precise results from data entered as mean, SD (or SEM) and N. Instead, Prism uses a simpler method called analysis of "unweighted means". This method is detailed in LD Fisher and G vanBelle, *Biostatistics*, John Wiley, 1993. If sample size is the same in all groups, and in some other special cases, this simpler method gives exactly the same results as obtained by analysis of the raw data. In other 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

Two-way ANOVA calculations with missing values

If some values are missing, two-way ANOVA calculations are challenging. Prism uses the method detailed in SA Glantz and BK Slinker, *Primer of Applied Regression and Analysis of Variance*, McGraw-Hill, 1990. This method converts the ANOVA problem to a multiple regression problem, and then displays the results as ANOVA. 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. These are called Type III sum-of-squares.

Prism cannot perform repeated measures two-way ANOVA with missing values.

Two-way ANOVA from unreplicated data

Prism can perform two-way ANOVA even if you have entered only a single replicate for each column/row pair. This kind of data does not let you test for interaction between rows and columns (random variability and interaction can't be distinguished unless you measure replicates). 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.

The concept of repeated measures doesn't apply when your data are unreplicated.

Repeated measures two-way ANOVA

Prism computes repeated measures two-way ANOVA calculations using the standard method explained especially well in SA Glantz and BK Slinker, Primer of Applied Regression and Analysis of Variance, McGraw-Hill, 1990.

Post tests following two-way ANOVA

Prism performs post tests following two-way ANOVA using the Bonferroni method as detailed in pages 741-744 and 771 in J Neter, W Wasserman, and MH Kutner, Applied Linear Statistical Models, 3rd edition, Irwin, 1990.

For each row, Prism calculates

$$t = \frac{\text{mean}_1 - \text{mean}_2}{\sqrt{MS_{\text{residual}} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

The numerator is the difference between the mean response in the two data sets (usually control and treated) at a particular row (usually dose or time point). The denominator combines the number of replicates in the two groups at that dose with the mean square of the residuals (sometimes called the mean square of the error), which is a pooled measure of variability at all doses.

Statistical significance is determined by comparing the t ratio with the t distribution for the number of df shown in the ANOVA table for MS_{residual} ,

applying the Bonferroni correction for multiple comparisons. The Bonferroni correction lowers the P value that you consider to be significant to 0.5 divided by the number of comparisons. This means that if you have five rows of data, the P value has to be less than 0.01 (0.5/5) for any particular row in order to be considered significant with $P < 0.05$. This correction ensures that the 5% probability applies to the entire family of comparisons, and not separately to each individual comparison.

Confidence intervals at each row are computed using this equation:

$$\text{Span} = t^* \cdot \sqrt{MS_{\text{residual}} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

$$95\% \text{ CI: } [(\text{mean}_2 - \text{mean}_1) - \text{Span}] \text{ to } [(\text{mean}_2 - \text{mean}_1) + \text{Span}]$$

The critical value of t is abbreviated t^* in that equation (not a standard abbreviation). Its value does not depend on your data, only on your experimental design. It depends on the number of degrees of freedom and the number of rows (number of comparisons).

Post tests following repeated measures two-way ANOVA use exactly the same equation if the repeated measures are by row. If the repeated measures are by column, use MS_{subject} rather than MS_{residual} in both equations above, and use the degrees of freedom for subjects (matching) rather than the residual degrees of freedom.

How to think about results from two-way ANOVA

Two-way ANOVA partitions the overall variance of the outcome variable into three components plus a residual (or error) term.

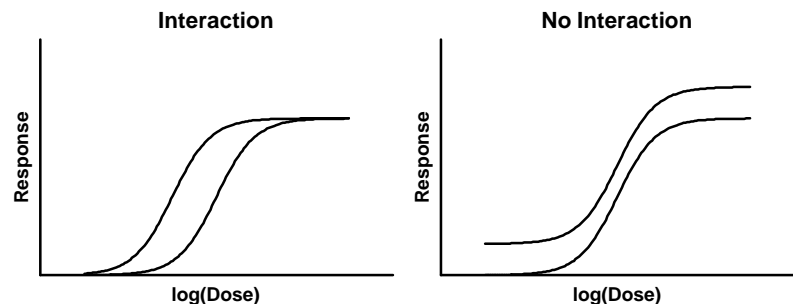
Interaction

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 each row represents a time or concentration, there is no interaction if the vertical difference between the curves is the same for all values of X. Some statistics books say that there is no interaction when the curves are

"parallel". But that term can be ambiguous. 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. Two-way ANOVA of such data 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.



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.

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 obtaining column means as different (or more so) than you have observed.

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 obtaining row means as different (or more so) than you have observed.

Subject (matching)

For repeated measures ANOVA, Prism tests the null hypothesis that the matching was not effective. You expect a low P value if the repeated measures design was effective in controlling for variability between subjects. If the P value was high, reconsider your decision to use repeated measures ANOVA.

How to think about post tests following two-way ANOVA

If you have two data sets (columns), Prism can perform post tests to compare the two means from each row.

For each row, Prism reports the 95% confidence interval for the difference between the two means. These confidence intervals adjust for multiple comparisons, so you can be 95% certain that all the intervals contain the true difference between means.

For each row, Prism also reports the P value testing the null hypothesis that the two means are really identical. Again, the P value computations take into account multiple comparisons. If there really are no differences, there is a 5% chance that any one (or more) of the P values will be less than 0.05. The 5% probability applies to the entire family of comparisons, not to each individual P value.

If the difference is statistically significant

If the P value for a post test is small, then it is unlikely that the difference you observed is due to a coincidence of random sampling. You can reject the idea that those two populations have identical means.

Because of random variation, the difference between the group means in this experiment is unlikely to equal the true difference between population means. There is no way to know what that true difference is. With most post tests (but not the Newman-Keuls test), Prism presents the uncertainty as a 95% confidence interval for the difference between all (or selected) pairs of means. You can be 95% sure that this interval contains the true difference between the two means.

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a difference between means that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial difference	Trivial difference	Although the true difference is not zero (since the P value is low) the true difference between means is tiny and uninteresting. The treatment had an effect, but a small one.
Trivial difference	Important difference	Since the confidence interval ranges from a difference that you think are biologically trivial to one you think would be important, you can't reach a strong conclusion from your data. You can conclude that the means are different, but you don't know whether the size of that difference is scientifically trivial or important. You'll need more data to draw a clear conclusion.
Important difference	Important difference	Since even the low end of the confidence interval represents a difference large enough to be considered biologically important, you can conclude that there is a difference between treatment means and that the difference is large enough to be scientifically relevant.

If the difference is not statistically significant

If the P value from a post test is large, the data do not give you any reason to conclude that the means of these two groups differ. Even if the true means were equal, you would not be surprised to find means this far apart just by coincidence. This is not the same as saying that the true means are the same. You just don't have evidence that they differ.

How large could the true difference really be? Because of random variation, the difference between the group means in this experiment is unlikely to equal the true difference between population means. There is no way to know what that true difference is. Prism presents the uncertainty as a 95% confidence interval (except with the Newman-Keuls test). You can be 95% sure that this interval contains the true difference between the two means. When the P value is larger than 0.05, the 95% confidence interval will start with a negative number (representing a decrease) and go up to a positive number (representing an increase).

To interpret the results in a scientific context, look at both ends of the confidence interval for each pair of means, and ask whether those differences would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial decrease	Trivial increase	You can reach a crisp conclusion. Either the means really are the same or they are different by a trivial amount. At most, the true difference between means is tiny and uninteresting.
Trivial decrease	Large increase	You can't reach a strong conclusion. The data are consistent with the treatment causing a trivial decrease, no change, or a large increase. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Large decrease	Trivial increase	You can't reach a strong conclusion. The data are consistent with a trivial increase, no change, or a decrease that may be large enough to be important. You can't make a clear conclusion without repeating the experiment with more subjects.
Large decrease	Large increase	You can't reach any conclusion. Repeat the experiment with a much larger sample size.

Problems with post tests following two-way ANOVA

Post test are often used to compare dose-response curves or time course curves. Using two-way ANOVA in this way presents two problems. One problem is that ANOVA treats different doses (or time points) exactly as it deals with different species or different drugs. ANOVA ignores the fact that doses or time points come in order. You could jumble the doses in any order, and get exactly the same ANOVA results. However, you did the experiment to observe a trend, so you should be cautious about interpreting results from an analysis method that doesn't recognize trends.

Another problem with the ANOVA approach is that it is hard to interpret the results. Knowing at which doses or time points the treatment had a statistically significant effect doesn't always help you understand the biology of the system, and rarely helps you design new experiments. Some scientists like to ask which is the lowest dose (or time) at which the effect of the treatment is statistically significant. The post tests give you the answer,

but the answer depends on sample size. Run more subjects, or more doses or time points for each curve, and the answer will change. Rather than two-way ANOVA, consider using linear or nonlinear regression to fit the curve to a model and then compare the fits.

Checklist. Is two-way ANOVA the right test for these data?

Before accepting the results of any statistical test, first think carefully about whether you chose an appropriate test. Before accepting results from a one-way ANOVA, ask yourself these questions:

Question	Discussion
Are the populations distributed according to a Gaussian distribution?	Two-way ANOVA assumes that your replicates are sampled from Gaussian distributions. While this assumption is not too important with large samples, it is important with small sample sizes, especially with unequal sample sizes. Prism does not test for violations of this assumption. 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 non-parametric two-way ANOVA. Prism does not offer this test. ANOVA also assumes that all sets of replicates have the same SD overall, and that any differences between SDs are due to random sampling.
Are the data matched?	Standard two-way ANOVA works by comparing the differences among group means with the pooled standard deviations of the groups. If the data are matched, then you should choose repeated measures ANOVA instead. If the matching is effective in controlling for experimental variability, repeated measures ANOVA will be more powerful than regular ANOVA.

Are the "errors" independent?	The term "error" refers to the difference between each value and the mean of all the replicates. The results of two-way ANOVA only make sense when the scatter is random – that whatever factor caused a value to be too high or too low affects only that one value. Prism cannot test this assumption. You must think about the experimental design. For example, the errors are not independent if you have six replicates, but these were obtained from two animals in triplicate. In this case, some factor may cause all values from one animal to be high or low. See "The need for independent samples" on page 5.
Do you really want to compare means?	Two-way ANOVA compares the means. It is possible to have a tiny P value – clear evidence that the population means are different – even if the distributions overlap considerably. In some situations – for example, assessing the usefulness of a diagnostic test – you may be more interested in the overlap of the distributions than in differences between means.
Are there two factors?	One-way ANOVA compares three or more groups defined by one factor. For example, you might compare a control group, with a drug treatment group and a group treated with drug plus antagonist. Or you might compare a control group with five different drug treatments. Prism has a separate analysis for one-way ANOVA. Some experiments involve more than two factors. For example, you might compare three different drugs in men and women at four time points. There are three factors in that experiment: drug treatment, gender and time. These data need to be analyzed by three-way ANOVA, also called three factor ANOVA. Prism does not perform three-way ANOVA.

Are both factors “fixed” rather than “random”?	Prism performs Type I ANOVA, also known as fixed-effect ANOVA. This tests for differences among the means of the particular groups you have collected data from. Different calculations are needed if you randomly selected groups from an infinite (or at least large) number of possible groups, and want to reach conclusions about differences among ALL the groups, even the ones you didn’t include in this experiment. See "Model I (fixed effects) vs. Model II (random effects) ANOVA" on page 98.
--	---

The circularity assumption in two-way repeated measures ANOVA

Repeated measures ANOVA assumes that the random error truly is random. A random factor that causes a measurement in one subject to be a bit high (or low) should have no affect on the next measurement in the same subject. This assumption is called *circularity* or *sphericity*. It is closely related to another term you may encounter, *compound symmetry*.

Repeated measures ANOVA is quite sensitive to violations of the assumption of circularity. If the assumption is violated, the P value will be too low. You’ll violate this assumption when the repeated measurements are made too close together so that random factors that cause a particular value to be high (or low) don’t wash away or dissipate before the next measurement. To avoid violating the assumption, wait long enough between treatments so the subject is essentially the same as before the treatment. Also randomize the order of treatments, when possible.

You only have to worry about the assumption of circularity when your experiment truly is a repeated measures experiment, with measurements from a single subject. You don’t have to worry about circularity with randomized block experiments, where you used a matched set of subjects (or a matched set of experiments).

Survival curves

Introduction to survival curves

With many kinds of experiments, the outcome is a survival time and you wish to compare the survival of two or more groups. Prism creates survival curves, which plot percent survival as a function of time, using the method of Kaplan and Meier. Prism can also compare two or more survival curves using the logrank test.

The name *survival curve* is a bit misleading. The methods described in this chapter can analyze any kind of experiment where the result is expressed as a time to a well-defined end point. Instead of death, the endpoint could be 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 end point must be a one-time event. Recurring events should not be analyzed with survival curves.

Some kinds of survival data are better analyzed with nonlinear regression. For example, don't use the methods in this chapter to analyze cell survival curves plotting percent survival (Y) as a function of various doses of radiation (X). The survival methods described in this chapter are only useful if X is time.

Entering survival data

What are censored data?

When creating a survival curve, you rarely know the survival time for each subject. Some subjects are still alive at the end of the study. You know how long they have survived so far, but don't know how long they will survive in the future. Others drop out of the study – perhaps they moved to a different city or wanted to take a medication disallowed on the protocol. You know they survived a certain length of time on the protocol, but don't know how long they survived after that (or do know, but can't use the

information because they weren't following the experimental protocol). In both cases, information about these patients is *censored*. When calculating and comparing survival curves, Prism automatically accounts for censored data.

Arranging survival data

Format the data sheet for single Y values (no replicates; no error bars). Enter each subject on a separate row in the table.

In the X column, enter time until censoring or death (or whatever event you are tracking). Use any convenient unit, such as days or months. Time zero does not have to be 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.

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. Every subject in a survival study either dies or is 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 shown in the figure below: 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. Place the X values for the second group of subjects beneath those for the first group (X values do not have to be sorted). Place the corresponding Y codes in the second Y column, leaving the first column blank. In the example below, data for group A were entered in the first 14 rows, and data for group B started in row 15.

10	46.0	1.0	
11	64.0	0.0	
12	78.0	0.0	
13	124.0	1.0	
14	127.0	0.0	
15	9.0		1.0
16	26.0		0.0
17	43.0		1.0
18	64.0		1.0

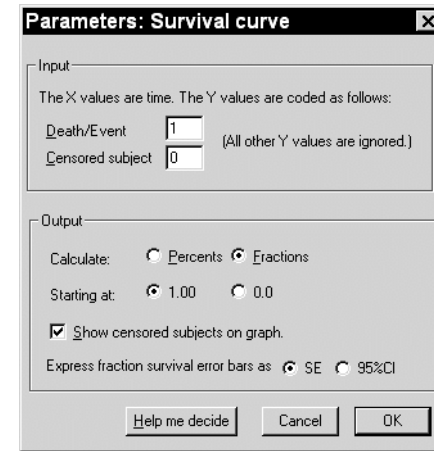
If the treatment groups are intrinsically ordered (perhaps increasing dose) maintain that order when entering data. Make the progression from column A to column B to column C follow the natural order of the treatment groups.

Common questions about entering survival data

Question	Answer
How do I enter data for subjects still alive at the end of the study?	Those subjects are said to be censored. You know how long they survived so far, but don't know what will happen later. X is the # of days (or months...) they were followed. Y is the code for censored observations, usually zero.
What if two or more subjects died at the same time?	Each subject must be entered on a separate row. Enter the same X value on two (or more) rows.
How do I enter data for a subject who died of an unrelated cause?	Different investigators handle this differently. Some treat a death as a death, no matter what the cause. Others treat death of an unrelated cause to be a censored observation. Ideally, this decision should be made in the study design. If the study design is ambiguous, you should decide how to handle these data before unblinding the study.
Do the X values have to be entered in order?	No. You can enter the data in any order you want.
How does Prism distinguish between subjects who are alive at the end of the study and those who dropped out of the study?	It doesn't. In either case, the observation is censored. You know the patient was alive and on the protocol for a certain period of time. After that you can't know (patient still alive), or can't use (patient stopped following the protocol) the information. Survival analysis calculations treat all censored subjects in the same way. Until the time of censoring, censored subjects contribute towards calculation of percent survival. After the time of censoring, they are essentially missing data.
I already have a life-table showing percent survival at various times. Can I enter this table into Prism?	No. Prism only can analyze survival data if you enter survival time for each subject. Prism can not analyze data entered as a life table.
Can I enter a starting and ending date, rather than duration?	No. You must enter the number of days (or months, or some other unit of time). Use a spreadsheet to subtract dates to calculate duration.

Choosing a survival analysis

To create and compare survival curves with Prism, start from your table of survival data (see "Entering survival data" on page 109). Press Analyze, and choose built-in analyses. Then choose Survival analysis from the list of statistical analyses.



Ordinarily, $Y=0$ indicates a censored subject and $Y=1$ indicates a death. You may define other codes (use digits, not letters).

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 death, it is traditional to show percent (or fraction) survival rather than percent 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.

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.

Choose whether you want to tabulate the uncertainty in survival as a standard error or as a 95% confidence interval.

Interpreting survival analysis

How survival analysis works

The fraction (or percent) survival at each time

Prism calculates survival fractions using the product limit or Kaplan-Meier method. For each X value (time) Prism shows the fraction still alive (or the fraction already dead, if you chose to begin the curve at 0.0 rather than 1.0). This table contains the numbers used to graph survival vs. time.

Prism also reports the uncertainty of the fractional survival as a standard error or 95% confidence intervals. Standard errors are calculated by the method of Greenwood. The 95% confidence intervals are computed as 1.96 times the standard error in each direction. In some cases the confidence interval calculated this way would start below 0.0 or end above 1.0 (or 100%). In these cases, the error bars are clipped to avoid impossible values.

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 in 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 logrank 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?

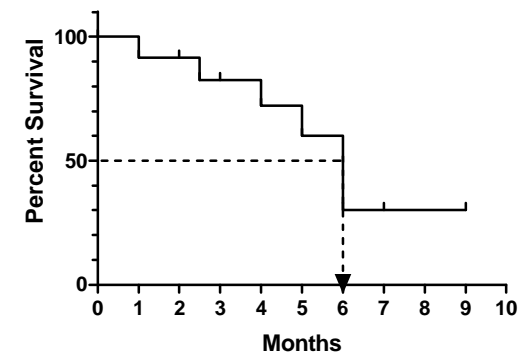
Prism always calculates two-tailed P values. If you wish to report a one-tailed P value, you must have predicted which group would have the longer median survival before collecting any data. If your prediction was correct, the one-tail P value is half the two-tail P value. If your prediction was wrong, the one-tail P value is greater than 0.50, and you must conclude that the difference was due to chance, no matter how large it is.

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.

Median survival

The median survival is the time at which half the subjects have died. The example survival curve below shows 50% survival at 6 months, so median survival is 6 months.



If survival exceeds 50% at the longest time points, then median survival cannot be computed. If the survival curve is horizontal at 50% survival, Prism reports the median survival as the average of the first and last times at which survival is 50%.

When comparing two survival curves, Prism also reports the ratio of the median survival times along with its 95% confidence interval. You can be 95% sure that the true ratio of median survival times lies within that range.

Hazard ratio

If you compare two survival curves, Prism reports the hazard ratio and its 95% confidence interval.

The hazard is the slope of the survival curve – a measure of how rapidly subjects are dying. The hazard ratio compares two treatments. If the hazard ratio is 2.0, then the rate of deaths in one treatment group is twice the rate in the other group.

The computation of the hazard ratio assumes that the ratio is consistent over time, and that any differences are due to random sampling. So Prism

reports a single hazard ratio, not a different hazard ratio for each time interval. If two survival curves cross, the hazard ratios are not equal.

Reference for survival calculations

Most of the calculations are quite standard. Prism follows the calculations as spelled out in detail in DG Altman, Practical Statistics for Medical Research, 1991, Chapman and Hall.

For calculating the logrank test, Prism uses the second of the two methods described in that reference (more difficult to calculate, but more accurate). Some books call this method the Mantel-Haenszel logrank test.

How to think about survival curve results

As with any analysis, you need to look both at the P value and at the confidence intervals.

This section applies only if you are comparing two survival curves. The logrank test computes a P value that answers this question: If the two populations have identical survival curves overall, what is that chance that random sampling of subjects would lead to as big a difference in survival (or bigger) as you observed.

To interpret the P value, you also need to consider a confidence interval. Prism reports two confidence intervals: the confidence interval for the ratio of median survival and the confidence interval for the hazard ratio. The discussion below assumes that you are focusing on the ratio of median survival times, but you can use the same logic if you focus on the hazard ratio.

If the difference is statistically significant – the P value is small

If the P value is small, then it is unlikely that the difference you observed is due to a coincidence of random sampling. You can reject the idea that the two populations have identical survival characteristics.

Because of random variation, the ratio of median survival times (or the hazard ratio) between the two groups in this experiment is unlikely to equal the true ratio. There is no way to know what that true ratio is. Prism presents the uncertainty as a 95% confidence interval for the ratio of median survival times (and the hazard ratio). You can be 95% sure that this interval contains the true ratio.

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent a ratio far enough from 1.0 to be scientifically important. How far is "far"? How close is "close"? That depends on the reasons you did the experiment. In some cases, you may think that a 10% increase or decrease in median survival would be very important. In other cases you may care only about a 50% change.

Lower confidence limit	Upper confidence limit	Conclusion
Close to 1.0	Close to 1.0	Although the true ratio is not 1.0 (since the P value is low) the difference between median survival is tiny and uninteresting. The treatment had an effect, but a small one.
Close to 1.0	Far from 1.0	Since the confidence interval ranges from a ratio that you think is trivially different than 1.0 to one you think is far from 1.0, you can't reach a strong conclusion from your data. You can conclude that the median survival times are different, but you don't know whether the size of that difference is scientifically trivial or important. You'll need more data to obtain a clear conclusion.
Far from 1.0	Far from 1.0	Since even the low end of the confidence interval represents a ratio far from 1.0 and thus considered biologically important, you can conclude that there is a difference between median survival times, and that the difference is large enough to be scientifically relevant.

If the difference is not statistically significant -- the P value is large (logrank test)

If the P value from the logrank is large, the data do not give you any reason to conclude that the median survival times of the two groups are really different. Even if the true median survival times were equal, you would not be surprised to find medians this far apart just by coincidence. This is not the same as saying that the true median times are the same. You just don't have evidence that they differ.

How large could the true difference really be? Because of random variation, the ratio of median survival times in this experiment is unlikely to equal the true ratio in the entire population. There is no way to know what that true ratio is. Prism presents the uncertainty as a 95% confidence interval (except with the Newman-Keuls test). You can be 95% sure that this interval contains the true ratio of median survival times. When the P

value is larger than 0.05, the 95% confidence interval will start with a ratio below 1.0 (representing a decrease) and go up to a ratio greater than 1.0 (representing an increase).

To interpret the results in a scientific context, look at both ends of the confidence interval, and ask whether they are close to or far from 1.0.

Lower confidence limit	Upper confidence limit	Conclusion
Close to 1.0	Close to 1.0	You can reach a crisp conclusion. Either the median survival times really are the same, or they are different by a trivial amount. At most, the true difference between median survival times is tiny and uninteresting.
Close to 1.0	Far from 1.0	You can't reach a strong conclusion. The data are consistent with the treatment causing a trivial decrease, no change, or a large increase. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Far from 1.0	Close to 1.0	You can't reach a strong conclusion. The data are consistent with a trivial increase, no change, or a decrease that may be large enough to be important. You can't make a clear conclusion without repeating the experiment with more subjects.
Far from 1.0	Far from 1.0	You cannot reach any conclusion. Your data are consistent with no change, or a difference in either direction that may be large enough to be scientifically relevant. You can't make a clear conclusion without repeating the experiment with more subjects.

Checklist for interpreting survival analyses

Question	Discussion
Are the subjects independent?	Factors that influence survival should either affect all subjects in a group, or just one subject. If the survival of several subjects is linked, then you don't have independent observations. For example, if the study pools data from two hospitals, the subjects are not independent, as it is possible that subjects from one hospital have different average survival times than subjects from another. You could alter the median survival curve by choosing more subjects from one hospital and fewer from the other. To analyze these data, use Cox proportional hazards regression which Prism cannot perform.
Were the entry criteria consistent?	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. Here's why: Patients die at the same age they otherwise would, but are diagnosed when they are younger, and so live longer with the diagnosis.
Was the end point defined consistently?	If the curve is plotting time to death, then there can be ambiguity about which deaths to count. In a cancer trial, for example, what happens to subjects who die in an automobile accident? Some investigators count these as deaths, others count them as censored subjects. Both approaches can be justified, but the approach should be decided before the study begins. If there is any ambiguity about which deaths to count, the decision should be made by someone who doesn't know which patient is in which treatment group. If the curve plots time to an event other than death, it is crucial that the event be assessed consistently throughout the study.

Is time of censoring is unrelated to survival?

The survival analysis only is valid when the survival times of censored patients are identical to the survival of subjects who stayed with the study. If a large fraction of subjects are censored, the validity of this assumption is critical to the integrity of the results. There is no reason to doubt that assumption for patients still alive at the end of the study. When patients drop out of the study, you should ask whether the reason could affect survival. 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.

Does average survival stay constant 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 the 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.

Is the assumption of proportional hazards reasonable?

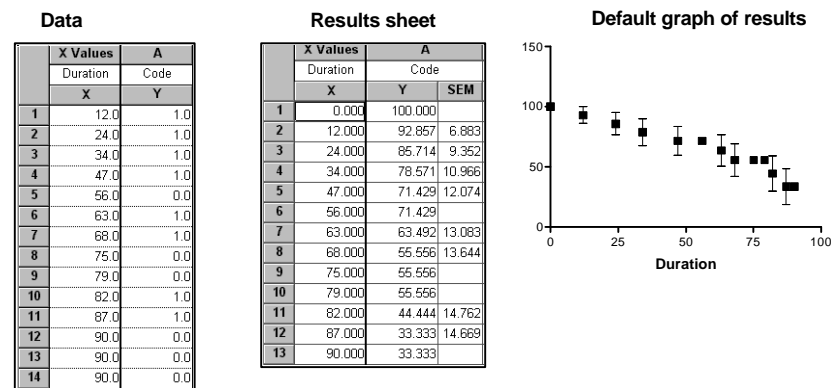
The logrank test is only strictly valid when the survival curves have proportional hazards. This means that the rate of dying in one group is a constant fraction of the rate of dying in the other group. This assumption has proven to be reasonable for many situations. It would not be reasonable, for example, if you are comparing a medical therapy with a risky surgical therapy. At early times, the rate of dying might be much higher in the surgical group. At later times the rate of dying might be greater in the medical group. Since the hazard ratio is not consistent over time (the assumption of proportional hazards is not reasonable), these data should not be analyzed with a logrank test.

Were the treatment groups defined before data collection began?

It is not valid to divide a single group of patients (all treated the same) into two groups based on whether they responded to treatment (tumor got smaller, lab tests got better). By definition, the responders must have lived long enough to see the response. And they may have lived longer anyway, regardless of treatment. When you compare groups, the groups must be defined before data collection begins.

Graphing survival curves

The results sheet for survival analysis has several views. One is a table of fractional survival as a function of time. Prism graphs this sheet to make a survival curve. Below is an example, showing the data, the results sheet, and the graph that Prism made automatically.

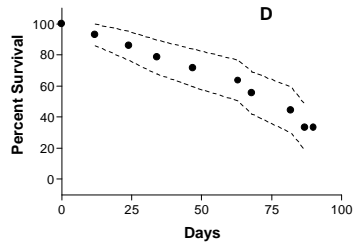
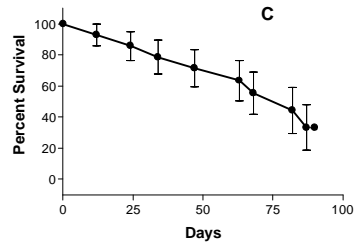
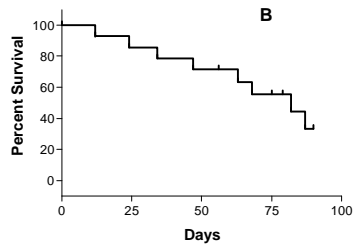
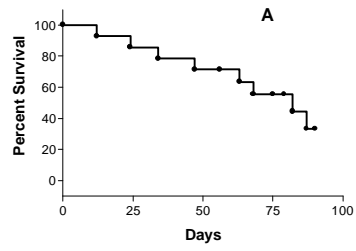


Choices on the analysis parameters dialog made the graph start at 100% and go down to zero, showing percent survival with its standard error at every time at which a subject died. It also shows percent survival (with no error bar) at the times at which a subject was censored (left the study, etc.). You could change an option on the parameters dialog to not show these censored points.

The automatic graph is not satisfactory. One problem is that the Y-axis extends too far. Double-click on the axis, and change the range to go from -10 to 110. Then double-click on one of the symbols to bring up the Format Symbols to change the appearance. In graph A below, the data points are shown as circles connected by a staircase. There are no error bars. In graph B, the data points are shown as ticks (the choice fifth from the bottom in the list of shapes). You only see ticks for censored subjects because the other ticks are superimposed on the staircase. These two kinds of survival graphs are commonly seen in the clinical literature.

Examples C and D include error bars to demonstrate how precisely you know the survival curve. In example C, the data points are circles connected by point-to-point lines. Error bars show plus and minus one standard error (they would show 95% confidence, if you had selected that choice in the analysis parameters dialog). Some statisticians dislike showing survival data with point-to-point lines, because the survival in your sample actually follows a staircase pattern – the survival changes only at the instant each subject dies. Example D does not connect the points at all, but shows

an error envelope. This is created by choosing "- - - -" in the list of error bar styles.



Contingency tables

Introduction to contingency tables

Contingency tables summarize results where the outcome is a categorical variable such as disease vs. no disease, pass vs. fail, artery open vs. artery obstructed.

Use contingency tables to display the results of four kinds of experiments.

In a **cross-sectional study**, you recruit a single group of subjects and then classify them by two criteria (row and column). As an example, let's consider how to conduct a cross-sectional study of the link between electromagnetic fields (EMF) and leukemia. To perform a cross-sectional study of the EMF-leukemia link, you would need to study a large sample of people selected from the general population. You would assess whether or not each subject has been exposed to high levels of EMF. This defines the two rows in the study. You then check the subjects to see whether or not they have leukemia. This defines the two columns. It would not be a cross-sectional study if you selected subjects based on EMF exposure or on the presence of leukemia.

A **prospective study** starts with the potential risk factor and looks forward to see what happens to each group of subjects. To perform a prospective study of the EMF-leukemia link, you would select one group of subjects with low exposure to EMF and another group with high exposure. These two groups define the two rows in the table. Then you would follow all subjects over time and tabulate the numbers that get leukemia. Subjects that get leukemia are tabulated in one column; the rest are tabulated in the other column.

A **retrospective case-control study** starts with the condition being studied and looks backwards at potential causes. To perform a retrospective study of the EMF-leukemia link, you would recruit one group of subjects with leukemia and a control group that does not have leukemia but is otherwise similar. These groups define the two columns. Then you would assess EMF exposure in all subjects. Enter the number with low exposure in one row,

and the number with high exposure in the other row. This design is also called a case control study

In an **experiment**, you manipulate variables. Start with a single group of subjects. Half get one treatment, half the other (or none). This defines the two rows in the study. The outcomes are tabulated in the columns. For example, you could perform a study of the EMF/leukemia link with animals. Half are exposed to EMF, while half are not. These are the two rows. After a suitable period of time, assess whether each animal has leukemia. Enter the number with leukemia in one column, and the number without leukemia in the other column. Contingency tables can also tabulate the results of some basic science experiments. The rows represent alternative treatments, and the columns tabulate alternative outcomes.

If the table has two rows and two columns, Prism computes P values using either Fisher's exact test or the chi-square test, and summarizes the data by computing the relative risk, odds ratio or difference in proportions, along with 95% confidence intervals. If the table has two rows and three or more columns (or three or more rows and two columns) Prism calculates both the chi-square test and the chi-square test for trend. With larger tables, Prism only calculates the chi-square test.

Entering data into contingency tables

You must enter data in the form of a contingency table. Prism cannot tabulate raw data to create a contingency table. Prism also cannot compare proportions directly. You need to enter the number of subjects in each category – you cannot enter fractions or percentages.

The values entered in a contingency table represent the number of subjects actually observed in this experiment. Tables of averages, percentages or rates are not contingency tables, and cannot be analyzed by chi-square and related tests. Note also that the columns define mutually exclusive categories, as do the rows. A subject can be in one or the other, but not both.

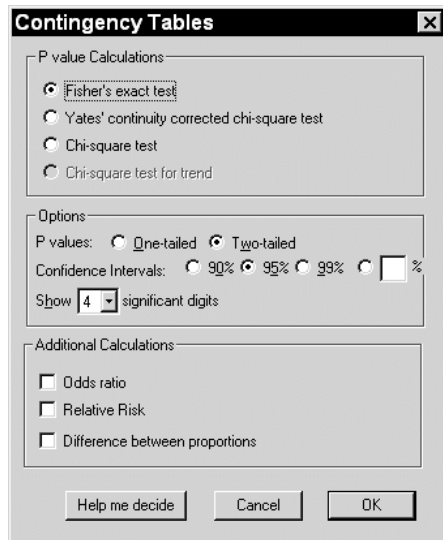
Most contingency tables have two rows (two groups) and two columns (two possible outcomes). For data from prospective and experimental studies, the top row usually represents exposure to a risk factor or treatment, and the bottom row is for controls. The left column usually tabulates the number of individuals with disease; the right column is for those without the disease. In case-control retrospective studies, the left column is for cases; the right column is for controls. The top row tabulates the number of individuals exposed to the risk factor; the bottom row is for those not exposed.

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.

Choosing how to analyze a contingency table

To analyze a contingency table, start from your data table (see "Entering data into contingency tables" on page 122). Press Analyze and choose built-in analyses. Then choose Contingency table analysis from the list of statistical tests.



Prism offers two methods for calculating a P value from tables with two rows and two columns: Fisher's exact test and the chi-square test. We recommend always picking Fisher's test, as it calculates a P value that is exactly correct. The only advantage of the chi-square test is that it is easier to calculate by hand, and so is better known. We don't recommend using it to analyze contingency tables with two rows and two columns.

If you choose a chi-square test, also choose whether to apply Yates' continuity correction. 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 select Fisher's exact test (recommended), Yates' correction is of no concern.

If your table includes very large numbers (thousands), Prism will automatically perform the chi-square test even if you select Fisher's test. This is because the Fisher's test calculations are slow with large samples. With large samples, the chi-square test is very accurate and Yates' continuity correction has negligible effect.

Choose a two-sided P value, unless you have a good reason to pick a one-sided P value. (With contingency tables, Prism refers to "two-sided" P values rather than "two-tail P value" – the distinction is subtle and not worth worrying about.) See "One- vs. two-tail P values" on page 10.

If your table has more than two rows or two columns, Prism always calculates the chi-square test. You have no choice. 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.

Prism can summarize data on a two by two table with the relative risk, difference between proportions (P1-P2) and/or the odds ratio. If your data came from an experimental, cross-sectional or prospective study, summarize the data by choosing the relative risk or P1-P2. If your data came from a retrospective case-control study, pick only the odds ratio.

Interpreting analyses of contingency tables

How analyses of 2x2 contingency tables work

If your table has two rows and two columns, Prism computes relative risk, odds ratio and P1-P2 using the equations below:

	Outcome 1	Outcome 2
Group 1	A	B
Group 2	C	D

$$\text{Relative Risk} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

$$P1 - P2 = \frac{A}{A+B} - \frac{C}{C+D}$$

$$\text{Odds Ratio} = \frac{\frac{A}{B}}{\frac{C}{D}}$$

If any of the four values in the contingency table are zero, Prism adds 0.5 to all values before calculating the relative risk, odds ratio and P1-P2 (to avoid dividing by zero).

The word "risk" is appropriate when the first row is the exposed or treated group and the left column is the bad outcome. With other kinds of data, the term "risk" isn't appropriate, but you may still be interested in the ratio of proportions. Prism calculates the 95% confidence interval for the relative risk using the approximation of Katz. You can be 95% certain that this range includes the true relative risk.

If your data are from a case-control retrospective study, neither the relative risk nor P1-P2 is meaningful. Instead, Prism calculates an odds ratio and the confidence interval of the odds ratio using the approximation of Woolf. If the disease is rare, you can think of an odds ratio as an approximation of the relative risk.

Prism computes the P value using either the chi-square test or Fisher's exact test.

How analyses of larger contingency tables work

If your table has two columns and more than two rows (or two rows and more than two columns), Prism will perform both the chi-square test for independence and the chi-square test for trend.

The **chi-square test for independence** asks whether there is an association between the variable that defines the rows and the variable that defines the columns.

Prism first computes the expected values for each value. These expected values are calculated from the row and column totals, and are not displayed in the results. The discrepancies between the observed values and expected values are then pooled to compute chi-square, which is reported. A large value of chi-square tells you that there is a large discrepancy. The P value answers this question: If there is really no association between the variable that defines the rows and the variable that defines the columns, then what is the chance that random sampling would result in a chi-square value as large (or larger) as you obtained in this experiment.

The P value from the **chi-square test for trend** answers this question: If there is no linear trend between row (column) number and the fraction of subjects in the left column (top row), what is the chance that you would happen to observe such a strong trend as a coincidence of random sampling? If the P value is small, you will conclude that there is a statistically significant trend.

For more information about the chi-square test for trend, see the excellent text, Practical Statistics for Medical Research by D. G. Altman, published in 1991 by Chapman and Hall.

How to think about the relative risk, odds ratio and P1-P2

To understand the differences between the relative risk, odds ratio and P1-P2 consider this example. There are two groups of subjects, denoted by two rows. There are two outcomes denoted by columns:

	X Labels	A	B
	X Labels	Progress	No Progress
	X	Y	Y
1	AZT	76.00000	399.00000
2	Placebo	129.00000	332.00000

Method	Description
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.

How to think about P values from a 2x2 contingency table

The P value answers this question: If there really is no association between the variable defining the rows and the variable defining the columns in the overall population, what is the chance that random sampling would result in an association as strong (or stronger) as observed in this experiment? Equivalently, if there really is no association between rows and columns overall, what is the chance that random sampling would lead to a relative

risk or odds ratio as far (or further) from 1.0 (or P1-P2 as far from 0.0) as observed in this experiment?

“Statistically significant” is not the same as “scientifically important”. Before interpreting the P value or confidence interval, you should think about the size of the relative risk, odds ratio or P1-P2 you are looking for. How large does the value need to be for you consider it to be scientifically important? How small a value would you consider to be scientifically trivial? Use scientific judgment and common sense to answer these questions. Statistical calculations cannot help, as the answers depend on the context of the experiment.

You will interpret the results differently depending on whether the P value is small or large.

If the P value is small

If the P value is small, then it is unlikely that the association you observed is due to a coincidence of random sampling. You can reject the idea that the association is a coincidence, and conclude instead that the population has a relative risk or odds ratio different than 1.0 (or P1-P2 different than zero). The association is statistically significant. But is it scientifically important? The confidence interval helps you decide.

Your data include the effects of random sampling, so the true relative risk (or odds ratio or P1-P2) is probably not the same as the value calculated from the data in this experiment. There is no way to know what that true value is. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true relative risk, odds ratio or P1-P2.

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent values that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial	Trivial	Although the true relative risk or odds ratio is not 1.0 (and the true P1-P2 is not 0.0) the association is tiny and uninteresting. The variables defining the rows is associated with the variable defining the columns, but weakly.
Trivial	Important	Since the confidence interval ranges from a relative risk (or odds ratio or P1-P2) that you think is biologically trivial to one you think would be important, you can't reach a strong conclusion from your data. You can conclude that the rows and columns are associated, but you don't know whether the association is scientifically trivial or important. You'll need more data to obtain a clear conclusion.
Important	Important	Since even the low end of the confidence interval represents an association large enough to be considered biologically important, you can conclude that the rows and columns are associated, and the association is strong enough to be scientifically relevant.

If the P value is large

If the P value is large, the data do not give you any reason to conclude that the relative risk or odds ratio differs from 1.0 (or P1-P2 differs from 0.0). This is not the same as saying that the true relative risk or odds ratio equals 1.0 (or P1-P2 equals 0.0). You just don't have evidence that they differ.

How large could the true relative risk really be? Your data include the effects of random sampling, so the true relative risk (or odds ratio or P1-P2) is probably not the same as the value calculated from the data in this experiment. There is no way to know what that true value is. Prism presents the uncertainty as a 95% confidence interval. You can be 95% sure that this interval contains the true relative risk (or odds ratio or P1-P2). When the P value is larger than 0.05, the 95% confidence interval includes the null hypothesis (relative risk or odds ratio equal to 1.0 or P1-P2 equal

to zero) and extends from a negative association ($RR < 1.0$, $OR < 1.0$, or $P1-P2 < 0.0$) to a positive association ($RR > 1.0$, $OR > 1.0$, or $P1-P2 > 0.0$)

To interpret the results in a scientific context, look at both ends of the confidence interval and ask whether they represent an association that would be scientifically important or scientifically trivial.

Lower confidence limit	Upper confidence limit	Conclusion
Trivial	Trivial	You can reach a crisp conclusion. Either there is no association between rows and columns, or it is trivial. At most, the true association between rows and columns is tiny and uninteresting.
Trivial	Large	You can't reach a strong conclusion. The data are consistent with the treatment causing a trivial negative association, no association, or a large positive association. To reach a clear conclusion, you need to repeat the experiment with more subjects.
Large	Trivial	You can't reach a strong conclusion. The data are consistent with a trivial positive association, no association, or a large negative association. You can't make a clear conclusion without repeating the experiment with more subjects.
Large	Large	You can't reach any conclusion at all. You need more data.

Checklist. Are contingency table analyses appropriate for your data?

Before interpreting the results of any statistical test, first think carefully about whether you have chosen an appropriate test. Before accepting results from a chi-square or Fisher's test, ask yourself these questions:

Question	Discussion
Are the subjects independent?	The results of a chi-square or Fisher's test only make sense if each subject (or experimental unit) is independent of the rest. That means that any factor that affects the outcome of one subject only affects that one subject. Prism cannot test this assumption. You must think about the experimental design. For example, suppose that the rows of the table represent two different kinds of preoperative antibiotics and the columns denote whether or not there was a postoperative infection. There are 100 subjects. These subjects are not independent if the table combines results from 50 subjects in one hospital with 50 subjects from another hospital. Any difference between hospitals, or the patient groups they serve, would affect half the subjects but not the other half. You do not have 100 independent observations. To analyze this kind of data, use the Mantel-Haenszel test or logistic regression. Neither of these tests are offered by Prism.
Are the data unpaired?	In some experiments, subjects are matched for age and other variables. One subject in each pair receives one treatment while the other subject gets the other treatment. These data should be analyzed by special methods such as McNemar's test (which Prism does not do, but GraphPad's StatMate program does). Paired data should not be analyzed by chi-square or Fisher's test.
Is your table really a contingency table?	To be a true contingency table, each value must represent numbers of subjects (or experimental units). If it tabulates averages, percentages, ratios, normalized values, etc. then it is not a contingency table and the results of chi-square or Fisher's tests will not be meaningful.
Does your table contain only data?	The chi-square test is not only used for analyzing contingency tables. It can also be used to compare the observed number of subjects in each category with the number you expect to see based on theory. Prism cannot do this kind of chi-square test. It is not correct to enter observed values in one column and expected in another. When analyzing a contingency table with the chi-square test, Prism generates the expected values from the data – you do not enter them.
Are the rows or columns arranged in a natural order?	If your table has two columns and more than two rows (or two rows and more than two columns), Prism will perform the chi-square test for trend as well as the regular chi-square test. The results of the test for trend will only be meaningful if the rows (or columns) are arranged in a natural order, such as age, duration, or time. Otherwise, ignore the results of the chi-square test for trend and only consider the results of the regular chi-square test.

The confidence interval of a proportion

When an experiment has two possible outcomes, the results are expressed as a proportion. Out of N experiments (or subjects), you observed one outcome (termed "success") in S experiments (or subjects) and the alternative outcome in N-S experiments. Success occurred in S/N of the experiments (or subjects), and we will call that proportion p. Since your data are derived from random sampling, the true proportion of success in the overall population is almost certainly not p. A 95% confidence interval quantifies the uncertainty. You can be 95% sure the overall proportion of success is within the confidence interval.

How to compute the 95% CI of a proportion

Prism does not compute the confidence interval of a single proportion, but does compute the confidence interval of two proportions when analyzing a 2x2 contingency table. Prism's companion program StatMate computes a confidence interval of a single proportion. Both programs (and many others) compute a confidence interval of a proportion using a method developed by Clopper and Pearson (*Biometrika* 26:404-413, 1934). The result is labeled an "exact" confidence interval (in contrast to the approximate intervals you can calculate conveniently by hand).

If you want to compute the 95% confidence interval by hand, most books present the Wald equation:

$$p = \frac{\# \text{ of "successes"}}{\# \text{ of experiments}} = \frac{S}{N}$$
$$\left[p - \left(1.96 \sqrt{\frac{p(1-p)}{N}} \right) \right] \text{ to } \left[p + \left(1.96 \sqrt{\frac{p(1-p)}{N}} \right) \right]$$

However, there is a better way. The Wald approximation is known to work well only with large N and proportions not too close to 0.0 or 1.0.

Computer simulations by several investigators demonstrate that the so-called exact confidence intervals are also approximations. They are wider than they need to be, and so generally give you more than 95% confidence. The discrepancy varies depending on the values of S and N. The so-called "exact" confidence intervals are not, in fact, exactly correct. For all values of S and N, you can be sure that you get at least 95% confidence, but the intervals may be wider than they need to be.

Agresti and Coull (*The American Statistician*. 52:119-126, 1998) recommend a method they term the modified Wald method. It is easy to compute by hand and is more accurate than the so-called "exact" method.

$$p' = \frac{\# \text{ "successes"} + 2}{\# \text{ of experiments} + 4} = \frac{S + 2}{N + 4}$$
$$\left[p' - \left(1.96 \sqrt{\frac{p'(1-p')}{N+4}} \right) \right] \text{ to } \left[p' + \left(1.96 \sqrt{\frac{p'(1-p')}{N+4}} \right) \right]$$

In some cases, the lower limit calculated using that equation is less than zero. If so, set the lower limit to 0.0. Similarly, if the calculated upper limit is greater than 1.0, set the upper limit to 1.0.

This method works very well. For any values of S and N, there is close to a 95% chance that it contains the true proportion. With some values of S and N, the degree of confidence can a bit less than 95%, but it is never less than 92%.

Where did the numbers 2 and 4 in the equation come from? Those values are actually z and z², where z is a critical value from the Gaussian distribution. Since 95% of all values of a normal distribution lie within 1.96 standard deviations of the mean, z = 1.96 (which we round to 2.0) for 95% confidence intervals.

Note that the confidence interval is centered on p', which is not the same as p, the proportion of experiments that were "successful". If p is less than 0.5, p' is higher than p. If p is greater than 0.5, p' is less than p. This makes sense as the confidence interval can never extend below zero or above one. So the center of the interval is between p and 0.5.

The meaning of "95% confidence" when the numerator is zero

If the numerator of a proportion is zero, the "95% confidence interval" really gives you 97.5% confidence. Here's why. When the proportion does not equal zero, we define the 95% confidence interval so that there is a 2.5% chance that the true proportion is less than the lower limit of the interval, and a 2.5% chance that the true proportion is higher than the upper limit. This leaves a 95% chance (100% - 2.5% - 2.5%) that the interval includes the true proportion. When the numerator is zero, we know that the true proportion cannot be less than zero, so we only need to compute an upper confidence limit. If we use the usual equations, we define the upper limit so that there is only a 2.5% chance that the true proportion is higher. Since the uncertainty only goes one way you'll actually have a 97.5% CI (100% - 2.5%). The advantage of this approach is consistency with CIs computed for proportions where the numerator is not zero.

If you don't care about consistency with other data, but want to really calculate a 95% CI, you can do that by computing a "90% CI". This is computed so that there is a 5% chance that the true proportion is higher

than the upper limit. If the numerator is zero, there is no chance of the proportion being less than zero, so the "90% CI" reported by StatMate (or other programs) really gives you 95% confidence (and StatMate tells you this). For the example above, StatMate says that the "90% confidence interval" for a proportion with the numerator = 0 and the denominator = 41 extends from 0.00% to 7.04%.

A shortcut equation for a confidence interval when the numerator equals zero

JA Hanley and A Lippman-Hand (J. Am. Med. Assoc., 249: 1743-1745, 1983) devised a simple shortcut equation for estimating the 95% confidence interval. If you observe zero events in N trials, you can be 95% sure that the true rate is less than $3/N$. To compute the usual "95% confidence interval" (which really gives you 97.5% confidence), estimate the upper limit as $3.5/N$. This equation is so simple, you can do it by hand in a few seconds.

Here is an example. You observe 0 dead cells in 10 cells you examined. What is the 95% confidence interval for the true proportion of dead cells. The "exact 95% CI" (calculated by StatMate) is 0.00% to 30.83. The adjusted Wald equation gives a "95%" confidence interval of 0.0 to 32.61%. The shortcut equation computes upper confidence limits of 35% ($3.5/10$). With such small N, the shortcut equation overestimates the confidence limit, but it is useful as a ballpark estimate.

Another example: You have observed no adverse drug reactions in the first 250 patients treated with a new antibiotic. What is the confidence interval for the true rate of drug reactions? StatMate tells us that the true rate could be as high as 1.46% (95% CI). The shortcut equation computes the upper limits as 1.40% ($3.5/250$). The adjusted Wald equation computes the upper limit as 1.87%. With large N, the shortcut equation is reasonably exact.

Correlation

Introduction to correlation

When two variables vary together, statisticians say that there is a lot of *covariation* or *correlation*. The correlation coefficient, r , quantifies the direction and magnitude of correlation.

Correlation is not the same as linear regression, but the two are related. Linear regression finds the line that best predicts Y from X. Correlation quantifies how well X and Y vary together. In some situations, you might want to perform both calculations.

Correlation only makes sense when both X and Y variables are outcomes you measure. If you control X (i.e., time, dose, concentration), don't use correlation, use linear regression. See "Linear regression" on page 141.

Correlation calculations do not discriminate between X and Y, but rather quantify the relationship between the two variables. Linear regression does discriminate between X and Y. Linear regression finds the best line that predicts Y from X by minimizing the sum of the square of the vertical distances of the points from the regression line. The X and Y variables are not symmetrical in the regression calculations. Therefore only choose regression, rather than correlation, if you can clearly define which variable is X and which is Y.

Entering data for correlation

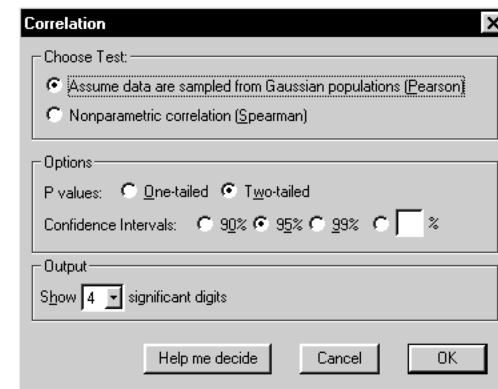
Format the data table with numbers for X and single values for Y.

If you enter Y values for several groups (into columns A, B, ...) Prism will report the correlation of X with each of the Y variables. Prism, however, cannot calculate a correlation matrix showing the correlation coefficient of each variable with each other variable.

If you format the Y columns for replicates (for example, triplicates), Prism averages each set of replicates and considers this list of averages to be Y. If you format the Y columns to enter mean and SD (or SEM) on each row, Prism analyzes only the means and ignores the SD or SEM values.

Choosing a correlation analysis

To calculate correlation, start from your data table. Click the Analyze button, select built-in analyses, then select Correlation from the list of statistical analyses. If you have entered replicate values, Prism computes the average of each set of replicates and uses the averages for correlation calculations. If you entered mean and SD (or SEM) on each row, Prism performs calculations with the mean values, but ignores the SD or SEM values.



Choose one of two correlation tests. Pearson 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 makes no assumption about the distribution of the values, as the calculations are based on ranks, not the actual values.

Prism can compute either a one-tailed or two-tailed P value. We suggest almost always choosing a two-tailed P value. You should only choose a one-tail P value when you have specified the anticipated sign of the correlation coefficient before collecting any data and are willing to attribute any correlation in the "wrong" direction to chance, no matter how large it is. See "One- vs. two-tail P values" on page 10.

Results of correlation

How correlation works

Correlation coefficient

The correlation coefficient, r , ranges from -1 to +1. The nonparametric Spearman correlation coefficient, abbreviated r_s , has the same range.

Value of r (or r_s)	Interpretation
$r = 0$	The two variables do not vary together at all.
$0 > r > -1$	The two variables tend to increase or decrease together.
$r = 1.0$	Perfect correlation.
$-1 > r > 0$	One variable increases as the other decreases.
$r = -1.0$	Perfect negative or inverse correlation.

If r or r_s 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.
- Another variable influences both X and Y.
- 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.

r^2

Perhaps the best way to interpret the value of r is to square it to calculate r^2 . Statisticians call this quantity the *coefficient of determination*, but scientists call it *r squared*. It 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 explained by variation in Y. Likewise, 59% of the variance in Y can be explained by (or goes along with) variation in X. More simply, 59% of the variance is shared between X and Y.

Prism only calculates an r^2 value from the Pearson correlation coefficient. It is not appropriate to compute r^2 from the nonparametric Spearman correlation coefficient.

P value

The P value answers this question: If there really is no correlation between X and Y in the overall population, what is the chance that random sampling

would result in a correlation coefficient as far from zero (or further) as observed in this experiment?

How to think about results of linear correlation

Look first at a graph of your data to see how X and Y vary together. Then look at the value of r (or r_s) which quantifies the correlation. Finally, look at the P value.

If the P value is small, you can reject the idea that the correlation is a coincidence. Look at the confidence interval for r . You can be 95% sure that the true population r lies somewhere within that range.

If the P value is large, the data do not give you any reason to conclude that the correlation is real. This is not the same as saying that there is no correlation at all. You just have no compelling evidence that the correlation is real and not a coincidence. Look at the confidence interval for r . It will extend from a negative correlation to a positive correlation. If the entire interval consists of values near zero that you would consider biologically trivial, then you have strong evidence that either there is no correlation in the population or that there is a weak (biologically trivial) association. On the other hand, if the confidence interval contains correlation coefficients that you would consider biologically important, then you couldn't make any strong conclusion from this experiment. To make a strong conclusion, you'll need data from a larger experiment.

Checklist. Is correlation the right analysis for these data?

To check that correlation is an appropriate analysis for these data, ask yourself these questions. Prism cannot help answer them.

Question	Discussion
Are the subjects independent?	Correlation assumes that any random factor affects only one subject, and not others. You would violate this assumption if you choose half the subjects from one group and half from another. A difference between groups would affect half the subjects and not the other half.
Are X and Y 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.
Were X values measured (not controlled)?	If you controlled X values (i.e. concentration, dose or time) you should calculate linear regression rather than correlation.
Is the covariation linear?	A correlation analysis would not be helpful if Y increases as X increases up to a point, and then Y decreases as X increases further. You might obtain a low value of r even though the two variables are strongly related. The correlation coefficient quantifies linear covariation only.
Are X and Y distributed according to Gaussian distributions?	To accept the P value from standard (Pearson) correlation, the X and Y values must each be sampled from populations that follow Gaussian distributions. Spearman nonparametric correlation does not make this assumption.

Linear regression

Introduction to linear regression

Linear regression analyzes the relationship between two variables, X and Y. For each subject (or experimental unit), you know both X and Y and you want to find the best straight line through the data. In some situations, the slope and/or intercept have a scientific meaning. In other cases, you use the linear regression line as a standard curve to find new values of X from Y, or Y from X.

The term "regression", like many statistical terms, is used in statistics quite differently than it is used in other contexts. The method was first used to examine the relationship between the heights of fathers and sons. The two were related, of course, but the slope is less than 1.0. A tall father tended to have sons shorter than himself; a short father tended to have sons taller than himself. The height of sons regressed to the mean. The term "regression" is now used for many sorts of curve fitting.

Prism determines and graphs the best-fit linear regression line, optionally including a 95% confidence interval or 95% prediction 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.

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.

Note that linear regression does not *test* whether your data are linear (except via 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.

Avoid Scatchard, Lineweaver-Burke and similar transforms

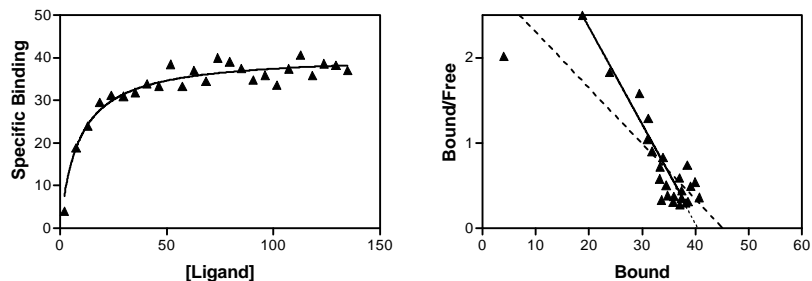
Before analyzing your data with linear regression, stop and ask yourself whether it might make more sense to fit your data with nonlinear regression. If you have transformed nonlinear data to create a linear relationship, you will probably be better off using nonlinear regression on the untransformed data.

Before nonlinear regression was readily available, the best way to analyze nonlinear data was to transform the data to create a linear graph, and then analyze 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.

The problem with these methods is that the 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 rarely true after transforming data. Furthermore, 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 assumption of linear regression that all uncertainty is in Y while X is known precisely. It doesn't make sense to minimize the sum of squares of the vertical distances of points from the line, if the same experimental error appears in both X and Y directions.

Since the assumptions of linear regression are violated, 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 technique for analyzing your data. Nonlinear regression produces the most accurate results.

This figure below 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 (see "Scatchard plots" on page 253). 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. Scatchard plots can be used to determine the receptor number (B_{\max} , determined as the X-intercept of the linear regression line) and dissociation constant (K_d , determined as the negative reciprocal of the slope). Since the Scatchard transformation amplified and distorted the scatter, the linear regression fit does not yield the most accurate values for B_{\max} and K_d .



Don't use linear regression just to avoid using nonlinear regression. Fitting curves with nonlinear regression is not difficult.

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 the results of a linear transform.

Entering data for linear regression

The data table must be formatted with numbers for the X column. You can choose one of several formats for Y columns.

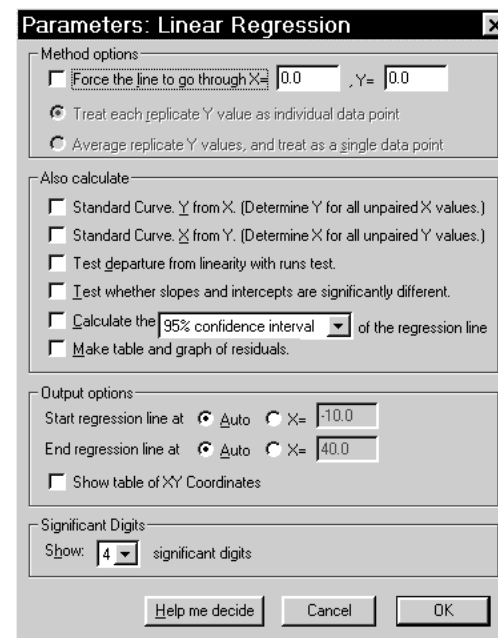
If you enter Y values for several groups (into columns A, B, ...) Prism will report the results of linear regression of X with each of the Y variables. Prism, however, cannot calculate multiple regression.

If you format the Y columns for replicates (for example, triplicates) Prism can average these and perform all calculations with the means. Or it can treat each replicate as a separate value to fit.

If you format the Y columns for entry of SD or SEM, Prism analyzes only the means and ignores the SD or SEM values.

Choosing a linear regression analysis

Start from a data table or graph (see "Entering data for linear regression" on page 143). Click on the Analyze button and choose built-in analyses. Then select Linear regression from the list of curves and regressions.



Force a regression line through the origin (or some other point)?

If you choose regression, you may force the line to go through a particular point such as the origin. In this case, Prism will determine only the best-fit slope, as the intercept will be fixed. Use this option when scientific theory tells you that the line must go through a particular point (usually the origin, $X=0, Y=0$) and you only want to know the slope. This situation arises rarely.

Use common sense when making your decision. For example, consider a protein assay. You measure optical density (Y) for several known concentrations of protein in order to create a standard curve. You then want to interpolate unknown protein concentrations from that standard curve. When performing the assay, you adjusted the spectrophotometer so that it reads zero with zero protein. Therefore you might be tempted to force the regression line through the origin. But this constraint may result in a line that doesn't fit the data so well. Since you really care that the line fits the standards very well near the unknowns, you will probably get a better fit by not constraining the line.

Most often, you should let Prism find the best-fit line without any constraints.

Fit linear regression to individual replicates or means?

If you collected replicate Y values at every value of X, there are two ways Prism can calculate linear regression. It can treat each replicate as a separate point, or average the replicate Y values, and treat the mean as a single point.

You should consider each replicate a separate point when the sources of experimental error are the same for each data point. If one value happens to be a bit high, there is no reason to expect the other replicates to be high as well. The errors are independent.

Average the replicates and treat the mean as a single value when the replicates are not independent. For examples, the replicates would not be independent if they represent triplicate measurements from the same animal, with a different animal used at each value of X (dose). If one animal happens to respond more than the others, that will affect all the replicates. The replicates are not independent.

For more examples, see "Replicates" on page 202.

Additional calculations with linear regression

Prism gives you several choices for performing additional calculations with linear regression. Read more about these choices:

- Reading unknowns from standard curves [on page 329](#).
- Runs test following linear regression [on page 150](#).
- Comparing slopes and intercepts [on page 150](#).
- Confidence or prediction interval of a regression line [on page 148](#).
- Residuals from a linear regression line [on page 149](#).

Output options for linear regression

By default Prism draws the regression line from the smallest X value in your data to the largest. Uncheck "auto" to enter the starting or stopping X values. This affects how the line is graphed, but does not affect any of the numerical results.

Results of linear regression

How linear regression works

Minimizing sum-of-squares

The goal of linear regression is to adjust the values of slope and intercept to find the line that best predicts Y from X. More precisely, the goal of

regression is to minimize the sum of the squares of the vertical distances of the points from the line. Why minimize the sum of the squares of the distances? Why not simply minimize the sum of the actual distances?

If the random scatter follows a Gaussian distribution, it is far more likely to have two medium size deviations (say 5 units each) than to have one small deviation (1 unit) and one large (9 units). A procedure that minimized the sum of the absolute value of the distances would have no preference over a line that was 5 units away from two points and one that was 1 unit away from one point and 9 units from another. The sum of the distances (more precisely, the sum of the absolute value of the distances) is 10 units in each case. A procedure that minimizes the sum of the squares of the distances prefers to be 5 units away from two points (sum-of-squares = 25) rather than 1 unit away from one point and 9 units away from another (sum-of-squares = 82). If the scatter is Gaussian (or nearly so), the line determined by minimizing the sum-of-squares is most likely to be correct.

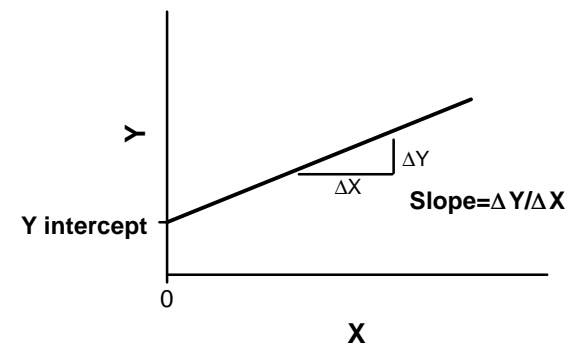
The calculations are shown in every statistics book, and are entirely standard.

Slope and intercept

Prism reports the best-fit values of the slope and intercept, along with their standard errors and confidence intervals.

The slope quantifies the steepness of the line. It equals the change in Y for each unit change in X. It is expressed in the units of the Y-axis divided by the units of the X-axis. If the slope is positive, Y increases as X increases. If the slope is negative, Y decreases as X increases.

The Y intercept is the Y value of the line when X equals zero. It defines the elevation of the line.

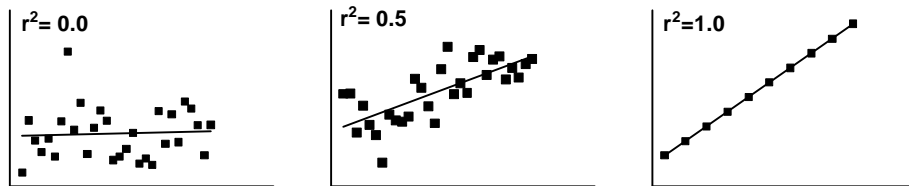


The standard error values of the slope and intercept can be hard to interpret, but their main purpose is to compute the 95% confidence intervals. If you accept the assumptions of linear regression, there is a 95% chance that the 95% confidence interval of the slope contains the true

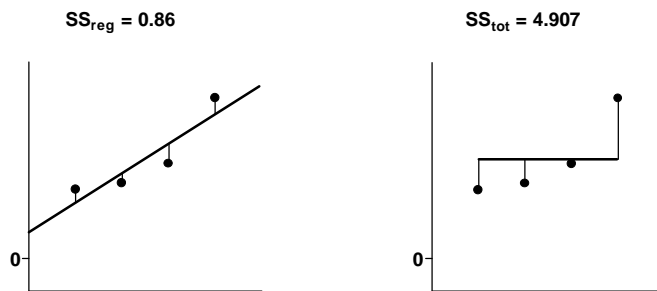
value of the slope, and that the 95% confidence interval for the intercept contains the true value of the intercept.

r^2 , a measure of goodness-of-fit of linear regression

The value r^2 is a fraction between 0.0 and 1.0, and has no units. An r^2 value of 0.0 means that knowing X does not help you predict Y. There is no linear relationship between X and Y, and the best-fit line is a horizontal line going through the mean of all Y values. When r^2 equals 1.0, all points lie exactly on a straight line with no scatter. Knowing X lets you predict Y perfectly.



This figure demonstrates how Prism computes r^2 .



$$r^2 = 1 - \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{0.86}{4.91} = 0.84$$

The left panel shows the best-fit linear regression line. This line minimizes the sum-of-squares of the vertical distances of the points from the line. Those vertical distances are also shown on the left panel of the figure. In this example, the sum of squares of those distances (SS_{reg}) equals 0.86. Its units are the units of the Y-axis squared. To use this value as a measure of goodness-of-fit, you must compare it to something.

The right half of the figure shows the null hypothesis – a horizontal line through the mean of all the Y values. Goodness-of-fit of this model (SS_{tot}) is also calculated as the sum of squares of the vertical distances of the points

from the line, 4.907 in this example. The ratio of the two sum-of-squares values compares the regression model with the null hypothesis model. The equation to compute r^2 is shown in the figure. In this example r^2 is 0.8428. The regression model fits the data much better than the null hypothesis, so SS_{reg} is much smaller than SS_{tot} , and r^2 is near 1.0. If the regression model were not much better than the null hypothesis, r^2 would be near zero.

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. In other words, r^2 is the fraction of the variation that is shared between X and Y.

In this example, 84% of the total variance in Y is “explained” by the linear regression model. The variance (SS) of the data from the linear regression model equals only 16% of the total variance of the Y values (SS_{tot})

Why Prism doesn't report r^2 in constrained linear regression

Prism does not report r^2 when you force the line through the origin (or any other point), because the calculations would be ambiguous. There are two ways to compute r^2 when the regression line is constrained. As you saw in the previous section, r^2 is computed by comparing the sum-of-squares from the regression line with the sum-of-squares from a model defined by the null hypothesis. With constrained regression, there are two possible null hypotheses. One is a horizontal line through the mean of all Y values. But this line doesn't follow the constraint – it does not go through the origin. The other null hypothesis would be a horizontal line through the origin, far from most of the data.

Because r^2 is ambiguous in constrained linear regression, Prism doesn't report it. If you really want to know a value for r^2 , use nonlinear regression to fit your data to the equation $Y = \text{slope} * X$. Prism will report r^2 defined the first way (comparing regression sum-of-squares to the sum-of-squares from a horizontal line at the mean Y value).

The standard deviation of the residuals, $s_{y,x}$

Prism doesn't actually report the sum-of-squares of the vertical distances of the points from the line (SS_{reg}). Instead Prism reports the standard deviation of the residuals, $s_{y,x}$.

The variable $s_{y,x}$ quantifies the average size of the residuals, expressed in the same units as Y. Some books and programs refer to this value as s_e . It is calculated from SS_{reg} and N (number of points) using this equation:

$$s_{y,x} = \sqrt{\frac{SS_{reg}}{N-2}}$$

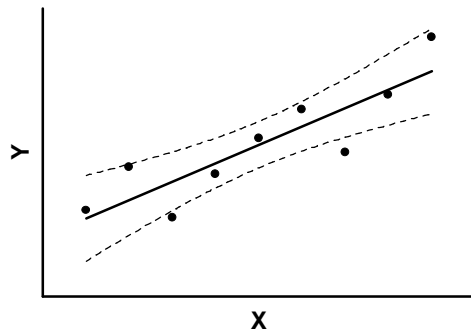
Is the slope significantly different than zero?

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 (or further) than you observed? The P value is calculated from an F test, and Prism also reports the value of F and its degrees of freedom.

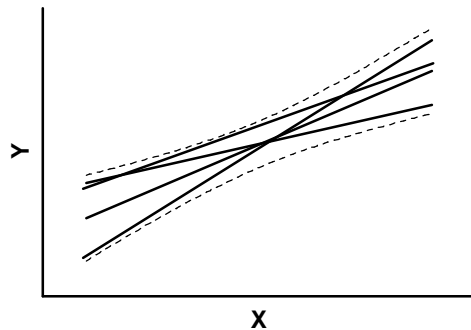
Additional calculations following linear regression

Confidence or prediction interval of a regression line

If you check the option box, Prism will calculate and graph either the 95% confidence interval or 95% prediction interval of the regression line. Two curves surrounding the best-fit line define the confidence interval.



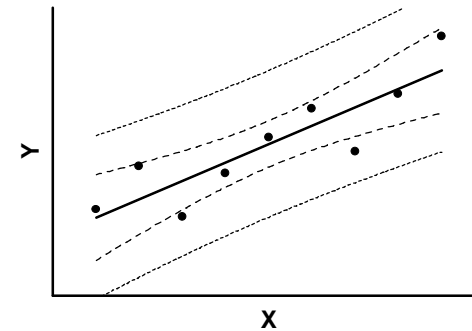
The dashed lines that demarcate the confidence interval are curved. This does not mean that the confidence interval includes the possibility of curves as well as straight lines. Rather, the curved lines are the boundaries of all possible straight lines. The figure below shows four possible linear regression lines (solid) that lie within the confidence interval (dashed).



Given the assumptions of linear regression, you can be 95% confident that the two curved confidence bands enclose the true best-fit linear regression line, leaving a 5% chance that the true line is outside those boundaries.

Many data points will be outside the 95% confidence interval boundary. The confidence interval is 95% sure to contain the best-fit regression line. This is not the same as saying it will contain 95% of the data points.

Prism can also plot the 95% *prediction interval*. The prediction bands are further from the best-fit line than the confidence bands, a lot further if you have many data points. The 95% prediction interval is the area in which you expect 95% of all data points to fall. In contrast, the 95% confidence interval is the area that has a 95% chance of containing the true regression line. This graph shows both prediction and confidence intervals (the curves defining the prediction intervals are further from the regression line).



Residuals from a linear regression line

Residuals are the vertical distances of each point from the regression line. The X values in the residual table are identical to the X values you entered. The Y values are the residuals. A residual with a positive value means that the point is above the line; a residual with 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. For an example, see "Residuals" on page 214.

Runs test following linear regression

The runs test determines whether your data differ significantly from a straight line. Prism can only calculate the runs test if you entered the X values in order.

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 consecutive series of points whose residuals are either all positive or all negative.

If the data points are randomly distributed above and below the regression line, it is possible to calculate the expected number of runs. If there are N_a points above the curve and N_b points below the curve, the number of runs you expect to see equals $[(2N_a N_b)/(N_a + N_b)] + 1$. If you observe fewer runs than expected, it may be a coincidence of random sampling or it may mean that your data deviate systematically from a straight line. The P value from the runs test answers this question: If the data really follow a straight line, what is the chance that you would obtain as few (or fewer) runs as observed in this experiment?

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, conclude that the data don't really follow a straight line, and consider using nonlinear regression to fit a curve.

Comparing slopes and intercepts

Prism can test whether the slopes and intercepts of two or more data sets are significantly different. It compares linear regression lines using the method explained in Chapter 18 of J Zar, *Biostatistical Analysis*, 2nd edition, Prentice-Hall, 1984.

Prism compares slopes first. It calculates a P value (two-tailed) testing the null hypothesis that the slopes are all identical (the lines are parallel). The P value answers this question: If the slopes really were identical, what is the chance that randomly selected data points would have slopes as different (or more different) than you 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 intersection point of two lines is:

$$X = \frac{\text{Intercept}_1 - \text{Intercept}_2}{\text{Slope}_2 - \text{Slope}_1}$$

$$Y = \text{Intercept}_1 + \text{Slope}_1 \cdot X = \text{Intercept}_2 + \text{Slope}_2 \cdot X$$

If the P value for comparing slopes is greater than 0.05, Prism concludes that the slopes are not significantly different and calculates a single slope for all the lines. Now the question is whether the lines are parallel or

identical. Prism calculates a second P value testing the null hypothesis that the lines are identical. If this P value is low, conclude that the lines are not identical (they are distinct but parallel). If this second P value is high, there is no compelling evidence that the lines are different.

This method is equivalent to an Analysis of Covariance (ANCOVA), although ANCOVA can be extended to more complicated situations.

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 329.

How to think about the results of linear regression

Your approach to linear regression will depend on your goals.

If your goal is to analyze a standard curve, you won't be very interested in most of the results. Just make sure that r^2 is high and that the line goes near the points. Then go straight to the standard curve results.

In many situations, you will be most interested in the best-fit values for slope and intercept. Don't just look at the best-fit values, also look at the 95% confidence interval of the slope and intercept. If the intervals are too wide, repeat the experiment with more data.

If you forced the line through a particular point, look carefully at the graph of the data and best-fit line to make sure you picked an appropriate point.

Consider whether a linear model is appropriate for your data. Do the data seem linear? Is the P value for the runs test high? Are the residuals random? If you answered no to any of those questions, consider whether it makes sense to use nonlinear regression instead.

Checklist. Is linear regression the right analysis for these data?

To check that linear regression is an appropriate analysis for these data, ask yourself these questions. Prism cannot help answer them.

Question	Discussion
Can the relationship between X and Y be graphed as a straight line?	In many experiments the relationship between X and Y is curved, making linear regression inappropriate. Either transform the data, or use a program (such as GraphPad Prism) that can perform nonlinear curve fitting.
Is the scatter of data around the line Gaussian (at least approximately)?	Linear regression analysis assumes that the scatter is Gaussian.
Is the variability the same everywhere?	Linear regression assumes that scatter of points around the best-fit line has the same standard deviation all along the curve. The assumption is violated if the points with high or low X values tend to be further from the best-fit line. The assumption that the standard deviation is the same everywhere is termed <i>homoscedasticity</i> .
Do you know the X values precisely?	The linear regression model assumes that X values are exactly correct, and that experimental error or biological variability only affects the Y values. 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.
Are the data points independent?	Whether one point is above or below the line is a matter of chance, and does not influence whether another point is above or below the line. See "Fit linear regression to individual replicates or means?" on page 145.
Are the X and Y values intertwined?	If the value of X is used to calculate Y (or the value of Y is used to calculate X) then linear regression calculations are invalid. One example is a Scatchard plot, where the Y value (bound/free) is calculated from the X value. See "Avoid Scatchard, Lineweaver-Burke and similar transforms" on page 142. Another example would be a graph of midterm exam scores (X) vs. total course grades (Y). Since the midterm exam score is a component of the total course grade, linear regression is not valid for these data.

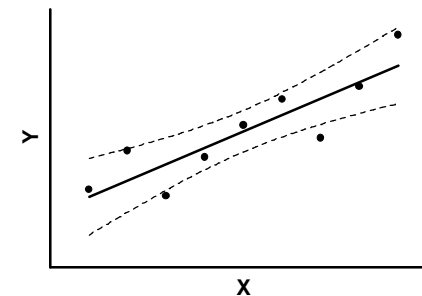
Graphing linear regression

When you fit a linear regression line with Prism, it automatically adds the best-fit line to a graph of the data. The best-fit line is simply a data set in the results sheet, and you can add this data set to other graphs or remove it from the graph where Prism placed it.

If you checked the option to create 95% confidence or prediction intervals, Prism includes those as error bars on the results sheet.

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.



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.

Fitting linear data with nonlinear regression

Prism's linear regression analysis does not have quite as much versatility as the nonlinear regression analysis. Since Prism's nonlinear analysis can fit data to any model, even a linear one, use the nonlinear regression analysis

when needed to take advantage of the increased power. For example, you could use the nonlinear regression analysis to compare an unconstrained linear regression line with one forced through the origin (or some other point). Or you could fit a linear regression line, but weight the points to minimize the sum of the relative distance squared, rather than the distance squared.

To fit linear data using Prism's nonlinear regression analysis:

1. From the data table or graph, click Analyze and choose nonlinear regression.
2. Choose the First order polynomial equation, which defines a straight line. Or choose one of the equations listed below from the equation library (or enter them yourself).
3. Click the Methods button to adjust weighting method or other options. Click the Output button to create a summary table.

Determining a point other than the Y intercept

The linear regression equation defines Y as a function of slope and Y intercept:

$$Y = \text{intercept} + \text{slope} \cdot X$$

This equation can be rewritten in a more general form, where we replace the intercept (Y value at X=0) with $Y_{X'}$, the Y value at X' , where X' is some specified X value. Now the linear regression equation becomes:

$$Y = Y_{X'} + \text{slope} \cdot (X - X')$$

Y at any particular X value equals the Y value at X' plus the slope times the distance from X to X' . If you set X' to zero, this equation becomes identical to the previous one.

This equation has three variables, $Y_{X'}$, X' , and slope. If you set either X' or $Y_{X'}$ to a constant value, you can fit the other (and the slope).

When you fit these equations using nonlinear regression, Prism insists that you enter rules for initial values (or directly enter initial values). Since the equation is in fact linear, Prism will find the best-fit values no matter what initial values you enter. Setting all initial values to zero should work fine.

Note. When you enter your equation, you must define either X' or $Y_{X'}$. The two are related, so it is impossible to fit both.

Example 1. You want to find the X value (with SE and confidence interval) where $Y = 50$ for linear data. Fit the data using nonlinear regression using this equation to determine the best-fit values of the slope and X_{50} .

$$Y = 50 + \text{slope} \cdot (X - X_{50})$$

In this equation, X_{50} is the value of X where $Y = 50$. Don't confuse this with an EC_{50} , which is the value of X where Y is halfway between minimum and maximum in a dose-response curve. A linear model does not have a minimum and maximum, so the concept of EC_{50} does not apply.

Example 2. You want to fit data to a straight line to determine the slope and the Y value at $X = 20$. Fit the data using nonlinear regression to this equation.

$$Y = Y_{20} + \text{slope} \cdot (X - 20)$$

Example 3. You want to determine the slope and the X intercept, with SE and confidence interval. Fit the data using nonlinear regression to this equation:

$$Y = \text{slope} \cdot (X - X_{\text{intercept}})$$

Fitting two line segments

Prism can fit two line segments to different portions of the data, so that the two meet at a defined X value. See "Example 3. Two linear regression segments that meet at the breakpoint" on page 192.

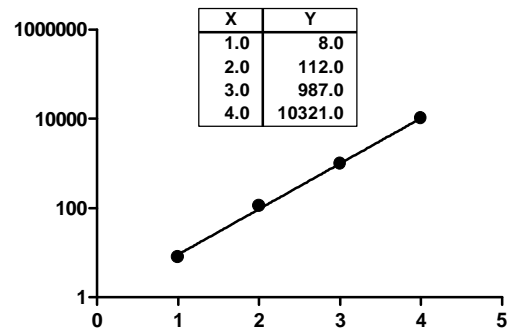
Fitting straight lines to semi-log graphs

You can start linear (or nonlinear) regression from a data table, results table or graph. Prism then fits a model to the data. If you start from a graph, Prism fits to the data plotted on that graph. Selecting a logarithmic axis (from the Axis dialog) does not change the data, so does not change the way Prism performs regression. If you plot a linear regression "line" on a graph with a logarithmic axis, the best-fit "line" will be curved. To fit a straight line on a semilog plot requires use of nonlinear regression.

If the Y-axis is logarithmic use this equation.

$$Y = 10^{(\text{Slope} \cdot X + Y_{\text{intercept}})}$$

Graphing on a log Y-axis is equivalent to taking the antilog. The antilog of the right side of the equation is the equation for a straight line: $\text{slope} \cdot X + Y_{\text{intercept}}$. So this equation appears linear when graphed on a logarithmic Y-axis. It is difficult to define rules for initial values that work for all data sets, so you'll need to enter initial values individually for each data set. Here is an example.



If the X-axis is logarithmic, then use this equation:

$$Y = \text{Slope} \cdot 10^X + \text{Yintercept}$$

If both axes are logarithmic (rare) then use this equation:

$$Y = 10^{(\text{Slope} \cdot 10^X + \text{Yintercept})}$$

Introducing curve fitting and nonlinear regression

Introduction to models

A mathematical model is a description of a physical, chemical or biological state or process. Using a model can help you think about chemical and physiological processes or mechanisms, so you can design better experiments and comprehend the results. When you fit a model to your data, you obtain best-fit values that you can interpret in the context of the model.

"A mathematical model is neither an hypothesis nor a theory. Unlike scientific hypotheses, a model is not verifiable directly by an experiment. For all models are both true and false... The validation of a model is not that it is "true" but that it generates good testable hypotheses relevant to important problems. " (R. Levins, Am. Scientist 54:421-31, 1966)

When you choose linear regression, you may not be thinking about a model. But linear regression fits data to the following model, which describes a linear relationship between X and Y.

$$Y = \text{Slope} \cdot X + \text{Intercept}$$

That is a special model that can be fit to data using very simple math. Other models require more difficult calculations, but the idea is the same.

In most circumstances, you'll be able to use standard models developed by others. Three examples follow. You will only need to develop new models if you work with new experimental systems, or need to extend conventional models to new situations.

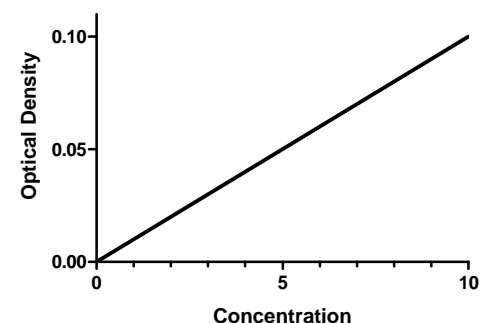
How models are derived

Models are derived using simple logic and algebra, along (sometimes) with some calculus. Here are three examples of commonly used models, with derivations.

Example model 1. Optical density as a function of concentration

Colorimetric chemical assays are based on a simple principle. Add appropriate reactants to your samples to initiate a chemical reaction whose product is colored. When you terminate the reaction, the concentration of colored product is proportional to the initial concentration of the substance you want to assay. Since optical density is proportional to the concentration of colored substances, the optical density will also be proportional to the concentration of the substance you are assaying.

$$\text{Optical Density} = Y = k \cdot [\text{substance}] = K \cdot X$$



Mathematically, the equation works for any value of X. However, the results only make sense with certain values. Negative X values are meaningless, as concentrations cannot be negative. The model may fail at high concentrations of substance where the reaction is no longer limited by the concentration of substance. The model may also fail at high concentrations if the solution becomes so dark (the optical density is so high) that little light reaches the detector. At that point, the noise of the instrument may exceed the signal. It is not unusual that a model works only for a certain range of values. You just have to be aware of the limitations, and not try to use the model outside of its useful range.

Example model 2. Exponential decay

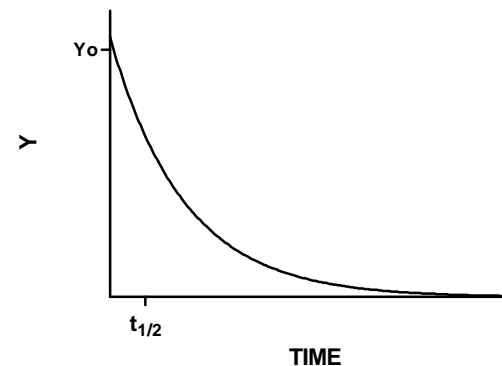
Exponential equations are used to model many processes. They are used whenever the rate at which something happens is proportional to the amount which is left. Here are three examples:

- When ligands dissociate from receptors, the number of molecules that dissociate in any short time interval is proportional to the number that were bound at the beginning of that interval. Equivalently, each individual molecule of ligand bound to a receptor has a certain probability of dissociating from the receptor in any small time interval. That probability does not get higher as the ligand stays on the receptor longer.
- When radioactive isotopes decay, the number of atoms that decay in any short interval is proportional to the number of undecayed atoms that were present at the beginning of the interval. This means that each individual atom has a certain probability of decaying in a small time interval, and that probability is constant. The probability that any particular atom will decay does not change over time. The total decay of the sample decreases with time because there are fewer and fewer undecayed atoms.
- When drugs are metabolized by the liver or excreted by the kidney, the rate of metabolism or excretion is often proportional to the concentration of drug in the blood plasma. Each drug molecule has a certain probability of being metabolized or secreted in a small time interval. As the drug concentration goes down, the rate of its metabolism or excretion goes down as well.

Define Y to be the number of ligand-receptor complexes still present (or the number of radioactive atoms that have not yet decayed, or the concentration of drug in the plasma) at any given time X . The rate of change of Y is proportional to Y . Expressed as a differential equation:

$$\frac{\Delta Y}{\Delta X} = \frac{dY}{dX} = -k \cdot Y$$

Shown as a graph:



Like most nonlinear regression programs, Prism doesn't let you enter a model expressed as a differential equation. Instead, you must enter the equation defining Y as a function of X . To do this, you need to remember a bit of calculus. There is only one function whose derivative is proportional to Y , the exponential function. Integrate both sides of the equation to obtain a new exponential equation that defines Y as a function of X , the rate constant k , and the value of Y at time zero, Y_0 .

$$Y = Y_0 \cdot e^{-k \cdot X} = Y_0 \cdot \exp(-k \cdot X)$$

The half-life is the time it takes for Y to drop by 50%. To find the half-life, set Y equal to one-half of Y_0 and solve the above equation for X . It equals the natural logarithm of 2 divided by the rate constant.

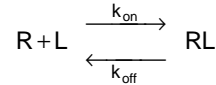
$$\text{Half life} = T_{1/2} = \frac{\ln(2)}{k} = \frac{0.693}{k}$$

In the case of radioactive decay, this model describes exactly what is going on physically. In the case of ligand binding, the model may be a simplification. The model considers that all receptors are either free or bound to ligand. In fact, binding is a complicated process with multiple points of contact between ligand and receptor, so there must be some states of partial binding. Even though the model is simplified, it predicts experimental data very well. Even very simple models can adequately predict the behavior of very complicated systems, and can yield constants (dissociation rate constant in this example) that have a physical meaning. Few models describe a physical process exactly. Models that simplify the true molecular or physiological mechanisms can be very useful, so long as they are not too simple.

Example model 3. Equilibrium binding

This example derives a very common model that describes equilibrium binding (or enzyme kinetics). Deriving this model does not require any calculus!

When a ligand interacts with a receptor, or when a substrate interacts with an enzyme, the binding follows the law of mass action.



In this equation R is the concentration of free receptor, L is the concentration of free ligand, and RL is the concentration of receptor ligand complex. In the case of enzyme kinetics, R is the enzyme and L is the substrate.

The association rate constant k_{on} is expressed in units of $M^{-1}min^{-1}$. The rate of RL formation equals $R \cdot L \cdot k_{on}$. The dissociation constant k_{off} is expressed in units of min^{-1} . The rate of RL dissociation equals $RL \cdot k_{off}$. At equilibrium, the backward (dissociation) reaction equals the forward (association) reaction so,

$$RL \cdot k_{off} = R \cdot L \cdot k_{on}$$

Binding studies measure specific binding, which is a measure of RL. Enzyme kinetic assays assess enzyme velocity, which is proportional to RL, the concentration of enzyme-substrate complexes. So you want to arrange the equation to obtain RL on the left.

$$RL = R \cdot L \cdot \frac{k_{on}}{k_{off}}$$

Define the equilibrium dissociation constant, K_d to equal k_{off}/k_{on} , which is in molar units. In enzyme kinetics, this is called the Michaelis-Menten constant, K_M . Rearrange the previous equation to define the concentration of receptor-ligand complexes at equilibrium:

$$RL = R \cdot L \cdot \frac{k_{on}}{k_{off}} = \frac{R \cdot L}{K_d}$$

Since you usually can't measure the free concentration of receptor, R, the equation won't be useful until that variable is removed. Fortunately, that's

easy to do. Since all receptors are either free or bound, we can express R as the total number of receptors minus the number bound: $R = R_{tot} - RL$.

Substitute this definitions of R into the previous equation.

$$RL = \frac{R \cdot L}{K_d} = \frac{(R_{tot} - RL) \cdot L}{K_d} = \frac{R_{tot} \cdot L - RL \cdot L}{K_d}$$

RL now appears on both sides of the equation. Rearrange to keep RL on the left.

$$\begin{aligned} RL \cdot K_d &= R_{tot} \cdot L - RL \cdot L \\ RL(K_d + L) &= R_{tot} \cdot L \\ RL &= \frac{R_{tot} \cdot L}{K_d + L} \end{aligned}$$

Since we vary L and measure RL, define Y to be RL (amount of specific binding, or enzyme activity) and X to be L (concentration of ligand or substrate). Finally call the total amount of binding B_{max} (instead of R_{tot}). The equation for equilibrium binding now becomes:

$$\text{Specific binding} = Y = \frac{B_{max} \cdot X}{K_d + X}$$

The graph of this equation (left panel below) is sometimes called a rectangular hyperbola or a binding isotherm. If you plot the same data on a semilog plot (the X-axis is log of ligand concentration) it becomes sigmoidal. The only difference between the left and right panel of the graph is whether the X-axis is linear or logarithmic.



The equation of enzyme velocity as a function of substrate concentration is identical except for the names of the variables.

$$\text{Enzyme Activity} = Y = \frac{V_{\max} \cdot X}{K_M + X}$$

Why Prism can't pick a model for you

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 Software) 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 approach can be useful in some situations (when you want a smooth curve, but don't care about models). Don't use this approach 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 as a way to avoid understanding your experimental system, or to avoid making scientific decisions.

Introduction to nonlinear regression

The difference between linear and nonlinear regression

Linear regression is described in every statistics book, and is performed by every statistics program. Nonlinear regression is mentioned in only a few books, and is not performed by all statistics programs. From a mathematician's point of view, the two procedures are vastly different. From a scientist's point of view, however, the two procedures are very similar. In many fields of science, nonlinear regression is used far more often than linear 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 equations used to do this can be derived with no more than high-school algebra (shown in many statistics books). Put the data in, and the answers come out. There is no chance for ambiguity. You could even do the calculations by hand, if you wanted to.

Nonlinear regression is more general than linear regression. It fits data to any equation that defines Y as a function of X and one or more parameters. It finds the values of those parameters that generate the curve that comes closest to the data. More precisely, nonlinear regression finds the values of the parameters that generates a curve that minimizes the sum of the squares of the vertical distances of the data points from the curve.

Except for a few special cases, it is not possible to directly derive an equation to compute the best-fit values from the data. Instead nonlinear regression requires a computationally intensive, iterative approach. You can't really follow the mathematics of nonlinear regression unless you are familiar with matrix algebra. But these complexities only pertain to performing the calculations, something Prism does for you. Using nonlinear regression to analyze data is only slightly more difficult than using linear regression. Your choice of linear or nonlinear regression should be based on the model you are fitting. Don't use linear regression just to avoid using nonlinear regression.

Sum-of-squares as a measure of goodness-of-fit in nonlinear regression

The goal of nonlinear regression is to adjust the values of the variables in the model to find the curve that best predicts Y from X. More precisely, the goal of regression is to minimize the sum of the squares of the vertical distances of the points from the curve. Why minimize the sum of the squares of the distances? Why not simply minimize the sum of the actual distances?

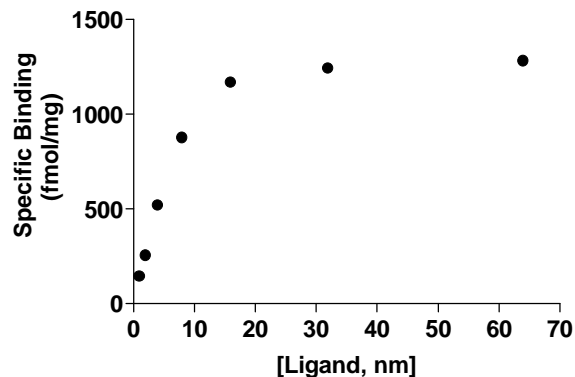
If the random scatter follows a Gaussian distribution, it is far more likely to have two medium size deviations (say 5 units each) than to have one small deviation (1 unit) and one large (9 units). A procedure that minimized the sum of the absolute value of the distances would have no preference over a curve that was 5 units away from two points and one that was 1 unit away from one point and 9 units from another. The sum of the distances (more precisely, the sum of the absolute value of the distances) is 10 units in each case. A procedure that minimizes the sum of the squares of the distances prefers to be 5 units away from two points (sum-of-squares = 25) rather than 1 unit away from one point and 9 units away from another (sum-of-squares = 82). If the scatter is Gaussian (or nearly so), the curve determined by minimizing the sum-of-squares is most likely to be correct.

How nonlinear regression works

You won't be able to understand the mathematical details of nonlinear regression unless you first master matrix algebra. But the basic idea is pretty easy to understand. Every nonlinear regression method follows these steps:

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). See "Sum-of-squares as a measure of goodness-of-fit in nonlinear regression" on page 163.
3. Adjust the variables to make the curve come closer to the data points. There are several algorithms for adjusting the variables, as explained below.
4. Adjust the variables again so that the curve comes even closer to the points. Repeat.
5. Stop the calculations when the adjustments make virtually no difference in the sum-of-squares. See "Convergence criteria" on page 203.
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.

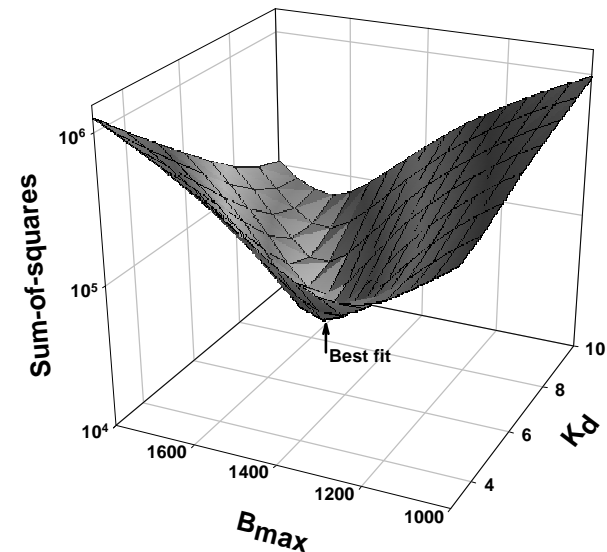
Step 3 is the only difficult one. Prism (and most other nonlinear regression programs) uses the method of Marquardt and Levenberg, which blends two other methods, the method of linear descent and the method of Gauss-Newton. The best way to understand these methods is to follow an example. Here are some data to be fit to a typical binding curve (rectangular hyperbola).



You want to fit a binding curve to determine B_{\max} and K_d using the equation

$$Y = \frac{B_{\max} \cdot X}{K_d + X}$$

How can you find the values of B_{\max} and K_d that fit the data best? You can generate an infinite number of curves by varying B_{\max} and K_d . For each of the generated curves, you can compute the sum-of-squares to assess how well that curve fits the data. The following graph illustrates the situation.



The X- and Y-axes correspond to two variables to be fit by nonlinear regression (B_{\max} and K_d in this example). The Z-axis is the sum-of-squares. Each point on the surface corresponds to one possible curve. The goal of nonlinear regression is to find the values of B_{\max} and K_d that make the sum-of-squares as small as possible (to find the bottom of the valley).

The **method of linear descent** follows a very simple strategy. Starting from the initial values try increasing each parameter a small amount. If the sum-of-squares goes down, continue. If the sum-of-squares goes up, go back and decrease the value of the parameter instead. You've taken a step down the surface. Repeat many times. Each step will usually reduce the sum-of-squares. If the sum-of-squares goes up instead, the step must have been so large that you went past the bottom and back up the other side. If this happens, go back and take a smaller step. After repeating these steps many time, you'll reach the bottom.

The **Gauss-Newton method** is a bit harder to understand. As with the method of linear descent, start by computing how much the sum-of-squares changes when you make a small change in the value of each parameter.

This tells you the slope of the sum-of-squares surface at the point defined by the initial values. If the equation really were linear, this is enough information to determine the shape of the entire sum-of-squares surface, and thus calculate the best-fit values of B_{\max} and K_d in one step. With a linear equation, knowing the slope at one point tells you everything you need to know about the surface, and you can find the minimum in one step. With nonlinear equations, the Gauss-Newton method won't find the best-fit values in one step, but that step usually improves the fit. After repeating many iterations, you reach the bottom.

This method of linear descent tends to work well for early iterations, but works slowly when it gets close to the best-fit values (and the surface is nearly flat). In contrast, the Gauss-Newton method tends to work badly in early iterations, but works very well in later iterations. The two methods are blended in the method of **Marquardt** (also called the **Levenberg-Marquardt method**). It uses the method of linear descent in early iterations and then gradually switches to the Gauss-Newton approach. Prism, like most programs, uses the Marquardt method for performing nonlinear regression.

To learn more about how nonlinear regression works, read Chapter 15 of [Numerical Recipes in C](#), Second Edition, WH Press, et. Al. , Cambridge Press, 1992 or Chapter 10 of [Primer of Applied Regression and Analysis of Variance](#) by SA Glantz and BK Slinker, McGraw-Hill, 1990.

Other kinds of regression

The difference between correlation and regression

Correlation quantifies how consistently the two variables 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 . Correlation calculations do not find a best-fit line. See "Correlation" on page 135

Polynomial regression

Polynomial regression fits data to this equation:

$$Y = A + B \cdot X + C \cdot X^2 + D \cdot X^3 + E \cdot X^4 \dots$$

You can include any number of terms. If you stop at the second (B) term, it is called a first-order polynomial equation, which is identical to the equation for a straight line. If you stop after the third (C) term, it is called a second-order, or quadratic, equation. If you stop after the fourth term, it is called a third-order, or cubic, equation.

If you choose a second, or higher, order equation, the graph of Y vs. X will be curved (depending on your choice of A, B, C...). Nonetheless, the polynomial equation is not strictly a *nonlinear* equation. Holding X and the other parameters constant, a graph of any parameter (A, B, C...) vs. Y would be linear. From a mathematical point of view, the polynomial equation is linear. This means little to scientists, but it means a lot to a mathematician, because it quite easy to write a program to fit data to linear equations. Because polynomial regression is related to linear regression, you don't have to enter any initial values.

But there is a fundamental problem with polynomial regression: Few biological or chemical models are described by polynomial equations. This means that the best-fit results can rarely be interpreted in terms of biology or chemistry.

Polynomial regression can be useful to create a standard curve for interpolation, or to create a smooth curve for graphing. But polynomial regression is rarely useful for fitting a model to biological data.

To perform polynomial regression with Prism, choose the nonlinear regression analysis but pick a polynomial equation. See "Polynomial equations" on page 177.

Multiple regression

Multiple regression fits data to a model that defines Y as a function of two or more independent (X) variables. For example, a model might define a biological response as a function of both time and concentration. The term *multiple regression* is usually used to mean fitting data to a linear equation with two or more X variables (X_1, X_2, \dots).

$$Y = A + B \cdot X_1 + C \cdot X_2 + D \cdot X_3 + E \cdot X_4 \dots$$

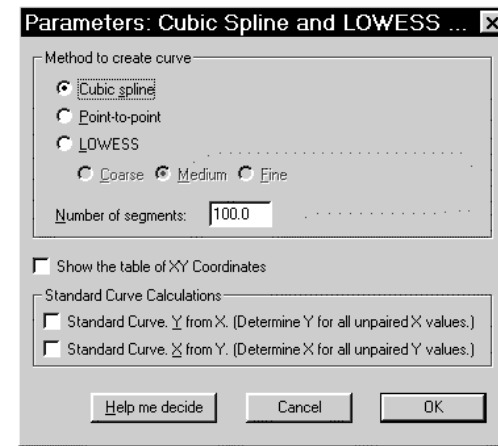
Nonlinear multiple regression models define Y as a function of several X variables using a more complicated equation. Prism cannot perform any kind of multiple regression. The companion program, GraphPad InStat, can perform basic multiple regression using the above equation. For details contact GraphPad software, or go to www.graphpad.com. No GraphPad programs can perform multiple nonlinear regression.

Logistic and proportional hazards regression

Linear, nonlinear and polynomial regression all fit data to models where Y is a continuous measured variable such as weight, concentration, receptor number, or enzyme activity.

If Y is a binomial outcome (for example male vs. female, pass vs. fail, viable vs. not viable) you need to use a different kind of regression, called *logistic regression*. Prism does not perform logistic regression. The only analysis it can do with binomial outcomes is to analyze contingency tables. See "Contingency tables" on page 121. By analyzing a contingency table, you can compare a binomial outcome in two or more groups. With logistic regression, you can compare outcomes after correcting for other differences between groups.

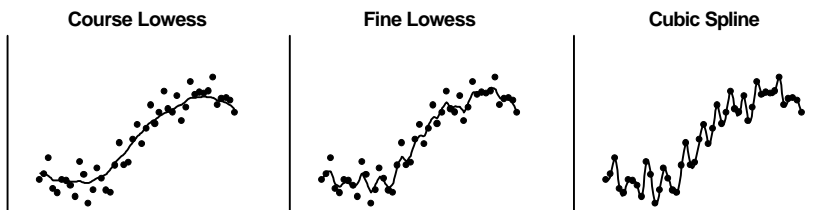
If Y is a survival time, you need to use yet another kind of regression, called proportional hazards regression. Prism can compare survival curves using the logrank test, but cannot perform proportional hazards regression. See "Survival curves" on page 109. Proportional hazards regression lets you compare survival curves after correcting for other differences.



Fitting a curve without choosing a model

The term *curve fitting* is more general than *regression*. Your approach to curve fitting depends on your goal. In some circumstances, your goal is simple. You don't care about models, and don't expect best-fit values that you can interpret. Instead, you just want to draw a smooth curve to make a graph look attractive, or to use as a standard curve.

Prism provides two approaches for fitting a curve without selecting a model. A **cubic spline curve** goes through every data point, bending and twisting as needed. A **lowess curve** follows the trend of the data. Lowess curves can be helpful when the data progresses monotonically, but are less helpful when there are peaks or valleys. Prism lets you choose between fine, medium and course lowess curves. The fine curve reveals the fine structure of the data, but tends to wiggle a lot. The coarse curve shows only the general trend, but obscures the detail.



To create a lowess or spline curve, click the Analyze button and choose Fit spline/lowess from the list of curves and regressions to bring up the dialog.

Prism generates lowess curves using an algorithm adapted from [Graphical Methods for Data Analysis](#), John Chambers et. al., Wadsworth and Brooks, 1983. Don't select a lowess curve unless you have well over twenty data points. Prism generates the curve as a series of line segments. Enter the number of segments you want, and check the option box if you need to see the XY coordinates of each point. To use the lowess, point-to-point, or spline curve as a standard curve, read "Reading unknowns from standard curves" on page 329.

Prism can also create a point-to-point "curve" – a series of line segments connecting all your data. 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 an option 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

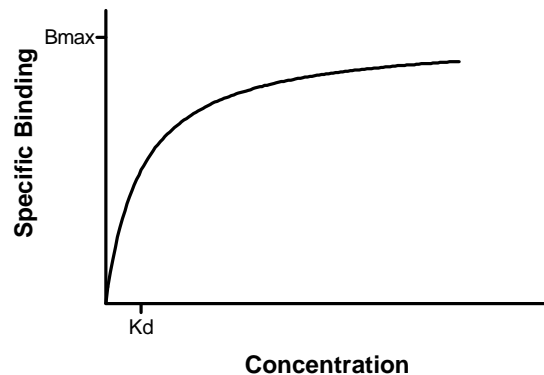
Spline curves can wiggle too much. Lowess curves can be too jagged. To get a smoother curve, consider using nonlinear regression and pick a model empirically. You don't have to pick a sensible model, and don't have to interpret the best-fit values. Use nonlinear regression to create a smooth curve, not as a method to analyze data. Polynomial models are often used for this purpose. See "Polynomial regression" on page 166.

Choosing or entering an equation (model)

Classic equations built-in to Prism

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 (usually 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.

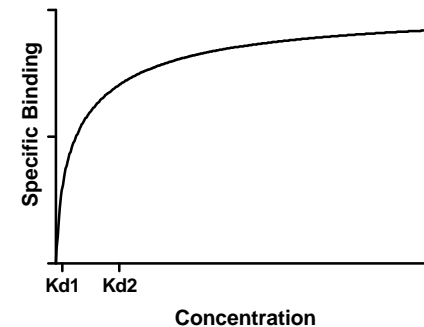
Note: Y should be the specific binding, not the total binding. To learn how Prism analyzes saturation binding curves, see "Analyzing saturation radioligand binding data" on page 249.

This equation also describes the activity of an enzyme as a function of substrate concentration. In this case, the variable labeled B_{max} is really V_{max}, the maximum enzyme activity, and the variable labeled K_d is really K_m, the Michaelis-Menten constant.

See also "Analyzing saturation radioligand binding data" on page 249, and "How to determine V_{max} and K_m" on page 321.

Two site binding

$$Y = \frac{B_{\max 1} \cdot X}{K_{d1} + X} + \frac{B_{\max 2} \cdot X}{K_{d2} + X}$$

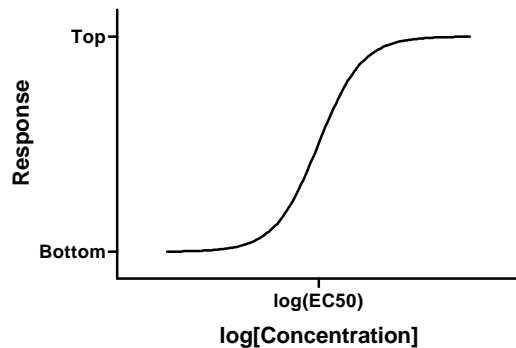


This equation is an extension of the one site binding curve. It shows the binding of a ligand to two receptors with different affinities (different K_d values). It also describes the enzyme activity as a function of substrate concentration when two isozymes are present. The curve in the example has K_d values that differ by a factor of ten, with equal B_{max} values. Even with such a large difference between K_d values, the curve is not obviously biphasic.

See "Determining K_d and B_{max} for two classes of binding sites" on page 251.

Sigmoidal dose-response

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + 10^{\text{LogEC50} - X}}$$



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. This equation is also called a *three-parameter logistic equation*.

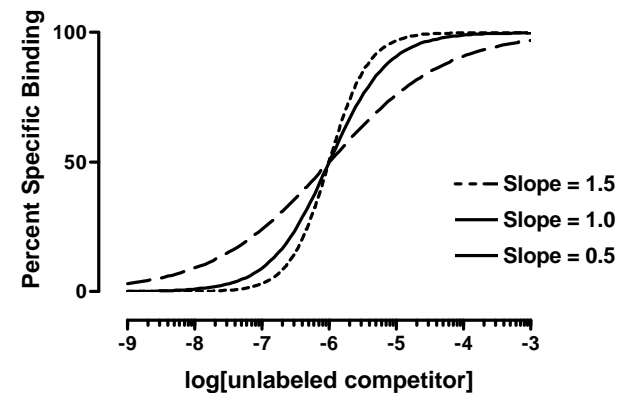
The variable Bottom is the Y value at the bottom plateau; Top is the Y value at the top plateau, and LogEC50 is the X value when the response is halfway between Bottom and Top. LogEC50 is the logarithm of the EC₅₀ (effective concentration, 50%). With different kinds of variables, this variable is sometimes called ED₅₀ (effective dose, 50%), or IC₅₀ (inhibitory concentration, 50%, used when the curve goes downhill).

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.

See "Analyzing dose-response curves" on page 297.

Sigmoidal dose-response (variable slope)

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + 10^{(\text{LogEC50} - X) \cdot \text{HillSlope}}}$$



This equation extends the previous equation, but allows for a variable slope. This equation is also called a *four-parameter logistic equation*.

The variable 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. With different kinds of variables, this variable is sometimes called ED₅₀ (effective dose, 50%), or IC₅₀ (inhibitory concentration, 50%, used when the curve goes downhill).

The variable HillSlope describes the steepness of the curve. This variable 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. A standard sigmoid dose-response curve (previous equation) has a Hill Slope of 1.0. When HillSlope is less than 1.0, the curve is more shallow. When HillSlope is greater than 1.0, the curve is steeper. The Hill slope has no units.

See "Analyzing dose-response curves" on page 297, and "The steepness of a dose-response curve" on page 299.

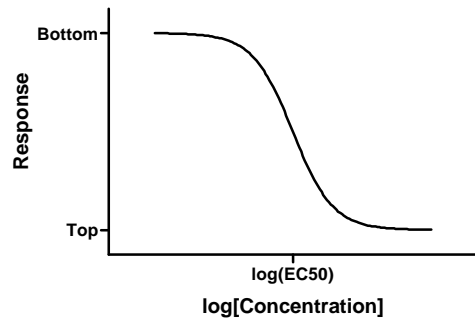
One site competition

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + 10^{X - \text{LogEC50}}}$$

This equation describes the competition of a ligand for receptor binding. It is identical to the sigmoid 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 sigmoid curves. The term IC₅₀ is used more frequently ("E" stands for effective; "I" stands for inhibitory).

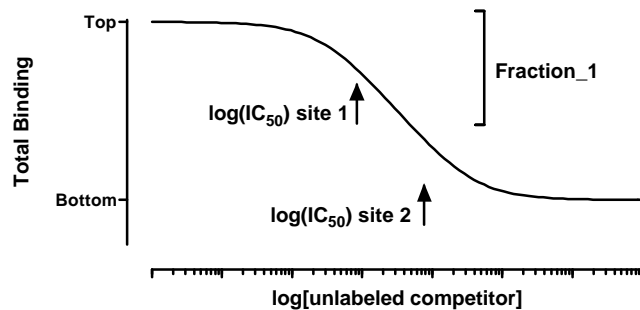
Usually the Y values are total binding. If you enter specific binding instead, fix BOTTOM to have a constant value of zero. If you enter percent specific binding, also set TOP to be a constant equal to 100.



See "Competitive binding data with one class of receptors" on page 266.

Two site competition

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) \left[\frac{\text{Fraction}_1}{1 + 10^{X - \text{LogEC50}_1}} + \frac{1 - \text{Fraction}_1}{1 + 10^{X - \text{LogEC50}_2}} \right]$$



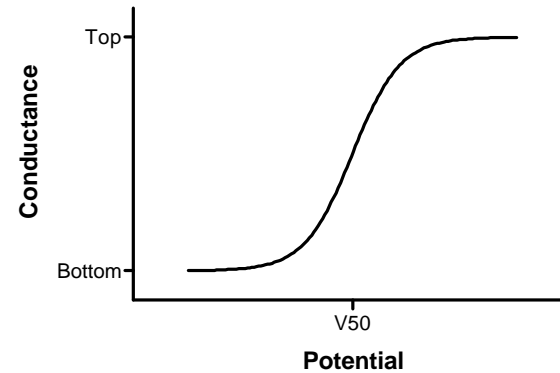
This equation describes the competition of a ligand for two types of receptors. The radioligand has identical affinities for both receptors, but the competitor has a different affinity for each.

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.

See "Competitive binding with two sites" on page 270.

Boltzmann sigmoid

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + \exp\left(\frac{V50 - X}{\text{Slope}}\right)}$$



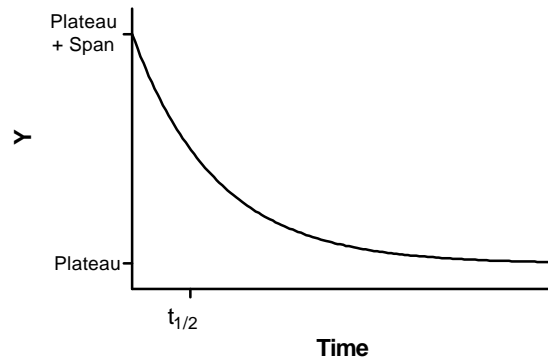
This equation describes voltage dependent activation of ion channels. It describes conductance (Y) as a function of the membrane potential (X). Conductance varies from BOTTOM to TOP. V50 is the potential at which conductance is halfway between BOTTOM and TOP. SLOPE describes the steepness of the curve, with a larger value denoting a shallow curve. Slope is expressed in units of potential, usually mV, and is positive for channels that activate upon depolarization.

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 constant, and z is the valence. Since $RT/F \approx -26$ mV at 25°C, $z = -26/\text{SLOPE}$.

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

$$Y = \text{Span} \cdot e^{-K \cdot X} + \text{Plateau}$$



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, the plateau equals zero. When fitting data to this equation, consider fixing the plateau to a constant value of zero.

See "Example model 2. Exponential decay" on page 158, and "Dissociation ("off rate") experiments" on page 287.

Two phase exponential decay

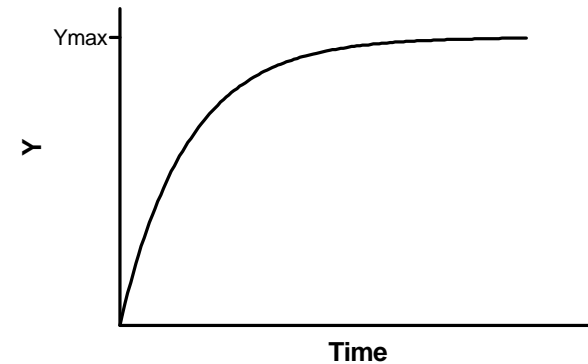
$$Y = \text{Span1} \cdot e^{-K_1 \cdot X} + \text{Span2} \cdot e^{-K_2 \cdot X} + \text{Plateau}$$



This equation describes a two phase exponential decay. Y starts out equal to $\text{Span1} + \text{Span2} + \text{PLATEAU}$ and decays to PLATEAU with fast and slow components. The two half-lives are $0.6932/K_1$ and $0.6932/K_2$. In the figure, the two rate constants differ tenfold, but the spans were equal. The curve is not obviously biphasic, and it takes a very practiced eye to see that the curve does not follow a single phase model.

One phase exponential association

$$Y = Y_{\text{max}} \cdot (1 - e^{-K \cdot X})$$



This equation describes the pseudo-first order association kinetics of the interaction between a ligand and its receptor, or a substrate and an enzyme. Y is either binding or enzyme activity. X is time.

Y starts out equal to zero and increases to a maximum plateau (at equilibrium) equal to YMAX. When X equals $0.6932/K$, Y equals $0.5 \cdot Y_{\text{MAX}}$.

See "Association binding experiments" on page 288.

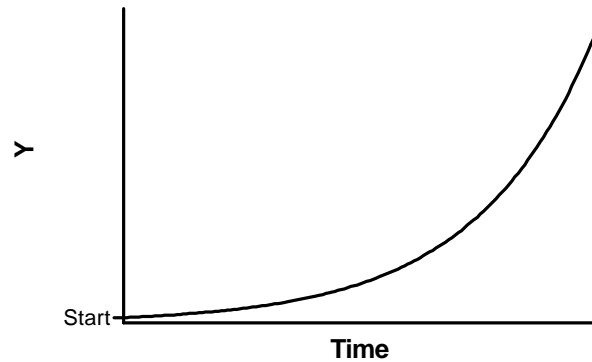
Two phase exponential association

$$Y = Y_{\text{max1}} \cdot (1 - e^{-K_1 \cdot X}) + Y_{\text{max2}} \cdot (1 - e^{-K_2 \cdot X})$$

This is an extension of the exponential association to two phases, corresponding to a radioligand binding to two independent sites.

Exponential growth

$$Y = \text{Start} \cdot e^{K \cdot X}$$



This describes an exponential growth curve. Y is population size (perhaps cell number) and X is time. At $X=0$, Y equals START. Y 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.

$$Y = A \cdot X^B + C \cdot X^D$$

Fitting data to a power series model can be difficult. The initial values generated automatically by Prism are not very helpful (all four parameters are set to 1.0). You'll probably need to enter better initial values in order to fit this equation to data. The initial values of B and D are important, because small changes in those values can make a huge change in Y.

The equation is not defined, and leads to a floating point error, if X equals zero and B or D are negative numbers or if X is negative and B or D are between 0.0 and 1.0.

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.

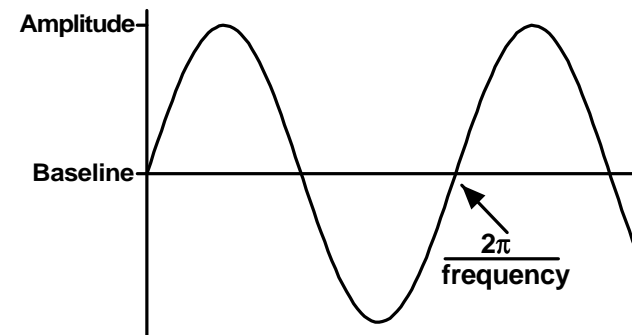
The "order" of a polynomial equation tells you how many terms are in the equation.

Order	Equation
First	$Y = A + B \cdot X$
Second	$Y = A + B \cdot X + C \cdot X^2$
Third	$Y = A + B \cdot X + C \cdot X^2 + D \cdot X^3$
Fourth	$Y = A + B \cdot X + C \cdot X^2 + D \cdot X^3 + E \cdot X^4$

You can enter a higher-order equation (up to 14th order) as a user-defined equation (or select one from the equation library).

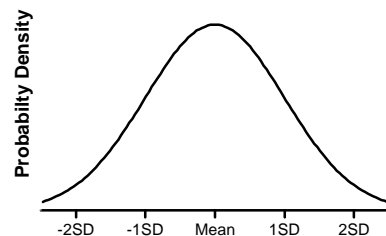
Sine wave

$$Y = \text{Baseline} + \text{Amplitude} \cdot \sin(\text{Frequency} \cdot X + \text{Offset})$$



X is in radians. In most cases, you'll want to fix BASELINE to a constant value of zero. AMPLITUDE is the maximum height of the curve away from the baseline. FREQUENCY is the number of complete oscillations per 1 X unit.

Gaussian distribution

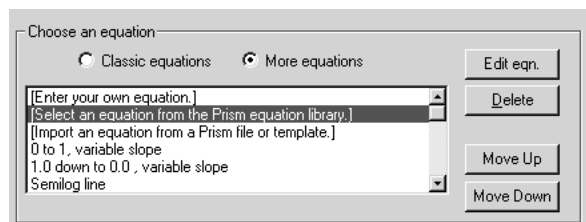


Cumulative probability distribution of a Gaussian bell-shaped distribution with specified mean and SD. The area under the entire curve is AREA. A standard probability distribution is scaled so that AREA equals 1.0. The units of the Y-axis are arbitrary, determined by your choice of AREA.

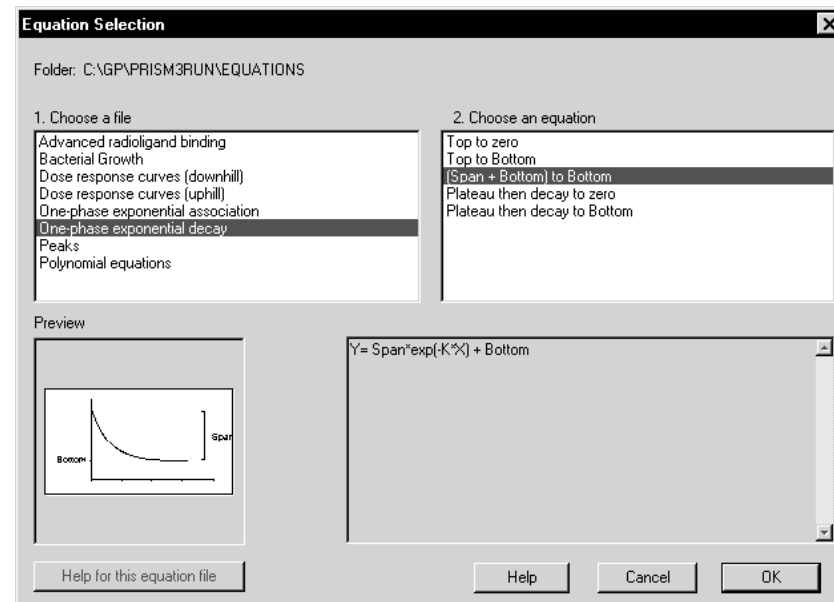
Importing equations and the equation library

Selecting from the Prism equation library

You are not limited to the classic equations built-in to Prism. You can also enter your own equation (see "User-defined equations" on page 181) or choose an equation from an equation library. To choose an equation from a library, click "more equations" (from the parameters dialog for nonlinear regression), then choose to select from the equation library.



On the Equation Selection dialog, first choose a library file (left panel) then choose an equation (right panel). Confirm your choice with the preview on the bottom of the dialog.



When you select an equation from the library, you transfer it from the library file to your own list of equations. You can then edit that copy of the equation without affecting the library. Prism stores the equation with every file that uses it, and also places the equation in your list of user-defined equations (the list you see when you choose "more equations"). Prism does not remember that the equation came from a library.

GraphPad may extend the equation library over time. Check www.graphpad.com for new equations.

You can add to the equation library. See "Creating equation library files" on page 187.

Importing an equation from another Prism project

You can import any equation from any project. Click "more equations" (from the parameters dialog for nonlinear regression). Choose to import an equation and select a file and an equation.

When you import an equation, you transfer to your own list of equations (the list you see when you choose "more equations"). Prism does not store any sort of link back to the file the equation was imported from.

Managing your list of equations

When you choose "more equations", Prism shows you a list of equations you have entered or imported. If you don't plan to use an equation again, select it and click Delete to erase it from the list. That won't affect any files

that use the equation you erased. If you open one of these files, and change the parameters of the nonlinear regression, Prism will automatically add the equation back to your list.

You can change the order of equations in your list by selecting an equation and then clicking Move up or Move down.

Modifying equations

You can edit any equation you entered yourself or imported (or chose from the equation library). From the nonlinear regression parameters dialog, select the equation from the list of "more equations" and then click Edit Eqn.

Classic equations cannot be modified. But you can create a new user-defined equation based on a classic equation.

To copy and paste a built-in equation:

1. Start from the Nonlinear Regression or Simulate Equation Parameters dialog.
2. Select a built-in classic equation, and click on the button labeled "Equation".
3. Press the button labeled "Copy All".
4. Cancel from that dialog.
5. Select "More equations", then "enter a new equation".
6. Enter an equation name. Then move the insertion point to the Equation block.
7. Press the button labeled "Paste".
8. Edit the equation.

User-defined equations

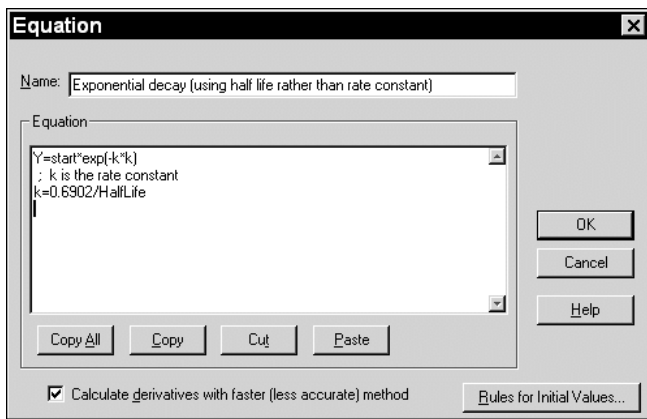
What kind of equations can you enter?

You can enter your own equations into Prism, subject to these limitations:

Limitation	Explanation
No implicit equations.	Y must be defined as a function of X and one or more parameters. The variable Y can only appear once, on the left side of the last line of the equation. If Y also appears on the right side of the equation, you have an implicit equation, which Prism cannot handle. In many cases, you'll be able to algebraically rearrange 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, so cannot fit models with two or more independent (X) variables. For example, Prism cannot fit a model that defines response as a function of both dose and time.
No equations with more than 14 variables.	Prism fits at most 14 variables, although you may use other intermediate variables when writing the equation.
No discontinuous equations.	If you enter a discontinuous equation (where an infinitesimal change in X can create a huge change in Y) the results of nonlinear regression may not be reliable.

Entering a user-defined equation

At the top of the parameters dialog for nonlinear regression (or simulate curve) select "More equations". Then select "Enter your own equation" to bring up the Equation dialog.



You must enter a name for the equation, which will then appear on the list of "more equations" in the nonlinear regression dialog.

In entering the equation itself, follow these guidelines (similar to those of the Basic computer language).

- 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.
- Prism does not distinguish between upper and lower case letters in variable names.
- Use an asterisk (*) to indicate multiplication and a caret (^) to indicate power. For example, "A*B" is the product of A times B and "A^B" is A to the B power.
- Use parentheses as necessary to show the order of operations. To increase readability, substitute brackets [like this] or braces {like this}. Prism interprets parentheses, brackets, and braces identically.
- Use a single equals sign to assign a value to a variable.
- You don't need any special punctuation at the end of a statement.
- 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.
- To enter a comment, type a semicolon (;) and type text. Comments can begin anywhere on a line.

You don't have to write your equation on one line. Use intermediate variables to simplify longer equations. Prism automatically distinguishes between intermediate variables and equation parameters that you can fit. If a variable is used first on the left side of an equals sign, then it is an intermediate variable. If a variable is used first on the right side of an equals sign, then it is an equation parameter.

Here are three examples of one-line equations:

$$Y=B_{max} * X / (K_d + X)$$

$$Y=A * (X+1)$$

$$Y=Start * exp[(-0.693 / Half_Life) * K]$$

Below is an example of a longer equation. 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. Note two comments, one on a line by itself and the other on the same line with equation code.

```
; One-phase exponential decay
K=0.693/HalfLife ;rate constant
Y=Start*exp(-K*X)
```

Available functions

When you enter your equations, you can use any of the functions listed below. Don't use any of those names for your variables. (The variables j and k in the descriptions below can be any variable or expression.)

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(k)	Converts k 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(k)	Converts k degrees to radians.
sgn(k)	Sign of k. If $k > 0$, $\text{sgn}(k) = 1$. If $k < 0$, $\text{sgn}(k) = -1$. If $k = 0$, $\text{sgn}(k) = 0$.
sin(k)	Sine. K is in radians.
sqrt(k)	Square root.
tan(k)	Tangent. K is in radians.

Using the IF function

Prism allows you to introduce some branching logic through use of the If function. The syntax is:

If (conditional expression, value if true, value if false)

You can precede a conditional expression with NOT, and can connect two conditional expressions with AND or OR. Examples of conditional expressions:

MAX > 100
 Ymax = Constraint
 (A < B or A < C)
 NOT(A < B AND A < C)
 FRACTION <> 1.0)
 X <= A and X >= B

Note: "<>" means not equal to. "<=" means less than or equal to. ">=" means greater than or equal to.

Here is an example.

Y= If (X<X0, Plateau, Plateau*exp(-K*X))

If X is less than X0, then Y is set equal to the variable Plateau. Otherwise Y is computed as Plateau*exp(-K*X).

You may also insert a conditional expression anywhere in an equation, apart from an If function. A conditional expression evaluates as 1.0 if true and 0.0 if false. Example:

Y=(X<4)*1 + (X>=4)*10

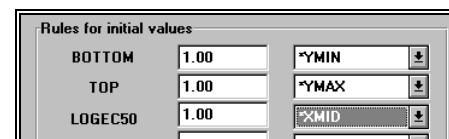
When X is less than 4, this evaluates to $1*1 + 0*10 = 1$. When X is greater than 4, this evaluates to $0*1 + 1*10 = 10$.

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. You can enter initial values at the time you fit curves, but it is helpful to define rules for generating the initial values at the time you enter a new equation. Then Prism will calculate the initial values automatically. If you don't enter rules for initial values, you will need to enter the initial values for every variable, for every data set, every time you fit data.

To define rules for initial values for user-defined equations:

1. While entering or editing a 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 two choices on the drop-down list are used to multiply or divide the number you entered by a value determined from the 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 sets the initial value to half of YMAX (which differs for each data set).

The first choice on the drop-down list is "(Initial value, to be fit)". This means that the value you entered will be the initial value for all data sets. The initial value will not depend on the range of the data.

The last choice is "(Hold constant)". This sets the parameter to a constant value, that won't be fit by nonlinear regression.

You can choose "(Hold constant)" even for parameters such as specific activity or protein concentration that have different values in each experiment. In the Rules for initial value dialog, choose "(Hold constant)" but leave the value blank. Each time the equation is used, Prism will prompt for a constant value before using the equation in nonlinear regression.

Creating equation library files

When you choose an equation for nonlinear regression (or simulations), Prism lets you import from the equation library. You can add to the library by following these steps:

1. Create a new Prism project. Use the Simulate Curve analysis and enter a new equation. Pick reasonable values for the minimum and maximum X values of the curve.
2. Customize the graph so it will be clear when seen as a preview (test this by looking at the graph gallery).
3. Repeat with any number of related equations that you want to store in one file.
4. Save the file into the Equations folder within the program folder.
5. If you are creating an equation file that will be used by others, consider creating a help file that explains the equations. Using any help compiler, create a help file with the same name as the file containing the equations but with the extension hlp. Place the help file in the same folder with the Equation file.

When you choose to select an equation from a library file, Prism displays a list of all files (and templates) in the Equations folder. When you pick one of those files, it shows a list of all user-defined equations used in the file, along with a thumbnail preview of the first graph linked to each equation. If a help file exists for the selected file, click "Help with this equation file" to open it.

Constraining variables in user-defined equations

Why constrain?

Prism has no common sense. When it finds the best-fit of an equation to your data, it doesn't know which values are reasonable. For example, if your data are quite scattered (or the initial values are way off) Prism might report that the best-fit value of a rate constant or a plateau is a negative number. Prism doesn't know that that negative numbers are impossible. All it knows is that setting that variable to a negative number reduces the sum-of-squares as much as possible. But your goal is not just to reduce the sum-of-squares, but to do so in a way that results in best-fit values that are scientifically sensible. If Prism reports results that are not sensible, consider adding constraints to your model (also consider using a different model).

Setting a parameter to a constant value

The most severe constraint is to set a parameter to a constant value. For example, you can set a plateau variable to zero if you have already subtracted out any baseline or nonspecific data. You can do this in two ways.

The direct approach is to edit the equation to set a parameter to a constant value. In the example below, the variable Plateau is fixed to a constant value, so its value is not altered by nonlinear regression.

```
Plateau=13.45
Y = Plateau + Span*exp(-K*X)
```

A simple approach is to use the Constants dialog. After selecting the equation in the nonlinear regression parameters dialog, click the Constants button to bring up a second dialog. In that dialog, you may set any of the parameters in the equation to a constant value.

Constraining to a positive value

A common constraint is to tell Prism that a variable must be positive. For example, you may wish to constrain the bottom plateau of an exponential dissociation curve to be positive. Here is an equation that does that using the IF function (see "Using the IF function" on page "185").

```
Bottom = IF(BASELINE<0, 0.0, Baseline)
Y = (TOP - BOTTOM)*exp(-K*X) + BOTTOM
```

The first line defines the intermediate variable Bottom to be zero if the variable Baseline is negative. Otherwise it sets Bottom equal to Baseline.

You can achieve the same effect using the MAX function. The following equation is equivalent to the previous one.

```
BOTTOM = MAX(BASELINE, 0.0)
Y = (TOP - BOTTOM)*exp(-K*X) + BOTTOM
```

The first line of the equation defines the intermediate variable BOTTOM to be the maximum 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).

There is no particular reason to favor using either the IF function or MAX functions. Use whichever you find easier to remember.

Constraining a variable to a range of values

You can constrain a variable to fall within a range of values. For example, the equation below fits a dose-response curve, constraining the bottom plateau to be between -10 to 10, and the top plateau to be between 90 to 110.

```
Top=IF(TopP<90,90, IF(TopP>110, 110, TopP))
Bottom=If(BottomP<-10, -10, If(BottomP>10,10, BottomP)
Y=Bottom + (Top-Bottom)/(1+10^((LogEC50-X)*HillSlope))
```

The logic of the nested If statements can be hard to follow. Here is the logic of the first line above, written in the Basic computer language (you cannot use this Basic syntax when entering models into Prism):

```
If TopP<90 then
  Top=90
ElseIf TopP>110 then
  Top=110
Else
  Top=TopP
EndIf
```

If you find the IF statements hard to follow, you might prefer to establish constraints using the Max and Min functions. The following equation is equivalent to the previous one. To understand the logic, try evaluating the first line setting the variable TopP to 85, 95, 105 and 115. The variable Top will equal 90, 95, 105 and 110 respectively.

```
Top=MAX(90, MIN(TopP, 110))
Bottom=MAX(-10, MIN(BottomP, 10))
Y=Bottom + (Top-Bottom)/(1+10^((LogEC50-X)*HillSlope))
```

An alternative approach to constraining a variable to a range of values

Hilborn and Mangel suggested an alternative approach to constraining a variable to a range of values ([The Ecological Detective](#), Princeton University Press, 1997).

Say you want to constrain the variable F to be between 0 and 1. Before using F in your user-defined equation, define it as a function of the variable FF. But rather than use an IF statement, or Min and Max functions, to constrain the variables, use this equation:

$$F = ((\pi/2) + \arctan(FF)) / \pi$$

Why does this work? No matter what value FF has, the arctangent is between $-\pi/2$ to $\pi/2$. Therefore, F will be between 0.0 and 1.0. Prism fits the value FF, which has no constraint. To make sense of the results, transform the best-fit value of FF into F, which will always be between 0.0 and 1.0. Do the same thing for both ends of the confidence interval for FF.

You can generalize this equation for use with any range of values.

$$F = ((ULimit - LLimit) * ((\pi/2) + \arctan(FF)) / \pi) + LLimit$$

ULimit and LLimit are the upper and lower limits of your constraint. Substitute values or constraint those parameters to be constant. Prism cannot fit ULimit and LLimit.

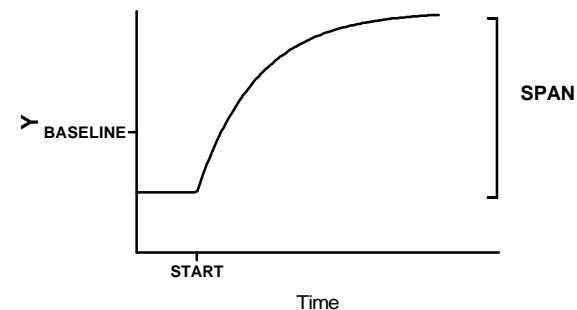
How to fit different portions of the data to different equations

In some situations you may wish to fit different models to different portions of your data. This often occurs in kinetic experiments where you add a drug or perform some sort of intervention while recording data. The values collected before the intervention follow a different model than those collected afterwards.

Although Prism has no built-in way to fit different equations to different portions of the data, you can achieve that effect using a user-defined equation containing the IF function.

Example 1. Plateau followed by exponential association

In this example, you collected data that established a baseline early in the experiment, up to "Start". You then added a drug, and followed the outcome (Y) as it increased towards a plateau. Prior to the injection the data followed a horizontal line; after the injection the data formed an exponential association curve.



```

Y1=BASELINE
Y2=BASELINE + SPAN*(1-exp(-K*(X-START)))
Y=IF( (X<START) , Y1 , Y2)

```

It is easiest to understand this equation by reading the bottom line first. For X values less than START, Y equals Y1, which is the baseline. Otherwise, Y equals Y2, which is defined by the exponential association equation.

This equation has two intermediate variables (Y1 and Y2). Prism can fit the four true variables: START, SPAN, K, and BASELINE.

In many cases, you'll make START a constant equal to the time of the experimental intervention. If you want Prism to fit START, choose an initial value carefully.

Example 2. Two linear regression segments

This example fits one linear regression line to points where X is less than some value X0 and another linear regression line to points where X is greater than X0. You can make X0 a constant (the time of an experimental intervention). Or you can ask Prism to find the best fit value of X0. The two lines will not necessarily meet at X0, and may intersect far from it.

Use this user-defined equation in Prism's nonlinear regression:

```

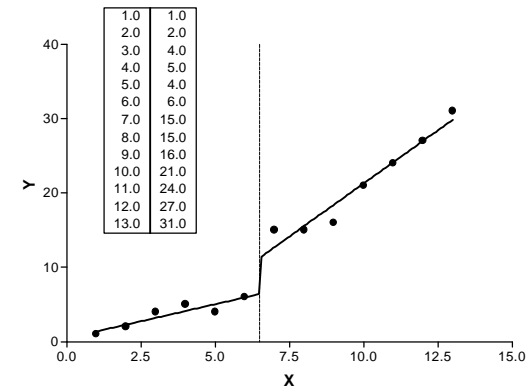
Y1=slope1*X + intercept1
Y2=slope2*X + intercept2
Y=IF(X<X0, Y1, Y2)

```

Y1 and Y2 define the two linear regression lines, each with their own slope and intercept. If X is less than X0, Y is computed using the Y1 formula, otherwise the Y2 formula is used.

If you want Prism to fit X0, give Prism a reasonable initial value. The rule that the initial value of X0 equals 1 times XMID works well for many data sets. The initial values of the two slopes and intercepts don't really matter (since this is a linear, rather than a nonlinear, equation). Prism will fit them accurately even if all initial values are set to zero.

This method fits two linear regression lines, but does not ensure that the lines meet at X0. The curve may be discontinuous, as shown below. In this example the best-fit value of X0 is 6.5. The two regression lines do not intersect at X=6.5, so the curve is discontinuous. The next example fits two linear regression lines that intersect at X0.



Example 3. Two linear regression segments that meet at the breakpoint

This equation fits two linear regression lines, ensuring that they intersect at X=X0.

```

Y1 = intercept1 + slope1*X
YatX0 = slope1*X0 + intercept1
Y2 = YatX0 + slope2*(X - X0)
Y = IF(X<X0, Y1, Y2)

```

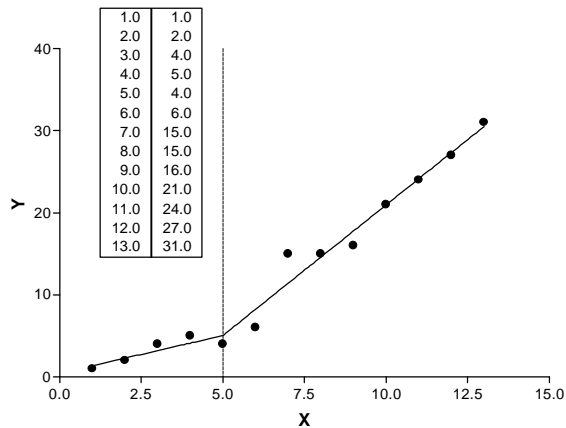
The first line of the equation defines the first line segment from its intercept and slope.

The second line of the equation computes the Y value of the first regression at the right end of that segment, when X=X0.

The third line of the equation computes the second regression segment. Since we want a continuous line, the Y value at the left end of the second segment must equal the Y value at the right end of the first segment (YatX0). The Y value at any other position along the second segment equals YatX0 plus the increase due to the second regression line. That increase equals the slope of the second segment (slope2) times the distance from X to X0.

The final line defines Y for all values of X. If X is less than X0 then Y is set equal to Y1. Otherwise Y is set equal to Y2.

Here are the results with sample data. The program found that the best-fit value of X0 was 5.00, and the two lines meet at that X value.



Example 4. Linear regression with three segments

This equation fits three linear regression lines, ensuring that they intersect at $X=X_0$ and $X=X_1$.

```

Y1 = intercept1 + slope1*X
YatX0 = intercept1 + slope1*X0
Y2 = YatX0 + slope2*(X - X0)
YatX1 = YatX0 + slope2*(X1-X0)
Y3 = YatX1 + slope3*(X - X1)
Y = IF(X<X0, Y1, IF(X<X1, Y2, Y3))

```

The logic is identical to that of the two segment regression line.

Use segmental linear regression cautiously

Segmental linear regression is appropriate when something happens at X_0 to change the slope of the line. For example, segmental linear regression might be appropriate if X is time, and you added a drug or changed a voltage at time X_0 .

If you didn't perform an intervention at X_0 , consider carefully whether segmental linear regression is the analysis of choice. Without an experimental intervention, segmental linear regression is rarely an appropriate method, as there usually is not a sharp break point. Instead, you probably want to fit some sort of curve.

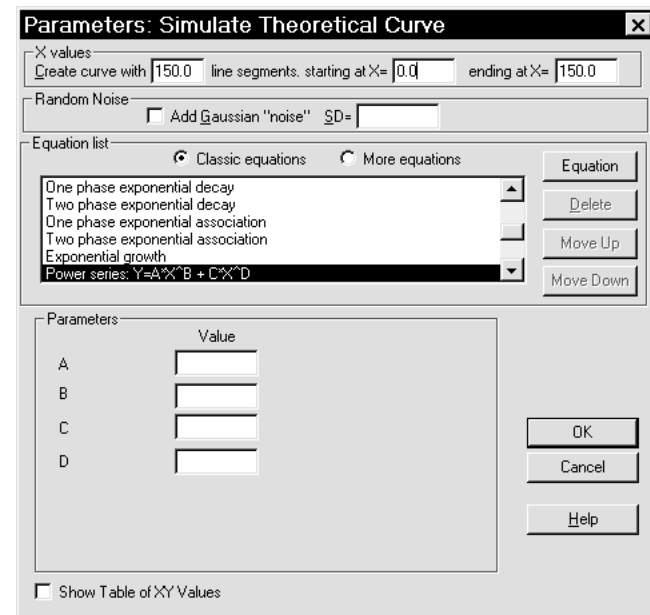
Warning. Segmental linear regression is definitely NOT an appropriate method to analyze a biphasic Scatchard plot. A biphasic Scatchard plot follows a curve, not two intersecting lines. There is no abrupt break point. You should fit the original data to a two-site binding curve instead.

How to simulate a theoretical curve

You'll find nonlinear regression most useful if you understand the models you have chosen. The best way to do this is to simulate a curve and then see what happens when you alter a parameter. Viewing graphs of simulated curves is a great way to learn about equations.

Prism can add Gaussian random error to each point in the simulated curve. This can help you test analysis methods. Create a curve with random error, and then analyze the simulated data.

To simulate a curve, start from a data table or graph. Click the Analyze button, select built-in analyses, and then select Simulate Theoretical Curve from the list of curve analyses. Select an equation (see "Choosing or entering an equation (model)" on page 171), enter a value for each parameter, and a range of X values. Check the option box to add random error to each simulated point. Prism generates random errors that follow a Gaussian (bell-shaped) distribution with a SD you enter.



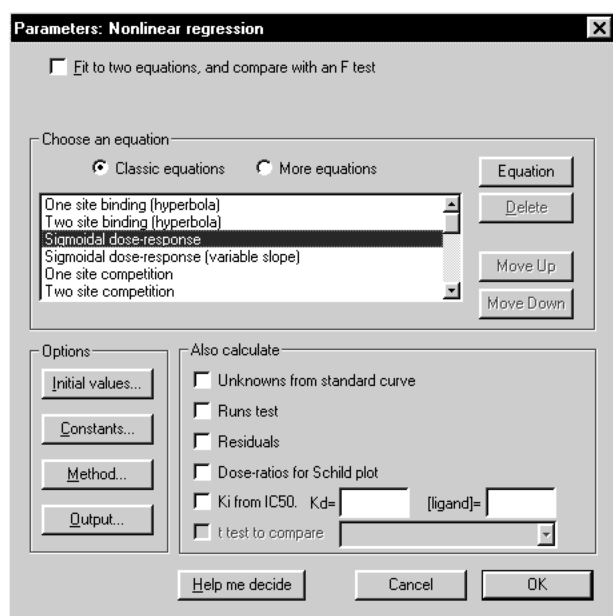
How Prism generates random numbers

Prism can add random values to each of the calculated Y values to simulate experimental error. Prism generates random numbers using routines adapted from Numerical Recipes in C, (W. H. Press et al, second edition, Cambridge Press, 1992; available online at www.nr.com). The function RAN3 (defined in Numerical Recipes) generates uniformly distributed random numbers and the function GASDEV transforms them to a Gaussian distribution with a mean of zero and a standard deviation you enter. 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”. The difference between truly random and pseudo-random numbers rarely creates a problem. For most purposes, computer generated random numbers are random enough to simulate data and test analytical methods.

Fitting curves with nonlinear regression

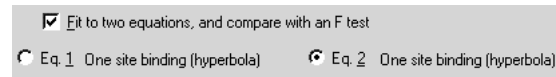
Fitting curves with nonlinear regression



To perform nonlinear regression

1. You may start either from a table of data or from a graph. Press the Analyze button and select built-in analyses. Then select Nonlinear regression from the list of curves and regressions.
2. Pick an equation. See "Classic equations built-in to Prism" on page 171, "Selecting from the Prism equation library" on page 178, and "User-defined equations" on page 181. For general information on creating models, see "Introduction to models" on page 157.

3. To compare two equations, select "Fit to two equations and compare with an F test." Check the option button for equation 1, and pick that equation. Then check the option button for equation 2 and pick the second equation. See "Comparing the fits of two models" on page 221.



4. Select the additional calculations you wish to perform.

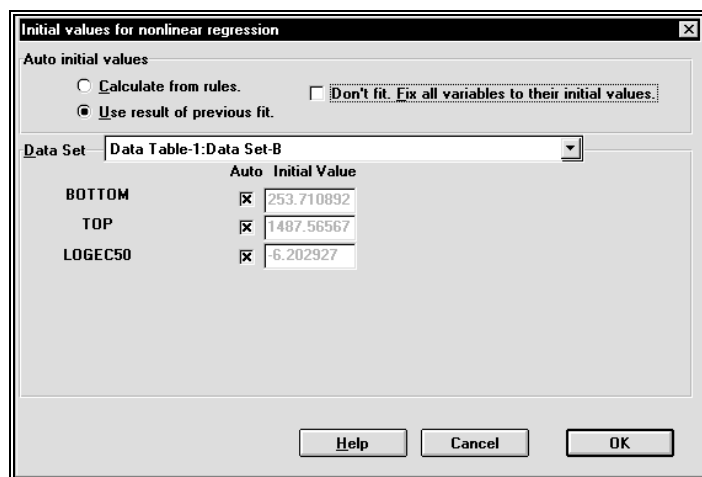
Calculation	Description
Calculate unknowns from a standard curve	After fitting a curve, Prism will interpolate unknown values from that curve. See "Reading unknowns from standard curves" on page 329.
Runs test	Tests whether the curve deviates systematically from your data. See "Runs test" on page 215.
Residuals	Helps you test the assumptions of nonlinear regression. See "Residuals" on page 214.
Dose-ratios for Schild plot	Computes the log of dose-ratios from a series of dose-response curves. See "Schild plot" on page 312
K _i from IC ₅₀	After fitting competitive binding curves, calculate the K _i from the IC ₅₀ . See "K _i from EC ₅₀ " on page 267.
t test to compare best-fit variables	If you fit two data sets, compare the best-fit values of each parameter using a t test. See "Compare two best-fit values from one experiment" on page 225. The results will be in the "Overview and comparison" view (page).

5. Set other options by clicking one of the four buttons at the bottom left of the nonlinear regression parameters dialog.

Option	Description
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 "For guidance interpreting the results, see "Approach to interpreting nonlinear regression results" on page 207. Initial values for nonlinear regression" on page 197.
Constants	Fix the value of one or more parameters in the equation to a constant value. See "Constants for nonlinear regression" on page 199.
Method	Choose how Prism weights data, handles replicates, and other options. See "Method options for nonlinear regression" on page 200.
Output	Configure the output by choosing the beginning and end of the curve, the number of significant digits, and more. See "Output options for nonlinear regression" on page 204.

For guidance interpreting the results, see "Approach to interpreting nonlinear regression results" on page 207.

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. It then adjusts these initial values to improve the fit.

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 186.

You'll find it easy to estimate initial values if you have looked at a graph of the data, understand the model, and understand the meaning of all the parameters in the equation. Remember that you just need an estimate. It doesn't have to be very accurate. If you are having problems estimating initial values, set aside your data and simulate curves using the model. Change the variables one at a time, and see how they influence the shape of the curve. Once you have a better feel for how the parameters influence the curve, you might find it easier to estimate initial values.

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. To change the initial value of a variable, deselect the Auto check box next to a variable and enter the new initial value.

After you fit data, Prism will use the best-fit values from that fit as the initial values for the next fit. If the changes you make to the data are minor, the fit will usually go much faster this way. But if you replace the data with values that are very different, the previous best-fit values may make very bad initial values, so bad that the nonlinear regression process may not converge on best-fit values. In this case, uncheck the option box "Use the results of the previous fit as initial values".

How much difference do initial values make?

When fitting a simple model to clean data, it won't matter much if the initial values are fairly far from the correct values. You'll get the same best-fit curve 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 or your model has many variables.

Viewing the curve generated by the initial values

If you aren't sure whether the initial values are reasonable, check "Don't fit, fix all variables to their initial values" on the initial values dialog. When you ok from the nonlinear regression dialog, Prism will not fit a curve but will instead generate a curve based on your initial values. If this curve is

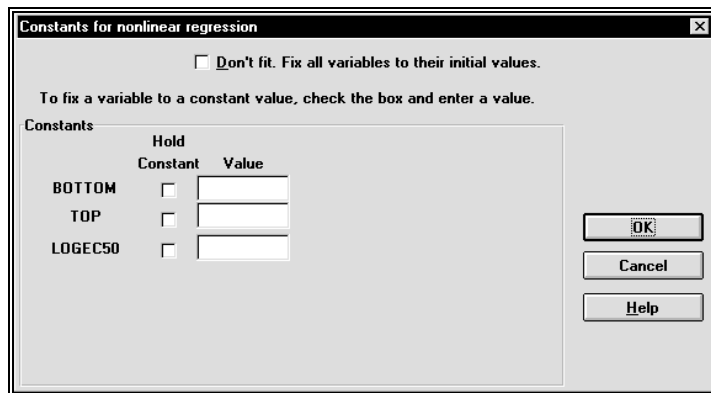
not generally in the vicinity of the data points, change the initial values before running nonlinear regression.

Constants for nonlinear regression

When performing nonlinear regression, you don't have to fit each parameter in the equation. Instead, you may fix one or more of the parameters to constant values.

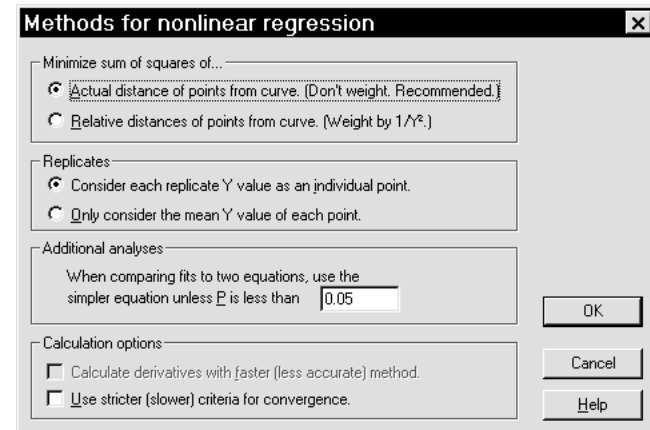
It is often helpful to define constants when you have only a few data points. For example, you might fix the bottom plateau of a sigmoid curve or exponential decay to zero.

Remember that Prism has no "common sense". Prism does not know how you did the experiment. Unless you tell it, Prism doesn't know that a curve has to plateau at zero



To set constants, click on the Constants button in the Nonlinear Regression Parameters dialog to bring up the Constants dialog. Check the "constant" box in front of the parameters 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. You cannot enter different constant values for each data set. When you set a parameter to a constant value, you fix it to that value for every data set analyzed.

Method options for nonlinear regression



Weighting method

Most often nonlinear regression is done without weighting. The program minimizes the sum-of-squares of the vertical distances of the data from the curve. This method gives equal weight to all points, as is appropriate when you expect experimental scatter to be the same in all parts of the curve.

If you expect experimental scatter to vary along the curve, Prism offers five ways to weight the points differently.

Relative weighting (weighting by 1/Y²)

The weighting method used most often is called weighting by 1/Y². It is easier to think of this method as minimizing the sum-of-squares of the relative distances of the data from the curve. This method is appropriate when you expect the average distance of the points from the curve to be higher when Y is higher, but the relative distance (distance divided by Y) to be a constant. In this common situation, minimizing the sum-of-squares is inappropriate because points with high Y values will have a large influence on the sum-of-squares value while points with smaller Y values will have little influence. Minimizing the sum of the square of the relative distances restores equal weighting to all points.

There are two ways to express the equation describing the quantity that nonlinear regression minimizes, shown below. The form on the left is, I think, easier to understand. You divide the distance of the data from the curve by the Y values of the data to obtain the relative distance, and then square that result. Most books on nonlinear regression use the equivalent form shown on the right – you first square the distance of the data from the curve, and then multiply that value times a weighting constant equal to

$1/Y^2$. That explains why relative weighting is often called weighting by $1/Y^2$.

$$\sum \left(\frac{Y_{\text{Data}} - Y_{\text{Curve}}}{Y_{\text{Data}}} \right)^2 = \sum \frac{1}{Y_{\text{Data}}^2} (Y_{\text{Data}} - Y_{\text{Curve}})^2$$

Weighting by $1/Y$

Weighting by $1/Y$ is a compromise between minimizing the actual distance squared and minimizing the relative distance squared. One situation where $1/Y$ weighting is appropriate is when the Y values follow a Poisson distribution. This would be the case when Y values are radioactive counts and most of the scatter is due to counting error. With the Poisson distribution, the standard error of a value equals the square root of that value. Therefore you divide the distance between the data and the curve by the square root of the value, and then square that result. The equation below shows the quantity that Prism minimizes, and shows why it is called weighting by $1/Y$.

$$\sum \left(\frac{Y_{\text{Data}} - Y_{\text{Curve}}}{\sqrt{Y_{\text{Data}}}} \right)^2 = \sum \frac{1}{Y_{\text{Data}}} (Y_{\text{Data}} - Y_{\text{Curve}})^2$$

Weighting by $1/X$ or $1/X^2$

The choices to weight by $1/X$ or $1/X^2$ are used rarely. These choices are useful when you want to weight the points at the left part of the graph more than points to the right.

Weighting by $1/SD^2$

Prism also offers the choice to weight by the reciprocal of the standard deviation squared. This means that data with little scatter (smaller standard deviation) get more weight than data with lots of scatter. This option will be useful if you understand how the scatter (or errors) arise in your experimental system, and can calculate appropriate weighting factors based on theory. Format a data table for entry of mean and SD, and enter (or paste) the weighting factors into the SD column. Don't use $1/SD^2$ weighting if the SD values are computed from a few replicates. Random scatter can cause some SD values to be high and some low, and these differences may not reflect consistent differences in variability. You want to choose a weighting scheme to account for systematic differences in the predicted amount of variability if you were to repeat the experiment many times. You should not choose weighting based on variability you happened to observe in one small experiment. If you choose to weight by $1/SD^2$, Prism minimizes this quantity:

$$\sum \left(\frac{Y_{\text{Data}} - Y_{\text{Curve}}}{SD} \right)^2 = \sum \frac{1}{SD^2} (Y_{\text{Data}} - Y_{\text{Curve}})^2$$

General comments on weighting

If the weighting scheme you chose would result in a division by zero for any value, Prism will not fit the dataset and reports "Weighting impossible" at the top of the results page.

Prism also considers sample size when weighting. If you entered individual replicates, and chose to treat the replicates separately, then no special calculations are needed. If you entered replicate values, but chose to fit to the mean of the replicates, then Prism always multiplies the weighting factor by N . The means computed from a large number of replicates get more weight than means computed from a few replicates. Similarly, if you enter mean, SD (or SEM) and N , Prism multiplies the weighting factor by N .

Replicates

If you collected replicate Y values at every value of X , there are two ways Prism can fit a model to the data. It can treat each replicate as a separate point, or average the replicate Y values, and treat the mean as a single point.

You should consider each replicate a separate point when the replicates are independent. Two examples:

- 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. The errors are independent.
- 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.

Average the replicates and treat the mean as a single value when the replicates are not independent. Two examples:

- You performed a binding experiment with a single tube at each concentration, but assessed the radioactivity in each tube three times. Those three values are not independent. Any experimental error while conducting the experiment would affect all the replicates.

- You performed a dose-response experiment, using a different animal at each dose with triplicate measurements. The three measurements are not independent. If one animal happens to respond more than the others, that will affect all the replicates. The replicates are not independent.

Calculation options

Threshold P value

Prism can compare the fits of two equations using an F test and choose the equation to graph 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.

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.

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 (or use a library equation), Prism evaluates the derivatives numerically, and you can choose the method Prism uses.

Ordinarily, Prism uses Richardson's method to evaluate the derivatives. This method calculates the derivative by determining Y after both increasing and decreasing the value of the variable a bit.

Check the option box to use a faster, but potentially less accurate, method (which only evaluates the equation after increasing the value).

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. You can then switch to the quick method for routine analyses if the speed of nonlinear regression matters to you. With small data sets and fast computers, nonlinear regression will seem instantaneous even if you pick the slower method.

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 five consecutive iterations each reduce the sum-of-squares by less than 0.000001%.

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. If you select the standard method, Prism will automatically switch to the stricter criteria if the R^2 is less than 0.3

Selecting the stricter criteria rarely affects the results but slows the calculations a bit (only noticeable with huge data sets or slow computers).

Stop calculating after 25 iterations

If this option is checked, Prism will stop nonlinear regression after 25 iterations. In most cases, nonlinear regression converges in fewer than 25 iterations. If the iterations continue beyond 25, it may be because you've picked an inappropriate equation, picked unhelpful initial values, or have very scattered data. This option insures that Prism won't spend a long time on calculations that won't be helpful. It is especially useful when you use a Prism script to fit many data sets.

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.

Number of significant digits

Choose the number of significant digits used to report results. This is especially useful if you embed the results table on a graph or layout.

Range of X values and number of line segments

After determining the best-fit values of the variables in the equation, Prism calculates and plots the best-fit curve. It automatically plots the curve starting at the X position of the first (lowest X) data point and ends at the last data point. You may enter different limits.

Curves are defined by many short line segments. You decide how many segments Prism will create. Prism initially creates all curves with 150 line segments. Increasing the number may improve the accuracy of standard curve calculations and make the curve appear smoother (especially if it has many inflection points).

Summary table and graph

When analyzing several data sets, the results table is rather lengthy. To display key results on a summary table, check the option box to create a summary table and select the variable you wish to summarize. Prism creates a summary table (as an additional results view) that shows the best-fit value of that variable for each data set, and graphs this 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. In some cases, you may analyze the summary table 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.

Warning. When Prism compares the fits of two equations, it shows only the results for the second equation. Since this may not be helpful, we suggest that you only make summary tables when fitting a single equation.

- Number of line segments to generate curves.

Note: Changing the analysis options changes the default settings for future nonlinear regression analyses. It will not change analyses you have already performed.

Default options for nonlinear regression

The nonlinear regression parameters dialog affects one particular nonlinear regression analysis. Change settings on the Analysis Options dialog to change default settings for future nonlinear regression analyses. To open this dialog, pull down the Edit menu and choose Options, then Analysis options. You can change these settings:

- 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?

Interpreting the results of nonlinear regression

Approach to interpreting nonlinear regression results

When evaluating the results of nonlinear regression, ask yourself:

Question	See page
Do the nonlinear regression results make sense?	208
How certain are the best-fit values?	209
How good is the fit?	212
Does the curve systematically deviate from the data?	214
Could the fit be a local minimum?	216
Have you violated an assumption of nonlinear regression?	217
Have you made a common error when using nonlinear regression?	218

The following sections provide additional help:

Topic	See page
Comparing the fits of two models	221
Comparing fits to two sets of data (same equation)	224
Analyzing radioligand binding data	237
Analyzing saturation radioligand binding data	249
Analyzing competitive binding data	263
Analyzing kinetic binding data	287
Analyzing dose-response curves	297
Analyzing enzyme kinetic data	317
Reading unknowns from standard curves	329

Do the nonlinear regression results make sense?

First check whether Prism actually fit a curve. Sometimes Prism cannot fit the data and reports an error instead, such as “floating point error” or “did not converge”. This does not mean there is a bug in Prism, just that Prism was not able to fit your data using the model and options you selected. The exact wording of the message is rarely helpful in diagnosing the problem. Consider the possibilities listed in “Have you made a common error” on page 218.

Next, look at a graph of the best-fit curve. In rare cases, the curve 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.

If the curve goes near the points, look at the best-fit values of the parameters to see if they make sense. Prism has no common sense and doesn’t know the context of your experiment. The curve fitting procedure can sometimes yield results that make no scientific sense. For example with noisy or incomplete data, Prism can report a best-fit rate constant that is negative, a best-fit fraction that is greater than 1.0, or a best-fit K_D value that is negative. All these results are scientifically meaningless, but Prism doesn't know that. Also check whether the best-fit values of the variables make sense in light of the range of the data. The results make no sense if the the top plateau of a sigmoid curve is far larger than the highest data point, or an EC_{50} is not within the range of your X values.

If the results make no scientific sense, they are unacceptable, even if the curve comes close to the points and R^2 is close to 1.0.

How certain are the best-fit values?

Along with the best-fit value of each variable in the equation, Prism reports its standard error and 95% confidence interval.

Standard errors of best-fit values

Prism reports a standard error of each best-fit value. If the SE is low, the best-fit value is “tight” – if you changed the variable a little bit, the curve would fit much worse. If the SE is high, the best-fit value is not so certain. You could change the value of the variable a lot without noticeably changing the goodness-of-fit.

The standard errors reported by Prism (and virtually all other nonlinear regression programs) are based on some mathematical simplifications. They are called “asymptotic” or “approximate” standard errors. They are

calculated assuming that the equation is linear, but are applied to nonlinear equations. This simplification means that the intervals can be too optimistic.

To understand how the standard errors are calculated requires mastering the matrix algebra of nonlinear regression (way beyond the scope of this manual). There is no simple formula. The standard errors are a function of the number of data points, the distance of the points from the curve, and the overall shape of the curve.

Some programs call these values standard deviations rather than standard errors. There really is no distinction between the standard error and standard deviation of a best-fit value. The term *standard error* refers to the standard deviation of a computed value. So the standard error of the mean is the same as the standard deviation of the mean (which is very different than the standard deviation of the data), and the standard error of a K_d or slope is the same as the standard deviation of a K_d or slope. Prism uses the term *standard deviation* to refer only to a measure of variability among values, and uses the term *standard error* to refer to the accuracy of a calculated value such as a mean or best-fit value.

By themselves, the SE values are difficult to interpret. They are used to calculate 95% confidence intervals, which are easier to interpret.

Confidence intervals of best-fit values

For each parameter that Prism fits, it reports a 95% confidence interval.

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 (but you won't know when this happens).

The confidence intervals are computed from the best-fit values and the SE of those best-fit values using this equation:

$$\text{BestFit} - t^* \cdot \text{SE} \text{ to } \text{BestFit} + t^* \cdot \text{SE}$$

The confidence interval is always centered at the best fit value and extends the same distance above and below it. That distance equals the SE of the best-fit value (also reported by Prism) times the critical value from the t distribution, abbreviated above as t^* . This value depends on the degree of confidence you want (usually 95%) and the number of degrees of freedom, which equals the number of data points minus the number of parameters fit by nonlinear regression.

Confidence intervals of constrained variables

If you constrained any variables with IF or MAX or MIN functions, Prism won't report standard errors or confidence for variables that have active constraints.

For example, the following statement constrains the variable top to be between 90 and 100. See "Constraining variables in user-defined equations" on page 187.

```
Top=IF(TopP<90,90, IF(TopP>110, 110, TopP))
```

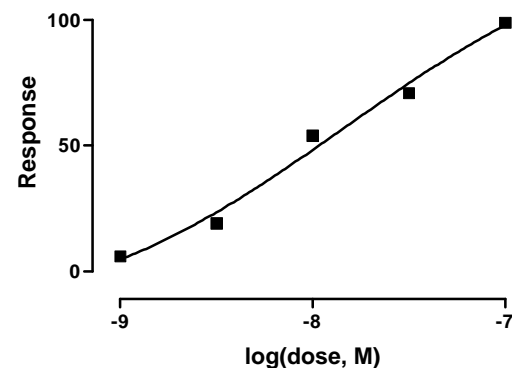
Prism doesn't report the best-fit value of Top, which is an intermediate variable. Instead it reports the best-fit value of TopP. If the best-fit value of TopP is between 90 and 110, then the constraint is not active and you can interpret the results as usual. If the best-fit value of TopP is over 110, then Top was constrained, so you can ignore the reported best fit value of TopP. Top equals 110. Similarly, if TopP is less than 90, ignore its value and report that Top equals 90. Prism doesn't report a standard error or confidence interval for variables with active constraints.

Explanations for high SE and wide confidence intervals

If the SE values are very high and the confidence intervals are very wide, the results of nonlinear regression won't be useful. The following four situations can cause confidence intervals to be wide:

Data collected over a too narrow range of X values

The confidence intervals of best-fit values provided by nonlinear regression will be wide if you have not collected data over a wide enough range of X values to fully define the curve. One example is a sigmoid dose-response curve with no data defining the top and bottom plateau.



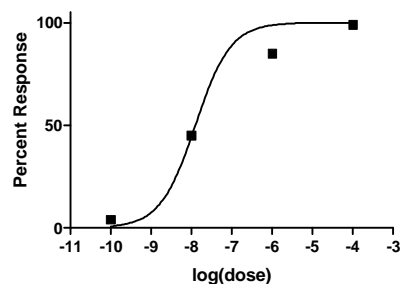
When these data were fit to a sigmoid dose-response curve, the 95% confidence interval for the EC_{50} extended over fifteen orders of magnitude!

The explanation is simple. The data were fit to a sigmoid equation with four variables: the top plateau, the bottom plateau, the slope, and the EC₅₀ (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, many curves (defined by different sets of parameters) would fit these data with similar sum of squares values.

In this example, it might make scientific sense to set the bottom plateau to 0% and the top plateau to 100% (assuming that 0% and 100% are well defined by control experiments). But Prism doesn't know this. It just finds values for Top and Bottom that make the curve fit well. In the example above, the best-fit value of the bottom plateau is -23 and the best-fit value of the top plateau is 137. Prism doesn't know that a negative value of Y makes no sense. If you defined the Y values to be percent of control, you could set the bottom plateau to a constant of zero and the top plateau to a constant of 100. If you do this, Prism fits only the EC₅₀ and the slope factor, and the confidence intervals will be narrower.

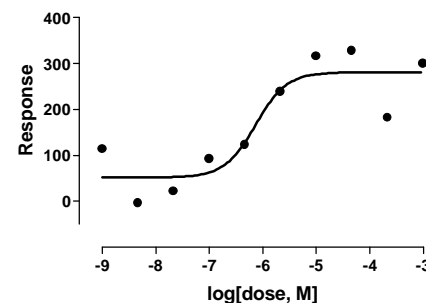
Note that the problem with the fit is not obvious by inspecting a graph, because the curve goes very close to the points. The value of R² (0.99) is also not helpful. That value just tells you that the curve comes close to the points, but does not tell you whether the fit is unique.

Data missing in an important part of the curve



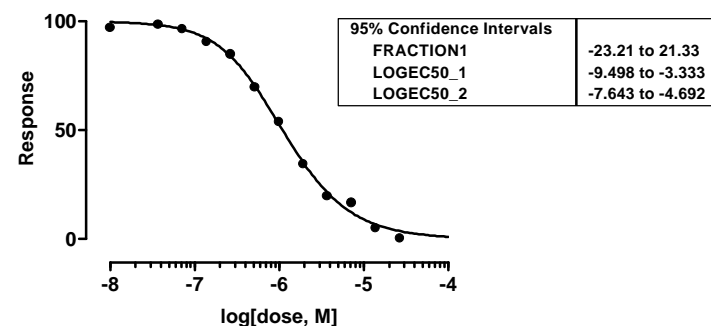
In fitting this dose-response curve, we set three parameters to constant values. We set the bottom plateau to equal zero, the top plateau to equal 100, and the slope to equal 1.0. There is only one variable to fit: the logEC₅₀. But the 95% CI for the EC₅₀ is quite wide, extending over almost an order of magnitude. The problem is simple. The EC₅₀ is the concentration at which the response is half-maximal, and there are few data points near that point.

Scattered data



Because of the scattered data, the confidence interval for the EC₅₀ in this experiment is wide, extending over a fiftyfold range.

The equation contains redundant variables



These data were fit to a competitive binding curve with two sites, constraining the bottom plateau to equal zero and the top plateau to equal 100. The curve comes very close to the points, so the R² is very high (0.997). But the confidence intervals for the two logEC₅₀ values all extremely wide, extending over several orders of magnitude. The confidence interval for the fraction of high-affinity sites extends to values that are nonsense (fractions must be between 0 and 1). The problem is that these data fit a one-site model just fine. There is no information in the data to support a two-site fit. When you choose a two-site model, the nonlinear regression procedure finds extremely wide confidence intervals.

How good is the fit?

Sum-of-squares from nonlinear regression

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.

If you chose to weight the values and minimize the relative distance squared, Prism reports both the absolute sum-of-squares (defined above) and the relative sum-of-squares, which is minimized.

S_{y,x}

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.

Prism calculates $s_{y,x}$ from the sum-of-squares (SS) and degrees of freedom (df, equal to number of data points minus the number of parameters fit) as:

$$s_{y,x} = \sqrt{\frac{SS}{df}}$$

R² from nonlinear regression

The value R^2 quantifies goodness of fit. It is a fraction between 0.0 and 1.0, and has no units. Higher values indicate that the model fits the data better. You can interpret R^2 from nonlinear regression very much like you interpret r^2 from linear regression. By tradition, statisticians use uppercase (R^2) for the results of nonlinear and multiple regression and lowercase (r^2) for the results of linear regression, but this is a distinction without a difference.

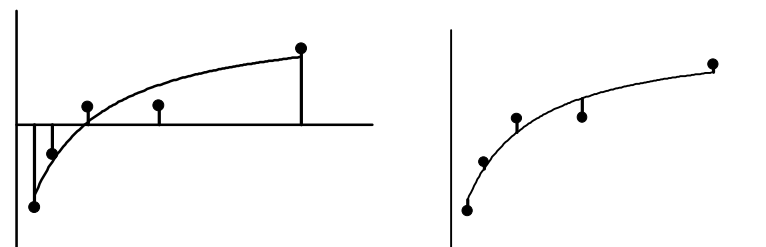
Tip: Don't make the mistake of using R^2 as your main criteria for whether a fit is reasonable. A high R^2 tells you that the curve came very close to the points. That doesn't mean the fit is "good" in other ways. The best-fit values of the parameters may have values that make no sense (for example, negative rate constants) or the confidence intervals may be very wide.

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).

R^2 is computed from the sum of the squares of the distances of the points from the best-fit curve determined by nonlinear regression. This sum-of-squares value is called SS_{reg} , which is in the units of the Y-axis squared. To

turn R^2 into a fraction, the results are normalized to the sum of the square of the distances of the points from a horizontal line through the mean of all Y values. This value is called SS_{tot} . If the curve fits the data well, SS_{reg} will be much smaller than SS_{tot} .

The figure below illustrates the calculation of R^2 . Both panels show the same data and best-fit curve. The left panel also shows a horizontal line at the mean of all Y values, and vertical lines showing how far each point is from the mean of all Y values. The sum of the square of these distances (SS_{tot}) equals 62735. The right panel shows the vertical distance of each point from the best-fit curve. The sum of squares of these distances (SS_{reg}) equals 4165.



R^2 is calculated using this equation.

$$R^2 = 1.0 - \frac{SS_{reg}}{SS_{tot}} = 1.0 - \frac{4165}{62735} = 1.0 - 0.0664 = 0.9336$$

Note that R^2 is not really the square of anything. If SS_{reg} is larger than SS_{tot} , R^2 will be negative. While it is surprising to see something called "squared" have a negative value, it is not impossible (since R^2 is not actually the square of R). R^2 will be negative when the best-fit curve fits the data worse than a horizontal line at the mean Y value. This could happen if you pick an inappropriate model, or fix a parameter to an inappropriate constant value (for example, if you fix the Hill slope of a dose-response curve to 1.0 when the curve goes downhill).

Even if you choose weighted nonlinear regression, Prism still computes R^2 using the formula above. In other words, it computes R^2 from the unweighted sum-of-squares.

If you want to compare the fit of two equations, don't just compare R^2 values. See "Comparing the fits of two models" on page 221.

Does the curve systematically deviate from the data?

If you've picked an appropriate model, and your data follow the assumptions of nonlinear regression, the data will be randomly distributed around the best-fit curve. You can assess this in three ways:

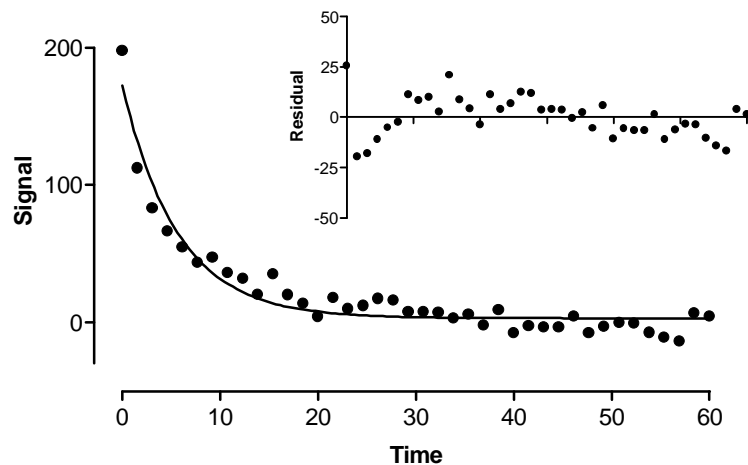
- The distribution of points around the curve should be Gaussian.
- The average distance of points from curve should be the same for all parts of the curve (unless you used weighting).
- Points should not be clustered. Whether each point is above or below the curve should be random.

Residuals and runs help you evaluate whether the curve deviates systematically from your data.

Residuals from nonlinear regression

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 the Y values are the vertical distances 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 at early and late times that are below the curve, and a cluster of points at middle time points that are above the curve. This is much easier to see on the graph of the residuals in the inset. The data are not randomly scattered above and below the X-axis.



Runs test from nonlinear regression

The *runs test* determines whether the curve deviates systematically from your data. 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 consecutive 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 N_a points above the curve and N_b points below the curve, the number of runs you expect to see equals $[(2N_a N_b)/(N_a + N_b)] + 1$. If you observe fewer runs than expected, it may be a coincidence or it may mean that you picked an inappropriate regression model and the curve systematically deviates from your data. The P value from the runs test answers this question: If the data really follow the model you selected, what is the chance that you would obtain as few (or fewer) runs as observed in this experiment?

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, conclude that the data don't really follow the equation you have selected.

In the example above, you expect 21 runs. There are 13 runs, and the P value for the runs test is 0.0077. If the data were randomly scattered above and below the curve, there is less than a 1% chance of observing so few runs. The data systematically deviate from the curve. Most likely, the data were fit to the wrong equation.

Testing whether the residuals are Gaussian

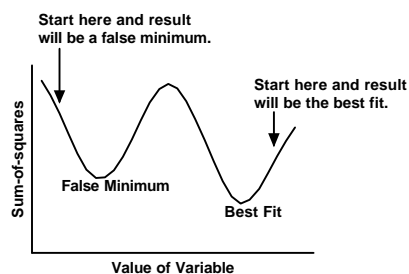
Nonlinear regression (as well as linear regression) assumes that the distribution of residuals follows a Gaussian distribution. You can test this assumption with Prism.

1. From the nonlinear regression parameters dialog, check the option box to show a table of residuals. If you have already fit your data, you can go back to this dialog by pressing Change Parameters from your results.
2. Go to the nonlinear regression results. Drop the list of results views (middle of third row of toolbar) and choose Residuals.
3. Press Analyze. Choose Column statistics from the list of statistical analyses.
4. On the column statistics parameters dialog, check the option box to test whether the distribution is Gaussian.

- Look at the results of the normality test on the column statistics results page. If the P value is small, the data fail the normality test.

Could the fit be 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 a 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 regression works iteratively to reduce the sum-of-squares. This is like walking downhill to find the bottom of the valley. See “How nonlinear regression works” on page 163. 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 test for the presence of a false minimum:

- Note the values of the variables and the sum-of-squares from the first fit.
- Make a large change to the initial values of one or more variables and run the fit again.
- Repeat step 2 several times.

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.

Have you violated an assumption of nonlinear regression?

Question	Discussion
Did you pick a sensible model?	Nonlinear regression adjusts the variables in the equation (model) for which you chose to minimize the sum-of-squares. It does not attempt to find a better equation.
Have you set appropriate parameters to constant values?	Prism doesn't have to fit all the parameters in a model. In many cases, it is appropriate to set one or more parameters to constant values. See "Constants for nonlinear regression" on page 199.
Is the scatter Gaussian?	Nonlinear regression assumes that the scatter is Gaussian. To test this assumption, see “Testing whether the residuals are Gaussian” on page 216.
Is the scatter consistent?	Nonlinear regression assumes that the scatter of points around the best-fit curve has the same standard deviation all along the curve. The assumption is violated if the points with higher (or lower) Y values also tend to be further from the best-fit curve. The assumption that the standard deviation is the same everywhere is termed <i>homoscedasticity</i> .
Is variability only in the Y direction?	The nonlinear regression model assumes that X values are exactly correct, and that experimental error or biological variability only affects the Y values. 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.
Does each data point contribute independent information	The deviation of each value from the curve should be random, and should <u>not</u> 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. You will also violate this assumption if you tell Prism to treat each replicate as a separate point, when they are not independent. See “Replicates” on page 202.

Have you made a common error when using nonlinear regression?

Consider these possibilities if you obtained an error message rather than results, or if the nonlinear regression results made no sense.

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 entered your own equation, check the rules for initial values. See "Defining rules for initial values in nonlinear regression" on page 186.
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 X or Y values are huge, change the units. Don't use values greater than about 10^4 .
Your numbers are too small.	If your X or Y values are tiny, change the units. Don't use values less than about 10^{-4} .
You've set a parameter to an inappropriate constant value.	Either don't set the parameter to a constant value, or change that value. Perhaps you made a simple mistake like setting a maximum plateau to 1.0 when it should be 100, or a Hill slope to +1.0 when it should be -1.0.

Comparing two curves

Comparing the fits of two models

Approach to comparing two models

When comparing the fits of two models, the first step is to examine the best-fit values of each model to make sure they are scientifically reasonable. Also make sure that the confidence intervals are not extremely wide. If the best-fit values of one model make no sense (for example, if a rate constant is negative) or if the confidence intervals are very wide, then reject that model and disregard the results of the F test.

If both models fit the data with sensible values, compare goodness-of-fit as quantified by sum-of-squares. If the more complicated equation fits worse (has higher sum-of-squares) than the simpler equation, then you should clearly reject the more complicated equation and conclude that the simpler equation fits better. This will happen rarely, as the curve generated by the more complicated equation (the one with more variables) will nearly always have a lower sum-of-squares, simply because it has more inflection points (it wiggles more).

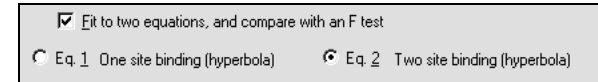
If both models fit the data with sensible values, and the more complicated model fits better, then you need to use statistical calculations to decide which model to accept. Do this with an F test to compare fits.

Comparing the fits of two equations using an F test

If you don't know which of two equations is more appropriate for your data, fit both and compare the results. For example, compare models with one and two classes of binding sites. Or compare a one-phase exponential dissociation curve with a two-phase curve.

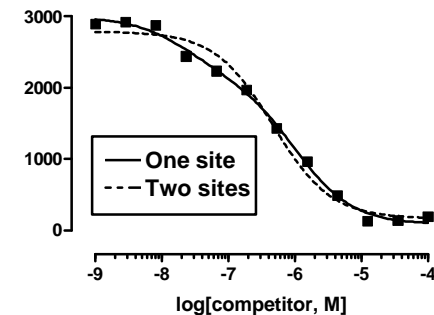
Prism can compute the F test automatically. To compare fits, simply, check the option box on top of the nonlinear regression parameters dialog. Check the option button for equation 1, and choose that equation along with its initial values. Also set any variables to constants, if necessary. Then check

the option button for equation 2, and choose an equation and adjust settings for initial values and constants.



Tip: It is not possible to automatically compare the fit of data to a built-in equation with a fit to a user-defined equation. Instead copy a built-in equation to the user-defined list. Then compare two user-defined equations.

The F test compares the fit of two equations, where the more complicated equation (the one with more parameters) fits better (has a smaller sum-of-squares) than the simple equation. 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 (detailed below) calculates 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, conclude that the more complicated model is significantly better than the simpler model. Most investigators accept the more complicated model if the P value is less than 0.05.



This graph shows the results of fitting both a one-site and two-site competitive binding curve to some data. Both fits are scientifically plausible, so it makes sense to perform compare the sum-of-squares with the F test.. Here is how Prism reports the comparison:

Comparison of Fits	
DFn, DFd	2, 7
F	11.29
P value	0.006452
Best Fit Equation	Eq. 2

The F ratio, as well as the two degrees of freedom values, are explained in the next section. The P value is 0.0065. If the one-site model is correct, there is only a 0.07% 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%, conclude that the fit of the two-site model is significantly better, statistically, than the one-site model.

The F test is only strictly valid the simpler equation is a special case of the more complicated equation.

How the F test works

To understand the F test, you need to compare the sum-of-squares and degrees of freedom for each fit. Here are the results for the example of the previous section.

	Two-site	One-site	% Increase
Degrees of freedom	7	9	28.57%
Sum-of-squares	52330	221100	322.51%

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 7 degrees of freedom (12 data points minus 5 variables), the degrees of freedom increased 28.6%. If the one-site model were correct, you'd expect the sum-of-squares to also increase about 28.6% just by chance. In fact, the one-site had a sum-of-squares 322% higher than the two-site model. The percent increase in sum-of-squares was 11.29 times higher than the increase in degrees of freedom (322.51/28.57). The F ratio is 11.29.

More generally, if the simpler model is correct you expect the relative increase in the sum of squares 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.

$$F = \frac{(SS1 - SS2)/SS2}{(DF1 - DF2)/DF2}$$

That equation is more commonly shown in an equivalent form:

$$F = \frac{(SS1 - SS2)/(DF1 - DF2)}{SS2/DF2}$$

F ratios are always associated with a certain number of degrees of freedom for the numerator and a certain number of degrees of freedom for the denominator. This F ratio has DF1-DF2 degrees of freedom for the numerator, and DF2 degrees of freedom for the denominator.

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. Otherwise, conclude that there is no compelling evidence supporting model 2, so accept the simpler model (model 1).

Comparing fits to two sets of data (same equation)

You'll use a different approach to compare nonlinear regression curve fits depending on whether you are comparing curves from one experiment or pool the results of several experiments.

Compare best-fit values pooled from several experiments

The best way to compare best-fit values is to repeat the experiment several times, and then analyze the pooled data. The first step is to focus on what you really want to know. For dose-response curves, you may want to test whether the two EC₅₀ values differ significantly, whether the maximum responses differ, or both. With kinetic curves, you'll want to ask about differences in rate constants or maximum response. With other kinds of experiments, you may summarize the experiment in other ways, perhaps as the maximum response, the minimum response, the time to maximum, the slope of a linear regression line, etc. Or perhaps you want to integrate the entire curve and use area-under-the-curve as an overall measure of cumulative response.

Once you've summarized each curve as a single value, compare those values using a paired t test.

For example, below are the results of a binding study to determine receptor number (B_{max}). The experiment was performed three times with control and treated cells side-by-side. Each value is a B_{max} determined by nonlinear regression.

Experiment	Control	Treated
1	1234	987
2	1654	1324
3	1543	1160

Treat the B_{max} values determined from nonlinear regression just as you'd treat any measurement, and compare the two groups with a t test. Because control and treated cells were treated side-by-side to control for experiment-to-experiment variability, analyze the data using a paired t test. Enter the data into a new data table formatted for single Y values and no X values. Prism determines the B_{max} for each curve, but you have to compile all the B_{max} values into one table. Press Analyze, choose t tests from the list of statistical comparisons, and choose a paired t test.

Prism reports that the two-tailed P value is 0.0150, so the effect of the treatment on reducing receptor number is statistically significant. The 95% confidence interval of the decrease in receptor number ranges from 149.70 to 490.30 sites/cell.

These calculations were based only on the best-fit B_{max} values, ignoring all the other results calculated by the curve-fitting program. You may be concerned that you are not making best use of the data, since the number of points and replicates do not appear to affect the calculations. But they do contribute indirectly. You'll get more accurate results if you use more concentrations of ligand in each experiment, so the results of the experiments will be more consistent. If there really are differences between control and treated curves, you'll get a higher t ratio and a lower P value if you use more concentrations.

If you have three or more treatment groups, use repeated measures one-way ANOVA rather than a paired t test.

Compare two best-fit values from one experiment

There are several approaches to comparing two curves obtained in one experiment. If your data are linear, see "Comparing slopes and intercepts" on page 150. to learn about comparing linear regression lines. If you want

to compare Y values at each X, without fitting a curve, see "Two-way analysis of variance" on page 93 to learn about post tests following two-way ANOVA.

Prism can compare the best-fit values of two curves with a t test. A t test compares a difference with the standard error of that difference. That standard error can come by pooling several experiments (as in the previous approach) or you can use the standard error reported by nonlinear regression. For example, in the first experiment in Approach 1, Prism reported these results for B_{max} :

	Best-fit B_{max}	SE	df
Control	1234	98	14
Treated	987	79	14

Prism can compare these curves automatically. On the nonlinear regression dialog, check the option to use a t test to compare fits. In the results view (page) called Overview and comparison, Prism reports the comparison of B_{max} values. The P value is 0.06. If there really were no difference between B_{max} values, you'd see a difference this large or larger in 6% of experiments of this size. Using the conventional threshold of $P=0.05$, the difference between B_{max} values in this example is not statistically significant.

Prism performs the unpaired t test using this equation:

$$t = \frac{B_1 - B_2}{\sqrt{SE_1^2 + SE_2^2}} = \frac{1234 - 987}{\sqrt{98^2 + 79^2}} = \frac{247}{125.88} = 1.96$$

The numerator is the difference between best-fit values. The denominator is an estimate of the standard error of that difference, computed as the square root of the sum of the squares of the two standard error values. This is a reasonable estimate if the number of data points in the two curves is equal, or nearly so. If the sample sizes are very different, don't rely on this calculation.

Prism determines the two-tailed P value from t. Each fit in this example has 14 degrees of freedom (equal to the number of data points minus the number of variables fit). The t test in this example has 28 degrees of freedom, the sum of the degrees of freedom in each fit.

The validity of this comparison depends on the assumptions of the t test (see "Checklist. Is an unpaired t test the right test for these data?" on page 49). A key assumption is that the distribution of best-fit values (if you were

to repeat the experiment many times) follows a Gaussian distribution. The next section discusses this assumption.

The Gaussian assumption and comparison of curves

The use of a t test to compare best-fit values (pooling several experiments or within one experiment) depends on the assumption that the distribution of best-fit values follows a Gaussian distribution. If you were to repeat the experiment many times, the distribution of the best-fit values must follow a Gaussian distribution. With linear regression, this assumption is sure to be valid if your data obey all the assumptions of the analysis. With nonlinear regression, the best-fit values may not be Gaussian, even if the data follow all the assumptions. The next chapter discusses this point in detail. See "The distributions of best-fit values" on page 229.

Compare entire curves using a F test

The method described in the previous section requires that you focus on one variable that you consider most relevant. If you don't wish to focus on one variable, compare entire curves using the following approach. You'll have to do some calculations manually, as this approach is not built-in to Prism.

First, use Prism to fit each data set to a model. Total the sum-of-squares and degrees of freedom from the two fits. If you are comparing a curve under control conditions with a curve after some treatment, then add the sum-of-squares from the best-fit of control data with the sum-of-squares of the best-fit for treated data. Do the same with the df. Since these values are obtained by fitting the control and treated data separately, label these values, SS_{separate} and DF_{separate} . For our example, the sums-of-squares equal 1261 and 1496 so SS_{separate} equals 2757. Each experiment had 14 degrees of freedom, so DF_{separate} equals 28.

Now do some copying and pasting to combine the control and treated data set into one big data set on a new table. Simply append one data set under the other, and analyze the data as if all the values came from one experiment. Its ok that X values are repeated. Fit the same equation to this combined data set. Label the sum-of-squares from this fit SS_{combined} and the number of degrees of freedom DF_{combined} . For the example, SS_{combined} is 3164 and DF_{combined} is 30 (32 data points minus two variables fit by nonlinear regression).

You expect SS_{separate} to be smaller than SS_{combined} even if the treatment had no effect simply because the separate curves have more degrees of freedom. The question is whether the difference between SS values is greater than you'd expect to see by chance. To find out, compute the F ratio using the equation below, and then determine the corresponding P value (there are

$DF_{\text{combined}} - DF_{\text{separate}}$ degrees of freedom in the numerator and DF_{separate} degrees of freedom in the denominator.

$$F = \frac{(SS_{\text{combined}} - SS_{\text{separate}}) / (DF_{\text{combined}} - DF_{\text{separate}})}{SS_{\text{separate}} / DF_{\text{separate}}}$$

For the above example, $F = 2.067$ with 2 df in the numerator and 28 in the denominator. To find the P value, use GraphPad StatMate or type this formula into an empty cell in Excel = **FDIST(2.06,2,28)** . The P value is 0.1463.

The P value tests the null hypothesis that there is no difference between the control and treated curves overall, and any difference you observed is due to chance. If the P value were small, you would conclude that the two curves are different – that the experimental treatment altered the curve. Since this method compares the entire curve, it doesn't help you focus on which parameter(s) differ between control and treated (unless, of course, you only fit one variable). It just tells you that the curves differ overall. In this example, the P value was fairly large, so we conclude that the treatment did not affect the curves in a statistically significant manner.

The distributions of best-fit values

Why care about the distribution of best-fit values

With linear regression, the distribution of slopes and intercepts will always be Gaussian, so long as you haven't violated any assumption of the analysis. If you repeated the experiment many times, you'd get different slopes and intercepts in each experiment, and these would follow a Gaussian distribution. This means that the 95% confidence intervals of best-fit values can be interpreted at face value, and that you can compare slopes (or Y-intercepts) with a t test.

With nonlinear regression, a parameter may not be Gaussian, even if you meet all the assumptions of nonlinear regression. It depends on the equation you select. There is no general way to know whether a parameter follows a Gaussian distribution or not. But it matters. If a parameter does approximate a Gaussian distribution, then you can:

- Accept the 95% confidence intervals at face value. There is a 95% chance that the interval contains the true value. If a parameter does not follow a Gaussian distribution, there may be less than a 95% chance that the interval contains the true value.
- Perform a t test to compare two treatments, either pooling several experiments or using the result of one pair of experiments (see "Comparing fits to two sets of data (same equation)" on page 224).

Using simulations to determine the distribution of a parameters

The only way to determine the distribution of best-fit parameters is to simulate many sets of data and then look at the distribution of best-fit values following this general procedure:

1. Generate a simulated data set by adding random Gaussian error to the ideal value at each value of X. You'll need to choose a model (equation) to simulate, the range of X values, the number of points, and the amount of scatter.
2. Use nonlinear regression to fit a curve to the simulated data, using the same model. Record the best-fit values of the parameters.
3. Repeat steps 2-5 many (thousands) of times.
4. Construct a frequency histogram of the parameter values.
5. Repeat using a different form of the equation.

The procedure employed for generating large numbers of simulated data sets with random error is referred to as *Monte Carlo* analysis. Arthur Christopoulos has done this with equations commonly used in analyzing pharmacological data (Trends Pharmacol. Sci. 19:351-357, 1998), and he is a co-author of this chapter.

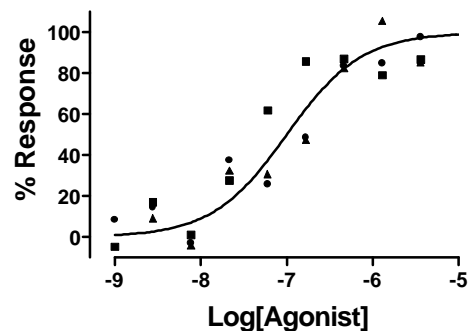
This procedure is explained in detail below (see "Detailed instructions for comparing parameter distributions" on page 234).

Example simulation 1. Dose-response curves.

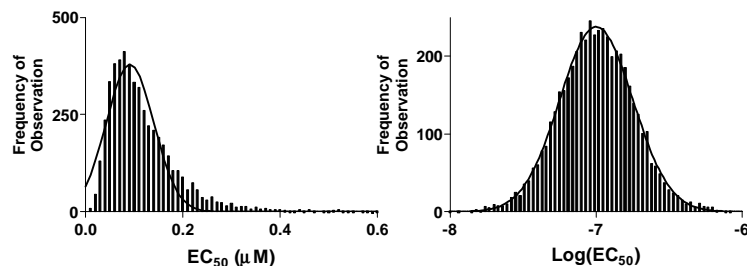
Dose-response curves. Which is more Gaussian, EC_{50} or $\log(EC_{50})$?

Scientists in many fields fit data to sigmoid dose-response curves. Is it better to find the best-fit value of EC_{50} or $\log(EC_{50})$? We used simulations to find out.

The simulated curves each had ten data points, equally spaced on a log scale from 1 nM to 10 μ M. The true curve had a bottom plateau at 0.0, a top plateau at 100, and an EC_{50} of 1 μ M. Random noise was added to each point, using a Gaussian distribution with a SD of 15, which simulates data in a system with lots of scatter. Three typical data sets are superimposed in the graph below.



Five thousand data sets were simulated and fit to the two alternative expressions of the dose-response equation. The distribution of the EC_{50} and $\log(EC_{50})$ values are shown below.



Clearly, the distribution of $\log(EC_{50})$ values is much closer to Gaussian. The normality test confirms this impression. The distribution of $\log(EC_{50})$ passes the normality test. The distribution of EC_{50} values fails the normality test with $P < 0.0001$.

How accurate are the confidence intervals of EC_{50} and $\log EC_{50}$?

If you express the equation in terms of EC_{50} , which is less Gaussian, only 91.20% of the 5000 "95%" confidence intervals contained the true value. In contrast, 94.20% of the "95%" confidence intervals of $\log(EC_{50})$ contained the true value. This is one reason to prefer an equation written in terms of $\log EC_{50}$.

The confidence intervals reported by Prism (and most other nonlinear regression programs) are always symmetrical around the best-fit value. When computing the confidence interval, Prism just applies the equation (see "Confidence intervals of best-fit values" on page 210) and doesn't have any common sense about whether the results make sense. If you express the results of these simulations as an EC_{50} value, in fact most of the results

do not make sense. In 78% of the simulations, the lower confidence limit was negative! Of course negative concentrations are impossible, so you would call the lower limit zero in these cases. This means that in most of our simulations, the 95% confidence interval gives you no information at all about the lower limit of the EC_{50} and only gives an estimate of the upper limit. In contrast, if you express the dose-response equation in terms of the $\log(EC_{50})$, it is impossible for the confidence interval to include negative values of the EC_{50} . The 95% CI of $\log EC_{50}$ gives you information about both the lower and upper limit of the value.

These simulations show a clear advantage to expressing the dose-response equation in terms of $\log(EC_{50})$. This makes sense. Since the concentrations of drug are equally spaced on a log scale, it makes sense that the uncertainty of the $\log(EC_{50})$ will be symmetrical, but the uncertainty of the EC_{50} will not be. Prism always reports symmetrical confidence intervals of best-fit values. If the true uncertainty is not symmetrical, then the 95% confidence interval reported by Prism will not be very useful.

Example simulation 2. Exponential decay.

Many biological and chemical events follow an exponential decay model. For more information on this model, see "Example model 2. Exponential decay" on page 158. We'll compare three ways to express this model. In all cases, one of the parameters is the starting point, which we will call Start. The second parameter quantifies how rapidly the curve decays. We will compare three ways to express this value, as a rate constant in units of inverse time, as a time constant in units of time, or as a $\log(\text{rate constant})$. The three equations are:

$$Y = \text{Start} \cdot e^{-k_{\text{rate}} \cdot t}$$

$$Y = \text{Start} \cdot e^{-(1/k_{\text{time}}) \cdot t}$$

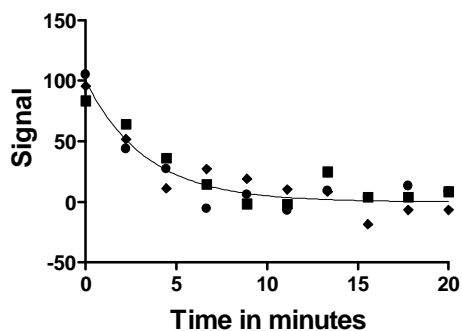
$$Y = \text{Start} \cdot e^{-10^{\log(k_{\text{rate}})} \cdot t}$$

Which distribution is closer to Gaussian, rate constants or time constants?

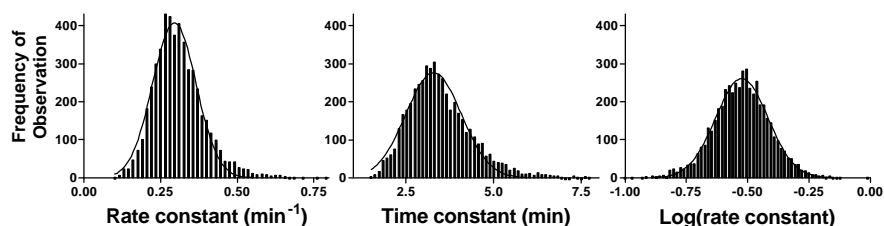
Is it better to express the exponential decay equation in terms of rate constant, time constant, or $\log(\text{rate constant})$? We'll answer the question by simulating data.

First, we need to choose some parameters. We chose a curve that starts at $Y = 100$ and decays exponentially towards 0 with a rate constant (k_{off}) of 0.3 min^{-1} and a half-life of a bit more than 2 minutes ($\ln(2)/k_{\text{off}}$). Our simulations generated 10 data points equally spaced between 0 and 20 minutes, adding

Gaussian random error with a standard deviation of 10. The graph below shows three sample simulations.



We simulated 5000 sets of data, and fit each data set to the exponential decay model expressed in three ways. The distribution of the rate constant, time constant, and log(rate constant) are shown in the following figures, which also superimpose ideal Gaussian distributions.



At first glance, all three distributions look roughly Gaussian. Looking more carefully, you can see that the distribution of time constants is skewed to the right. Careful scrutiny reveals that the rate constant distribution is also a bit skewed. These impressions can be confirmed by a normality test. The results are shown in the following table.

Model	Rate constant	Time constant	Log(Rate constant)
KS	0.06359	0.07169	0.01339
P value	P < 0.0001	P < 0.0001	P > 0.10

The KS value is the largest discrepancy between the actual cumulative distribution and an ideal cumulative Gaussian distribution (expressed as fractions). See "The results of normality tests" on page 29. None of the distributions are far from Gaussian. The distribution of time constants is the furthest from Gaussian; the distribution of log(rate constant) is closest. The P value answers the following question: If the true distribution is Gaussian,

what is the chance of obtaining a KS value as large as, or larger than, we observed. The distribution of both rate constants and time constants deviated significantly from the Gaussian ideal, while the distribution of log(rate constants) is indistinguishable from Gaussian.

Because we simulated so many data sets, the KS test has the power to detect even modest deviations from a Gaussian distribution.

How accurate are the confidence intervals of time constants and rate constants?

Before understanding the complexities of the confidence intervals reported by nonlinear regression, first review the meaning of confidence intervals reported by linear regression.

For example, assume that a linear regression presented the CI of the slope as 8.2 to 11.3. If you can accept all the assumptions of the analysis, this means that you can be 95% certain this range includes the true slope. More precisely, if you analyze many data sets, you'd expect that 95% of the confidence intervals will contain the true value, and 5% will not. When analyzing a particular experiment, the true value is unknown so you can't know whether or not the confidence interval includes the true value. All you can know is that there is a 95% chance that the interval contains the true value.

With nonlinear regression, the situation is not so simple. The confidence intervals reported by Prism (and virtually all other nonlinear regression programs) are based on some mathematical simplifications. They are called "asymptotic" or "approximate" confidence intervals. They are calculated assuming that the equation is linear, but are applied to nonlinear equations. This simplification means that the intervals can be too optimistic. While labeled 95% confidence intervals, they may contain the true value less than 95% of the time.

Using simulated data, we can ask how often the reported 95% confidence interval contains the true value. To do this, we changed the script to also save the SE of the best-fit value. We then imported the best-fit values and standard errors into Microsoft Excel, computed the high and low limit of the confidence interval (see "Confidence intervals of best-fit values" on page 210) and asked whether or not each interval contained the true value. When analyzing data, you don't know the true value so can't know whether the confidence interval contains the true value or not. With simulated data, you know the true value, so can answer that question. The results are:

Model	Fraction of "95% confidence intervals" that contain the true value.
Rate constant	93.72%
Time constant	94.44%
Log(rate constant)	94.94%

No matter how we expressed the model, the confidence intervals contained the true value almost 95% of the time. The difference between 95% confidence and 93% confidence is unlikely to alter your interpretation of experimental results.

Detailed instructions for comparing parameter distributions

The two examples above showed the general approach to comparing distributions. Here are step-by-step instructions for comparing parameter distributions using Prism, using the exponential decay example.

1. Generate simulated data with random error

1. Create a new file. From an empty data table, click Analyze, choose built-in analyses, and then choose to simulate a theoretical curve.
2. Click More equations, and enter the first equation, for example.

$$Y = \text{Start} * \exp(-K_{\text{rate}} * X)$$
3. Enter the ideal values. Choose values that are similar to those you would expect to see with real data. For this example, we enter Start = 100 and $K_{\text{rate}} = 0.3$. These values will simulate a dissociation curve that starts at 100 units and decays exponentially towards 0 with a rate constant (k_{off}) of 0.3 min^{-1} and a half-life of a bit more than 2 minutes ($\ln(2)/k_{\text{off}}$).
4. Choose to generate 10 data points (line segments) starting at 0 and ending at 20 (minutes).
5. Check the option box to add Gaussian "noise" with a standard deviation (SD) of 10. Choose a value that is similar to the scatter you expect to see with real data.

Click ok and Prism will generate a simulated data set. Click on the Graphs folder tab to see a graph of these data (plotted as line segments). Change to data points by clicking Change... Symbols and lines. Now click Change.. analysis parameters to bring up the dialog for simulating curves. Don't change any of the parameters, but simply click ok. Prism will generate new random values, and you'll see that the graph has changed.

2. Analyze the simulated data

The next step is to analyze the simulated data to determine the best-fit values. Follow these steps.

1. While still on the Results page with the simulated data, click Analyze, and select nonlinear regression.
2. Select the same built-in equation used to generate the theoretical points. Click Initial values and enter the ideal values. Start = 100, $K_{\text{rate}} = 0.3$. Click ok twice to fit the curve.
3. Note the best-fit values.

3. Simulate thousands of data sets using a Prism script

Create a script to instruct Prism to generate and analyze 5000 sets of data, and record the best-fit values. To learn details about scripting, read the chapter on scripts in the Prism User's Guide.

Using Notepad or some other text editor, create a file with the following script. Enter the lines shown in the left column. The right column explains what each command does, and should not be typed. Give the file a name with the extension pzc.

Script line	Explanation
Shortlog	Don't add a line to the log for each iteration.
Setpath c:\sims	Use this folder. You'll want to change the folder name from "sims" to something else.
Open disskin.pzm	Open a Prism file. You'll want to change the file name.
OpenOutput kinfit.txt	Create a text file to hold the best-fit values.
ForEach 5000	Loop 5000 times.
Goto R 1	Go to the first results page.
Regenerate	Create new random numbers.
Goto R 2	Go to the second results page.
WCell 4,1	Write the value in the fourth row of the first column into the output file. This is the best-fit value of the rate constant.
Next	Loop again.

Run this script from Prism, by choosing the Run Automation File command from the Tools menu. Selecting the script you just created, and click Run to

execute it. Depending on the speed of your computer, this task should take a few minutes to execute.

4. Create a frequency distribution of the best-fit values

Now let's look at the distribution of best-fit values.

1. Create a new data table, formatted with no X column and a single column of values for Y.
2. Click in the first cell of the first data column on the table.
3. From the File menu, choose to Import. Choose the file with the best-fit values (the example script created kinfit.txt).
4. Click ok to accept all the defaults in the Import dialog.
5. You'll see the 5000 best-fit values in the data table.
6. To construct a frequency distribution of the data, click Analyze and select the Frequency distribution option in the Statistical Analyses section. You may need to adjust the bin width.
7. Inspect the graph. To show the graph with histogram "spikes", choose one of the bottom four choices of symbol shape.
8. To assess whether the best-fit values follow a Gaussian distribution, choose Analyze and then select Column Statistics from the Statistical Analyses option. From the Column Statistics dialog, choose to test whether the distribution is Gaussian.
9. To superimpose a Gaussian distribution, press Analyze and choose nonlinear regression. Choose the equation for a Gaussian distribution, the last choice in the list of classic equations.

5. Repeat with other forms of the equation

You can now repeat all the steps, but enter the equation in a different form. For this example, we switched from a rate constant to a time constant or $\log(\text{rate constant})$.

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 site 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, discussed in the next three chapters.

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, B_{max} , and the ligand affinity, K_d . See "Analyzing saturation radioligand binding data" on page 249.

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. See "Analyzing competitive binding data" on page 263.

Kinetics experiments measure binding at various times to determine the rate constants for radioligand association and dissociation. See "Analyzing kinetic binding data" on page 287.

Prism makes it easy to analyze and display data from all three kinds of experiments.

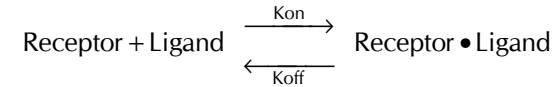
For more information on analysis of radioligand binding data, see:

LE Limbird, Cell surface receptors: A short course on theory and methods, Martinus Nijhoff Publishers, second edition, 1996.

HI Yamamura, et al, Methods in Neurotransmitter receptor analysis, Raven Press, 1990.

Law of mass action

Analysis of radioligand binding experiments is based on a simple model, called the law of mass action. This model assumes that binding is reversible.



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 is:

Number of binding events per unit of time = $[\text{Ligand}] \cdot [\text{Receptor}] \cdot k_{\text{on}}$.

Once binding has occurred, the ligand and receptor remain bound together for a random amount of time. 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. The rate of dissociation is:

Number of dissociation events per unit time = $[\text{ligand} \cdot \text{receptor}] \cdot k_{\text{off}}$.

After dissociation, the ligand and receptor are the same as at they were before binding. If either the ligand or receptor is 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:

$$[\text{Ligand}] \cdot [\text{Receptor}] \cdot k_{\text{on}} = [\text{Ligand} \cdot \text{Receptor}] \cdot k_{\text{off}}$$

Rearrange that equation to define the equilibrium dissociation constant K_d .

$$\frac{[\text{Ligand}] \cdot [\text{Receptor}]}{[\text{Ligand} \cdot \text{Receptor}]} = \frac{k_{\text{off}}}{k_{\text{on}}} = K_d$$

The K_d has a meaning that is easy to understand. Set $[\text{Ligand}]$ equal to K_d in the equation above. The K_d terms cancel out, and you'll see that $[\text{Receptor}] / [\text{Ligand} \cdot \text{Receptor}] = 1$, so $[\text{Receptor}]$ equals $[\text{Ligand} \cdot \text{Receptor}]$. Since all the receptors are either free or bound to ligand, this means that half the receptors are free and half are bound to ligand. In other words, when the concentration of ligand equals the K_d , half the receptors will be occupied at equilibrium. If the receptors have a high affinity for the ligand, the K_d will be low, as it will take a low concentration of ligand to bind half the receptors.

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.

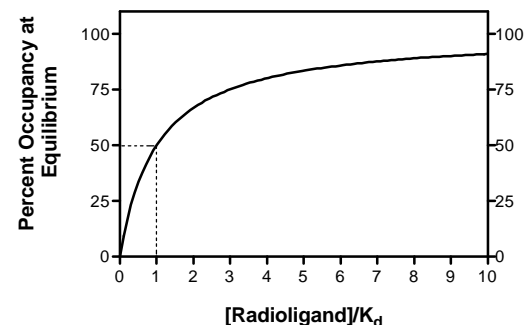
$$\text{Fractional occupancy} = \frac{[\text{Ligand} \cdot \text{Receptor}]}{[\text{Total Receptor}]} = \frac{[\text{Ligand} \cdot \text{Receptor}]}{[\text{Receptor}] + [\text{Ligand} \cdot \text{Receptor}]}$$

This equation is not useful, because you don't know the concentration of unoccupied receptor, [Receptor]. A bit of algebra creates a useful equation. See "Example model 3. Equilibrium binding" on page 160.

$$\text{Fractional occupancy} = \frac{[\text{Ligand}]}{[\text{Ligand}] + K_d}$$

This equation assumes equilibrium. To make sense of it, think about a few different values for [Ligand].

[Ligand]	Fractional Occupancy
0	0%
$1 K_d$	50%
$4 K_d$	80%
$9 K_d$	90%
$99 K_d$	99%



Note that when $[\text{Ligand}] = K_d$, 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:

- All receptors are equally accessible to ligands.
- Receptors are either free or bound to ligand. It doesn't allow for more than one affinity state, or states of partial binding.
- Binding does not alter the ligand or receptor.
- Binding is reversible.

Despite its simplicity, the law of mass action has proven to be very useful in describing many aspects of receptor pharmacology and physiology.

Nonspecific binding

In addition to binding to receptors of interest, radioligands may also bind to other sites. Binding to the receptor of interest is called *specific binding*, while binding to the other sites is called *nonspecific binding*. This means that nonspecific binding can represent several phenomena:

- In most cases, the bulk of nonspecific binding represents some sort of interaction of the ligand with membranes. The molecular details are unclear, but nonspecific binding depends on the charge and hydrophobicity of a ligand – but not its exact structure.
- Nonspecific binding can also be binding to receptors and transporters not of interest to the investigator, for example binding of epinephrine to serotonin receptors or metabolic enzymes.
- Nonspecific binding can also be binding to the filters used to separate bound from free ligand.

Nonspecific binding is usually (but not necessarily) 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.

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.

Which unlabeled drug should you use? The obvious answer is to use the same compound as the radioligand, but unlabeled. In many cases, this is necessary, as no other drug is known to bind to the receptors. But most investigators avoid using the same compound as the hot and cold ligand and prefer to define nonspecific binding with a drug chemically distinct from the radioligand.

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 binding. If you are studying a well-characterized receptor, a useful rule-of-thumb is to use the unlabeled compound at a concentration equal to 100 times its K_d for the receptors, or 100 times the highest concentration of radioligand, whichever is higher.

Ideally, you should get the same results defining nonspecific binding with a range of concentrations of several drugs, and you should test this when possible. In many assay systems, nonspecific binding is only 10-20% of the total radioligand binding. If the nonspecific binding makes up more than half of the total binding, you'll find it hard to get quality data. If your system has a lot of nonspecific binding, try different kinds of filters, a larger volume of washing buffer, warmer washing buffer, or a different radioligand.

Ligand depletion

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 binds 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. 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 possible if you use centrifugation or equilibrium dialysis, but is quite 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 few chapters explain several such methods.

Calculations with radioactivity

Efficiency of detecting radioactivity

It is not possible to detect every radioactive disintegration. The fraction of radioactive disintegrations detected by your counter is called *efficiency*. Determine efficiency by counting a standard sample under conditions identical to those used in your experiment.

It is relatively easy to detect gamma rays emitted from isotopes such as ^{125}I , so efficiencies are usually over 90%, depending on the geometry of the counter. The detector doesn't entirely surround the tube, so a small fraction of gamma rays (photons) miss the detector.

With ^3H , the efficiency of counting is much lower, often about 40%. 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 can 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. If the electron has insufficient energy, it is not captured by the fluor and is not detected.

Since the decay of a large fraction of tritium atoms does not lead to a detectable number of photons, the efficiency of counting is much less than 100%. This low efficiency is a consequence of the physics of decay, and you can't increase the efficiency much using a better scintillation counter or a better scintillation fluid. However, poor technique can reduce the efficiency. Electrons emitted by tritium decay have very little energy, so can only be detected when they immediately encounter a fluor molecule.

Radioactivity trapped in tissue chunks won't be detected. Nor will radioactivity trapped in an aqueous phase not well mixed into the scintillation fluid.

Specific radioactivity

When you buy radioligands, the packaging usually states the specific radioactivity as Curies per millimole (Ci/mmol). Since you measure counts per minute (cpm), the specific radioactivity is more useful if you change the units to be in terms of cpm rather than Curie. 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 × 10¹² disintegrations per minute. The following equation converts Z Ci/mmol to Y cpm/fmol when the counter has an efficiency (expressed as a fraction) equal to E.

$$Y \frac{\text{cpm}}{\text{fmol}} = Z \frac{\text{Ci}}{\text{mmole}} \cdot 2.22 \times 10^{12} \frac{\text{dpm}}{\text{Ci}} \cdot 10^{-12} \frac{\text{mmole}}{\text{fmole}} \cdot E \frac{\text{cpm}}{\text{dpm}} = Z \cdot 2.22 \cdot E$$

$$Y \frac{\text{cpm}}{\text{fmol}} = Z \frac{\text{Ci}}{\text{mmole}} \cdot 2.22 \cdot E$$

In some countries, radioligand packaging states the specific radioactivity in Gbq/mmol. A Becquerel, abbreviated bq, equals one radioactive disintegration per second. A Gbq is 10⁹ disintegration per second. To convert from Gbq/mmol to cpm/fmol, use this equation:

$$Y \frac{\text{cpm}}{\text{fmol}} = Z \frac{\text{Gbq}}{\text{mmole}} \cdot \frac{10^9 \text{ bq}}{\text{Gbq}} \cdot 60 \frac{\text{dpm}}{\text{bq}} \cdot 10^{-12} \frac{\text{mmole}}{\text{fmole}} \cdot E \frac{\text{cpm}}{\text{dpm}} = Z \cdot 0.06 \cdot E$$

$$Y \frac{\text{cpm}}{\text{fmol}} = Z \frac{\text{Ci}}{\text{mmole}} \cdot 0.06 \cdot E$$

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 the equation below. 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).

$$\text{Concentration in pM} = \frac{C \text{ cpm} / Y \text{ cpm/fmol}}{V \text{ ml}} \cdot \frac{0.001 \text{ pmol/fmol}}{0.001 \text{ liter/ml}} = \frac{C/Y}{V}$$

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. Therefore decay follows an exponential decay as explained in "Example model 2. Exponential" on page 158. If you start with N₀ radioactive atoms, the number remaining at time t is:

$$N_t = N_0 \cdot e^{-K_{\text{decay}} \cdot t}$$

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_{1/2}) is the time it takes for half the isotope to decay. Half-life and decay rate constant are related by this equation:

$$t_{1/2} = \frac{\ln(2)}{K_{\text{decay}}} = \frac{0.693}{K_{\text{decay}}}$$

This table below 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 ¹²⁵I and ³²P. 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	28.7 Ci/mmol
¹²⁵ I	59.6 days	0.0116/day	2190 Ci/mmol
³² P	14.3 days	0.0485/day	9128 Ci/mmol
³⁵ S	87.4 days	0.0079/day	1493 Ci/mmol

You can calculate radioactive decay from a date where you (or the manufacturer) knew the concentration and specific radioactivity using this equation.

$$\text{Fraction Remaining} = e^{-K_{\text{decay}} \cdot \text{Time}}$$

For example, after ¹²⁵I 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.

Counting errors and the Poisson distribution

The decay of a population of 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:

$$95\% \text{ Confidence Interval: } (C - 1.96\sqrt{C}) \text{ to } (C + 1.96\sqrt{C})$$

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 number of counts ranges approximately between 80 and 120 (using the equation here) or between 81.37 and 121.61 (using the exact equation programmed into 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 the advantages of counting your samples for a longer time. For example, the table below 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

The table below shows percent error as a function of the number of counts. Percent error is defined as the width of the confidence interval divided by the number of counts.

Counts	Percent Error
50	27.72%
100	19.60%
200	13.86%
500	8.77%
1000	6.20%
2000	4.38%
5000	2.77%
10000	1.96%
25000	1.24%
50000	0.88%
100000	0.62%
C	$100 \cdot \frac{1.96 \cdot \sqrt{C}}{C}$

The GraphPad radioactivity web calculator

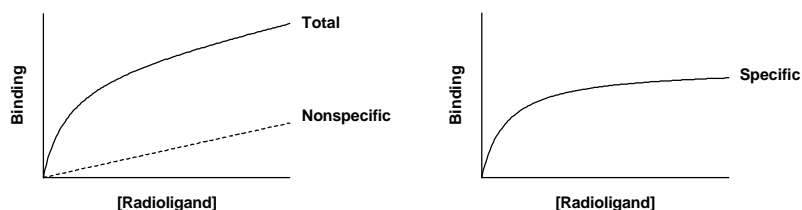
GraphPad Software provides a free radioactivity calculator on our web site at <http://www.graphpad.com/www/radcalc.htm>

Use it to perform seven common calculations.

Calculation	Description
Isotope decay	Calculates radioactive decay during a specified number of days. Select one of the common isotopes, or enter the half-life of another isotope.
Conc. of stock	<p>Enter mCi/ml and Ci/mmol, which should be on the label. If you are using a molecule labeled with ^{125}I, the specific activity equals 2200 Ci/mmol if each molecule is labeled with one iodine.</p> <p>Also enter the percent of the original isotope remaining (calculated above). The calculations assume that the decay product is not biologically active, so the concentration of stock that is biologically active decreases over time.</p>
Dilution of stock	Enter the concentration in your stock solution, after accounting for decay. Also enter the concentration and volume you want. The result is the volume of stock you need to use.
Specific activity (cpm/fmol)	<p>Enter the specific radioactivity as Ci/mmol which should be on the label. If you are using a molecule labeled with ^{125}I, the specific activity equals 2200 Ci/mmol if each molecule is labeled with one iodine.</p> <p>Also enter the counter efficiency - the fraction of radioactive disintegrations that are detected. The efficiency depends on the isotope and instrumentation. With low energy isotopes such as tritium, the efficiency also depends on the experimental details such as the choice of scintillation fluid, the amount of water in the sample, and the presence of any colored substances in the sample.</p>
Cpm to fmol/mg	Enter the specific radioactivity as cpm/fmol, the number of cpm counted, and the protein content of the sample in mg. The result is the number of binding sites in fmol/mg protein.
Cpm to sites/cell	Enter the specific radioactivity as cpm/fmol, the number of cpm counted, and the cell count. The result is the number of binding sites per cell.
Cpm to nM	Enter the specific radioactivity as cpm/fmol, the number of cpm counted, and the volume counted. The result is the concentration of radioligand in nM.

Analyzing saturation radioligand binding data

Introduction to saturation binding experiments



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

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. 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 nonspecific binding 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.
7. 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. Instead, you should fit total binding only as explained in "Analyzing total binding data" on page 251.

Fitting a curve to determine B_{max} and K_D

Analyzing specific binding data

To fit a binding curve:

1. Start from a table where X is the 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.

- From the curves and regression choices, choose Nonlinear regression.
- On the nonlinear regression dialog, choose One 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 (commonly cpm, sites/cell or fmol/mg). K_d is expressed in the same units as the X values, usually nM or pM. If the K_d is low, the affinity is high.

Analyzing total binding data

In many systems, nonspecific binding is proportional to radioligand concentration over the concentration range used in binding studies. If this is true for your system, you can determine K_d and B_{\max} by fitting total binding only. Nonspecific binding is determined via curve fitting of total binding, and is not determined experimentally.

To fit total binding, use this user-defined equation:

```
Specific = X*Bmax / (Kd + X)
Nonspecific = NS*X
Y = Specific + Nonspecific
```

Variable	Comment
X	Radioligand concentration. Usually expressed in nM or pM.
Y	Total binding in cpm, sites/cell, fmol/mg or some other units
Bmax	The maximum specific binding to be fit. A reasonable initial value might be that B_{\max} equals 0.5 times YMAX .
Kd	The equilibrium dissociation constant. A reasonable initial value might be 0.2*YMAX
NS	Multiply NS times X to calculate nonspecific binding at that concentration of radioligand. A reasonable rule for initial value is to set NS equal to 0.2 * (YMAX-YMIN)/(XMAX-XMIN).

To get useful results with this approach, you will need high quality data and at least ten data points including some well above the ligand K_d .

Determining K_d and B_{\max} for two classes of binding sites

If the radioligand binds to two classes of binding sites, fit the specific binding data to the two-site binding equation (found in Prism's list of built-in equations). This equation simply sums the binding to two sites, each with its own B_{\max} and K_d .

$$Y = \frac{B_{\max 1} \cdot X}{K_{d1} + X} + \frac{B_{\max 2} \cdot X}{K_{d2} + X}$$

This equation assumes that the radioligand binds to two independent noninteracting binding sites, and that the binding to each site follows the law of mass action. To compare the one-site and two-site fits, see "Comparing the fits of two models " on page 221.

You will only get meaningful results from a two-site fit if you have ten or more (preferably a lot more) data points spaced over a wide range of radioligand concentrations. Be sure that you measure binding at radioligand concentrations below the high-affinity K_d and above the low affinity K_d .

Checklist for saturation binding

When evaluating results of saturation binding analyses, ask yourself these questions:

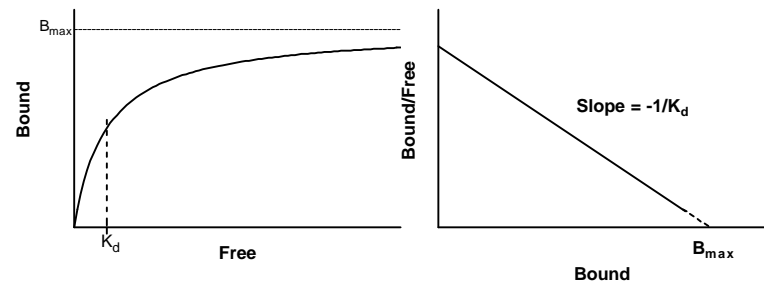
Question	Comment
Did only a small fraction of the radioligand bind?	The analysis assumes that the free concentration is almost identical to the concentration you added. You can test this by comparing the total cpm that bound to the total cpm added to the tube. If more than 10% of the ligand bound (at any ligand concentration) then the standard analysis won't work. Either change the experimental protocol (increase the volume) or use a method that accounts for depletion of radioligand – see "Analyzing saturation binding with ligand depletion" on page 257.
Did the binding equilibrate?	The tubes with the lowest concentration of radioligand take the longest to equilibrate. So test equilibration time with a low concentration of radioligand.

Did you use high enough concentrations of radioligand?	Calculate the ratio of the highest radioligand concentration you used divided by the K_d reported by the program (both in nM or pM). The highest concentration should be at least 10 times the K_d , so that occupancy exceeds 90%.
Is the B_{max} reasonable?	Typical values for B_{max} are 10-1000 fmol binding sites per milligram of membrane protein, 100-10000 sites per cell or 1 receptor per square micron of membrane. If you use cells transfected with receptor genes, then the B_{max} may be many times larger than these values.
Is the K_d reasonable?	Typical values for K_d of useful radioligands range between 10 pM and 10 nM. If the K_d is much lower than 10 pM, the dissociation rate is probably very slow and it will be difficult to achieve equilibrium. If the K_d is much higher than 10 nM, the dissociation rate will probably be fast, and you may be losing binding sites during separation of bound ligand from free radioligand.
Are the standard errors too large? Are the confidence intervals too wide?	Divide the SE of the B_{max} by the B_{max} , and divide the SE of the K_d by the K_d . If either ratio is much larger than about 20%, look further to try to find out why.
Is the nonspecific binding too high?	Divide the nonspecific binding at the highest concentration of radioligand by the total binding at the highest concentration. Nonspecific binding should usually be less than 50% of the total binding.

Scatchard plots

What is a Scatchard plot?

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), shown in the right panel below.



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, the Scatchard transformation distorts the experimental error, and thus violates several assumptions of linear regression (see "Avoid Scatchard, Lineweaver-Burke and similar transforms" on page 142). The B_{max} and K_D values you determine by linear regression of Scatchard transformed data may be far from their true values.

Tip. You should analyze saturation binding data with nonlinear regression **not** with Scatchard plots. Use Scatchard plots to display data, not to analyze data.

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 equals $-1/K_D$).

Transforming data to create a Scatchard plot

Prism cannot plot a Scatchard plot automatically, but it is very easy to transform the data to create a Scatchard plot. Use the same instructions for one- and two-site binding.

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 and choose built-in analyses.
3. From the data manipulation section, choose Transforms.
4. If your table has more than one data set, choose to analyze selected data sets only and choose one data set. The Scatchard transform can only be used with one data set at a time.
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 new graph.

If you do Scatchard transforms frequently, save these steps into a method or template. See "Analyzing repeated experiments" on page 353.

Plotting the line that corresponds to nonlinear regression analyses

If there is one class of receptors, the Scatchard plot will be linear. Some people use linear regression to draw a line through the points. To do this, start from either the graph or a table of the Scatchard transformed data. Click the analyze button and choose linear regression. You may need to change the limits of the regression line to create an attractive graph.

The linear regression line should NOT be used to analyze the data. The X-intercept of the regression line will be near the B_{max} and the negative inverse of the slope will be near the K_d . However, the B_{max} and K_d values determined directly with nonlinear regression will be more accurate.

It isn't hard to draw the Scatchard line that corresponds to the nonlinear regression determination of B_{max} and K_d . The discussion below assumes that the "bound" units for the Y axis are the same units used for the X-axis and in which you want to express B_{max} (sites/cell or fmol/mg.), and the "free" units are the same as the units you want to use for the K_d (nM or pM). Since the X intercept of the Scatchard line is the B_{max} , the Scatchard line ends at $X = B_{max}$, $Y = 0$. Since the slope of the Scatchard line equals $-1/K_d$, the Y-intercept equals the B_{max} divided by the K_d . So the Scatchard line begins at $X = 0$, $Y = B_{max}/K_d$.

To create a Scatchard line corresponding to the nonlinear regression fit:

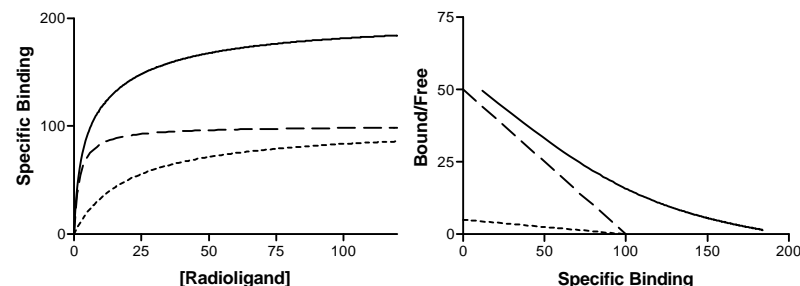
1. Create a new data table, with numerical X values and single Y values.
2. Into row 1 enter $X = 0$, $Y = B_{max}$ (previously determined by nonlinear regression).
3. Into row 2 enter $X = B_{max}/K_d$ (both previously determined by nonlinear regression) and $Y = 0$.
4. Note the name of this data table. Perhaps rename to something appropriate.
5. Go to the Scatchard graph.
6. Press Change, then Data on graph.
7. Add the new data table to the graph.
8. Press Change, then Symbols and lines.
9. Drop down the list of data sets, and select the one you noted in step 4.
10. Choose to plot no symbols, but to connect with a line.

Scatchard plots of binding to two sites

The appearance of a two-site Scatchard plot

The left panel below shows binding of a radioligand to two independent binding sites present in equal concentrations, but with a tenfold difference in K_d . The two individual curves are shown as dotted and dashed curves. When you do the experiment, you can't observe the individual components, but observe the sum, which is shown as a solid curve. Note that this curve is not obviously biphasic.

The right panel shows the same data plotted on a Scatchard plot. The binding to each receptor is shown as a straight line (dotted, or dashed). The total binding, plotted on a Scatchard plot, is curved. Note that the two lines that represent binding to each type of receptor are NOT the asymptotes of the curve.



Graphing data on a two-site Scatchard plot

See "Transforming data to create a Scatchard plot" on page 254. You don't need to do anything differently when there are two kinds of receptors.

Graphing the two lines of a two-site Scatchard plot

To plot the two straight lines that correspond to the nonlinear regression fit, adapt the instructions for plotting a Scatchard plot for one-site binding. See "Plotting the line that corresponds to nonlinear regression analyses" on page 255. Create a new data table that defines the two lines as shown below, using B_{\max} and K_d values determined by nonlinear regression.

X	A	B
0	$B_{\max 1}/K_{d1}$	
$B_{\max 1}$	0	
0		$B_{\max 2}/K_{d2}$
$B_{\max 2}$		0

Go to the graph of the Scatchard transformed data and add the new table to that graph. Use the Symbols dialog to plot the two data sets from the table using connecting lines but no symbols.

Graphing the curve on a two-site Scatchard plot

If the radioligand binds to two binding sites, the Scatchard plot will be curved but you cannot create the curve using nonlinear regression. This approach is not appropriate, because the specific binding data appear on the X-axis and as part of the Y-axis, and this violates a major assumption of nonlinear regression.

You should determine K_d and B_{\max} of both receptor types using nonlinear regression to fit the specific binding data. See "Determining K_d and B_{\max} for two classes of binding sites" on page 251. Then transform the resulting curve to Scatchard axes. Follow these steps:

1. Use nonlinear regression to fit the specific binding data to a two-site binding curve. While on the nonlinear regression parameters dialog, click the Output options button and check the option box to view the table of XY coordinates defining the curve.
2. Go to the nonlinear regression results. Drop the view list and select Curve.
3. Click Analyze, and choose Transform from the list of data manipulations.

4. To do a Scatchard transformation, check the option box to interchange X and Y values. Also check the box to transform Y and choose the Y transform: $Y = X/Y$. Do not check the box to make a new graph. The results will be the best-fit curve transformed to fit Scatchard axes.
5. Make a note of the name of the sheet with the Scatchard transformed curve data. Consider renaming the sheet.
6. Go to the graph of the Scatchard transformed data.
7. Click Change and choose Data on graph.
8. Add the data set that contains the Scatchard transformation of the best-fit curve (created in step 4).
9. Click Change and choose Symbols and lines.
10. Drop the list of data sets, and choose the Scatchard transformation of the best-fit curve.
11. Choose to plot no symbols, but to connect with a line.

Analyzing saturation binding with ligand depletion

The standard methods of analyzing saturation binding assume that a tiny 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.

In some experimental situations, where the receptors are present in high concentration and have a high affinity for the ligand, that assumption is not true. A large fraction of the radioligand binds to receptors, so the free concentration added is quite a bit lower than the concentration you added. 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 increasing the volume of the assay without changing the amount of tissue. The drawback is that you'll need more radioligand, which is usually expensive or difficult to synthesize.

If you can't avoid radioligand depletion, you need to account for the depletion in your analyses. The obvious 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 affect the free ligand concentration you calculate.

Error in Y would affect X, which violates an assumption of nonlinear regression.

- 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 equation that defines total binding as a function of added ligand, accounting for nonspecific binding and 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 equations shown below are not exactly the same as in Swillens' paper, but the ideas are the same.

From the law of mass action, total binding follows this equation.

$$\text{Total Binding} = \text{Specific} + \text{Nonspecific}$$

$$\text{Total Binding} = \frac{B_{\text{max}} \cdot [\text{Free Ligand}]}{K_d + [\text{Free Ligand}]} + [\text{Free Ligand}] \cdot \text{NS}$$

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.

This equation is not useful, because you don't know the concentration of free ligand. What you know is that the free concentration of ligand equals the concentration you added minus the concentration that bound (specific and nonspecific). Defining X to be the amount of ligand added and Y to be total binding, the system is defined by two equations:

$$Y = \frac{B_{\text{max}} \cdot [\text{Free Ligand}]}{K_d + [\text{Free Ligand}]} + [\text{Free Ligand}] \cdot \text{NS}$$

$$[\text{Free Ligand}] = X - Y$$

Combining the two equations:

$$Y = \frac{B_{\text{max}} \cdot (X - Y)}{K_d + (X - Y)} + (X - Y) \cdot \text{NS}$$

X, Y and B_{max} are expressed in units of cpm. To keep the equation consistent, therefore, K_d must also be converted to cpm units (the number of cpm added to each tube when the total concentration equals the K_d).

You cannot enter that equation into Prism for nonlinear regression because Y appears on both sides of the equal sign. But simple algebra rearranges it into a quadratic equation. The solution is shown below 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=Kd nM * 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)

```

Determining nonspecific binding experimentally in saturation binding experiments with ligand depletion

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 investigators 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.

$$\text{Nonspecific binding} = Y = \text{NS} \cdot [\text{Ligand}]$$

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:

$$Y = (X - Y) \cdot \text{NS}$$

$$Y = X \cdot \frac{\text{NS}}{\text{NS} + 1}$$

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 by nonlinear regression to this user-defined equation:
 $Y = X \cdot \text{NS} / (\text{NS} + 1)$
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.

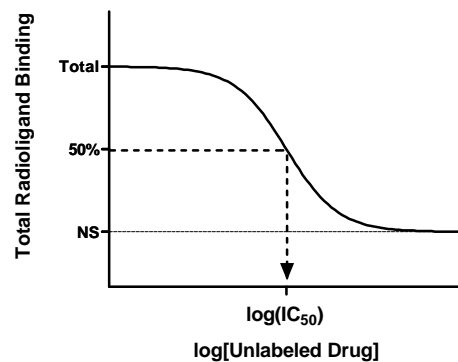
If you think the value of NS determined here is more accurate than the NS determined from analysis of the total binding, refit the total binding data holding NS constant (equal to the value determined from the nonspecific binding analysis).

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + 10^{X - \text{LogEC}_{50}}}$$

Analyzing competitive binding data

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. Ideally, 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 the same 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).

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 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 in the data table, rather than the logarithm of concentration, transform the data before performing nonlinear regression. Follow these steps:

1. Enter the data with X as concentration of ligand 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.
7. Check the option box to make a new graph of the transformed data.

Note: When fitting a curve with nonlinear regression, be sure to fit to the new transformed data table or the new graph.

Decisions to make before fitting the data

Weighting

When analyzing the data, you need to decide whether to minimize the sum of the squares of the absolute distances of the points from the curve or to minimize the sum of the squares of the relative distances. See "Weighting method" on page 200. 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 fit 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 IC_{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 Prism 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.

Competitive binding data with one class of receptors

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. From the nonlinear regression dialog, choose the option K_i from IC_{50} and enter values for the K_d of the radioligand and its concentration. Enter both in nM (or any concentration units; only the ratio matters). Enter concentrations, not the logarithm of concentration. The K_d must be known from previous saturation binding experiments

Checklist for competitive binding results

When evaluating results of competitive binding, ask yourself these questions:

Question	Comment
Is the logIC ₅₀ reasonable?	The IC ₅₀ should be near the middle of the curve, with at least several concentrations of unlabeled competitor on either side of it.
Are the standard errors too large? Are the confidence intervals too wide.	The SE of the logIC ₅₀ should be less than 0.5 log unit (ideally a lot less).
Are the values of TOP and BOTTOM reasonable?	TOP should be near the binding you observed in the absence of competitor. BOTTOM should be near the binding you observed in the presence of a maximal concentration of competitor. If the best-fit value of BOTTOM is negative, consider fixing it to a constant value equal to nonspecific binding determined in a control tube.
Has binding reached equilibrium?	Competitive binding incubations take longer to equilibrate than saturation binding incubations. You should incubate for 4-5 times the half-life for radioligand dissociation.
Does only a small fraction of the radioligand bind?	The equations are based on the assumption that the free concentration of labeled ligand is essentially identical to the concentration you added. Compare the total binding in the absence of competitor in cpm, to the amount of ligand added in cpm. If the ratio is greater than 10% at any concentration, then you've violated this assumption. Try to revise your experimental protocol, perhaps using a large incubation volume.
Does the curve have the expected steepness?	The competitive binding curve has a Hill slope (or slope factor) of -1. If your data form a curve shallower than this, see "Shallow competitive binding curves" on page 268.

K_i from EC₅₀

Prism first fits the curve to find the EC₅₀, the concentration of competitor that competes for half the specific binding. This is the same as the IC₅₀.

The value of the EC₅₀ is determined by three factors:

- The affinity of the receptor for the competing drug. If the affinity is high, the EC₅₀ will be low. The affinity is usually quantified as the equilibrium dissociation constant, K_i. The subscript i is used to indicate that the competitor inhibited radioligand binding. You can interpret it the same as you interpret a K_d. The K_i 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. If the K_i is low, the affinity of the receptor for the inhibitor is high.
- 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 half the radioligand binding sites.
- The affinity of the radioligand for the receptor (K_d). It takes more unlabeled drug to compete for a tightly bound radioligand (low 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(EC₅₀), so Prism finds the best-fit value of the log(EC₅₀) along with its SE and 95% CI. Prism also reports the EC₅₀ and its 95% CI. It does this by taking the antilog of the log(EC₅₀) and of both ends of the 95% CI. Since the confidence interval is symmetrical on the log scale, it is not symmetrical when converted to EC₅₀.

If the concentrations of unlabeled compound are equally spaced on a log scale, the uncertainty of the log(EC₅₀) will be symmetrical, but the uncertainty of the EC₅₀ will not be. That is why the equation is written in terms of log(EC₅₀).

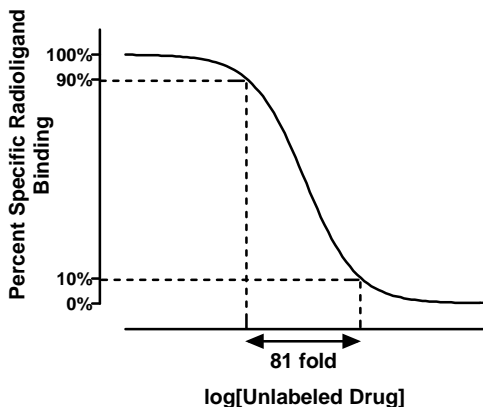
If you average together results from several experiments, it is better to average the log(K_i) values, rather than the K_i values. If you average K_i values, one value that is far from the rest will have too much influence on the mean. See "Why Prism fits the logEC₅₀ rather than EC₅₀" on page 301.

Shallow competitive binding curves

The slope factor or Hill slope

If the labeled and unlabeled ligands compete for a single class of binding site, the competitive binding curve will have a shape determined by the

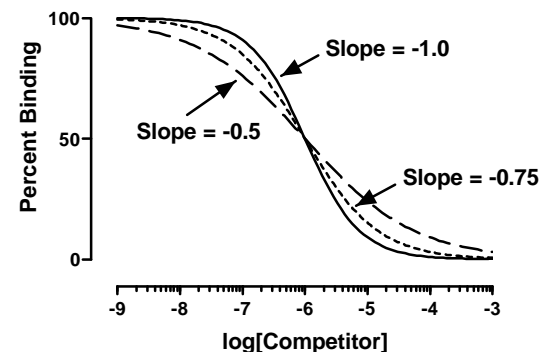
law of mass action. In this case, the curve will descend from 90% specific binding to 10% specific binding over an 81-fold increase in the concentration of the unlabeled drug. More simply, virtually the entire curve will cover two log units (100-fold change in concentration).



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 shallower, the slope factor will be a negative fraction, perhaps -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. There is no advantage to determining the Hill slope this way – it is more difficult and less accurate.



Explanations for shallow binding curves include:

Explanation	Comment
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. You may need to set the variables Top and Bottom to constant values.
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	When agonists interact with a G protein whose availability is limited, curves will be shallow. This is a common explanation for shallow binding. See "Receptors linked to G proteins" on page 271.

Competitive binding with two sites

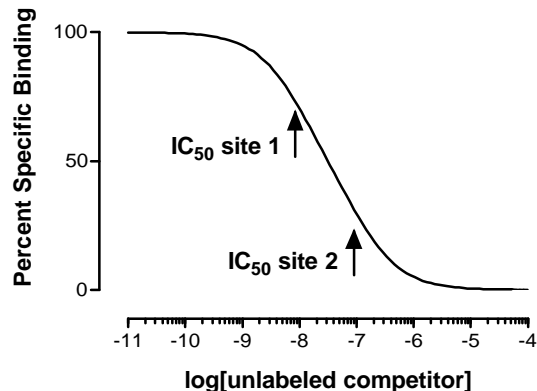
Included in the list of built-in equations of Prism is “Competitive binding (two sites)”. This equation fits data for the fairly common situation where:

- 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 receptor types (" on page 282.)
- 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 extremely different. More often you will see a shallow curve with the two components blurred together. For example, the graph below 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 fits 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. The resulting P value 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 the fits of two models " on page 221.

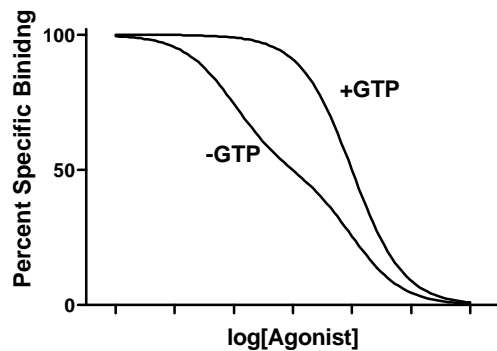
Before looking at Prism's comparison of the two equations, you should look at both fits yourself. Sometimes the two-site fit gives results that are clearly nonsense. For example, disregard a two-site fit when:

- 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 variable FRACTION is negative or greater than 1.0.
- The best-fit values for BOTTOM or TOP are far from the range of Y values observed in your experiment.

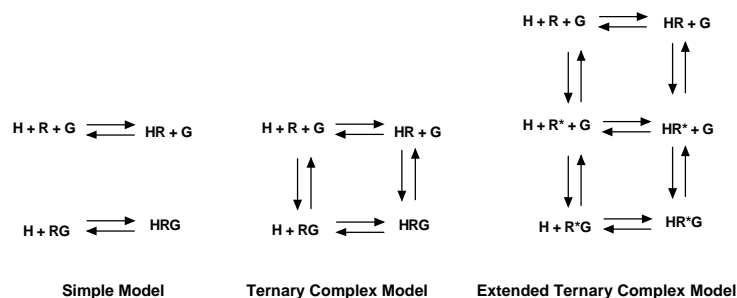
If the results don't make sense, don't believe them. Only pay attention to the comparison of two fits when each of the fits makes sense.

Receptors linked to G proteins

A well studied example of agonist binding is the interaction of agonists with receptors linked to G proteins. This is studied by comparing the competition of agonists with radiolabeled antagonist binding in the presence and absence of GTP (or its analogues). These experiments are done in membrane preparations to wash away endogenous GTP. Without added GTP, the competitive binding curves tend to be shallow. When GTP or an analog is added, the competitive binding curve is of normal steepness. This figure shows an idealized experiment.



The extended ternary complex model, shown in the figure below, can partially account for these findings (and others). In this model, receptors can exist in two states, R and R*. The R* state has a high affinity for agonist and preferentially associates with G proteins. Although some receptor may exist in the R* state in the absence of agonist, the binding of agonist fosters the transition from R to R* and thus promotes interaction of the receptor with G protein. The extended ternary complex model is shown on the right. Even this extended ternary complex model may be too simple, as it does not allow for receptors in the R state to interact with G. A cubic ternary complex model adds these additional equilibria. For details, see Kenakin, *Pharmacologic Analysis of Drug-Receptor Interaction*, 3rd edition, Lippincott-Raven, 1997.



The agonist binding curve is shallow (showing high and low affinity components) in the absence of GTP because some receptors interact with G proteins and others do not. The receptors that interact with G proteins bind agonist with high affinity, while those the receptors that don't interact with G proteins bind agonist with low affinity. Why don't all receptors bind to G proteins? The simplest answer is that there are fewer G proteins than receptors, but biochemical evidence disputes this idea. Other possible explanations include heterogeneous receptors and membrane compartmentation (so some receptors are sequestered from G proteins). For a review of these issues, see RR Neubig, *Faseb J.* 8:939-946, 1994.

If all the receptors could interact with G proteins, you'd expect to see an entirely high affinity binding curve in the absence of GTP. In the presence of GTP (or an analog) the HR*G complex is not stable, the G protein dissociates into its α_{GTP} and $\beta\gamma$ subunits, and the receptor is uncoupled from G. With GTP present, only a tiny fraction of receptors are coupled to G at any given time, so the agonist competition curves are of low affinity and normal steepness, as if only R was present and not RG.

Although the extended ternary complex model is very useful conceptually, it is not very useful when analyzing data. There are simply too many variables! The simpler ternary complex model shown in the middle of the figure has fewer variables, but still too many to reliably fit with nonlinear regression. For routine analyses, most investigators fit data to the much simpler two-state model shown on the left of the figure. This model allows for receptors to exist in two affinity states (R and RG), but does not allow conversion between R and RG. It is easy to fit data to this simpler model using a two-site competition curve model. Since we know the model is too simple, the high and low affinity dissociation constants derived from the model should be treated merely as empirical descriptions of the data, and not as true molecular equilibrium constants.

Homologous competitive binding curves

Introducing homologous competition

The most common 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 chemical, but not radioactively labeled. Since the hot (radiolabeled) and cold (unlabeled) ligands are chemically identical, this is called a homologous competitive binding experiment.

Most analyses of homologous competition data are based on these assumptions:

Assumption	Comments
The receptors have identical affinity for the labeled and unlabeled ligand.	This is a nontrivial assumption. With tritiated ligands, there is no reason to doubt it, since tritium doesn't greatly alter the conformation of a molecule. However iodination can change conformation and alter the binding affinity. Don't assume that iodinated and noniodinated compounds bind with the same affinity. If your radioligand is labeled with radioactive iodine, then you should use a competitor that is the same compound iodinated with nonradioactive iodine.
There is no cooperativity.	This means that binding of ligand to one binding site does not change its affinity at other site(s).
No ligand depletion.	The basic analysis methods assume that only a small fraction of ligand binds. In other words, the method assumes that free concentration of hot (and cold) equals the concentration you added. Since homologous competition curves are best performed with low concentrations of radioligand, it may be difficult to comply with this assumption. If a large fraction of radioligand binds, you can lower the fractional binding by increasing the incubation volume (without increasing the amount of tissue). A later section in this chapter explains how to analyze data when this assumption is not valid.
Nonspecific binding is proportional to the concentration of labeled ligand.	We assume that a certain fraction of hot ligand binds nonspecifically, regardless of the concentration of unlabeled ligand. This assumption has proven to be true in many systems.

Theory of homologous competition binding

Start with the equation for equilibrium binding to a single class of receptors.

$$\text{Specific Binding} = \frac{[\text{Ligand}] \cdot B_{\max}}{[\text{Ligand}] + K_d}$$

Set [Ligand] equal to the sum of the labeled (hot) and unlabeled ligand (cold). Specific binding you measure (specific binding of the labeled ligand) equals specific binding of all ligand times the fraction of the ligand that is labeled. This fraction, hot/(hot + cold), varies from tube to tube. Therefore specific binding of labeled ligand follows this equation:

$$\begin{aligned} \text{Sp. Binding of hot ligand} &= \text{Sp. binding of all ligand} \cdot \text{Fraction of ligand that is hot} \\ &= \frac{([\text{Hot}] + [\text{Cold}]) \cdot B_{\max}}{[\text{Hot}] + [\text{Cold}] + K_d} \cdot \frac{[\text{Hot}]}{[\text{Hot}] + [\text{Cold}]} \\ &= \frac{B_{\max} \cdot [\text{Hot}]}{[\text{Hot}] + [\text{Cold}] + K_d} \end{aligned}$$

Specific binding and B_{\max} are in the same units, usually cpm, sites/cell or fmol/mg. [Hot], [Cold] and K_d are in concentration units. Those units cancel so it doesn't matter if you use molar, nM, or some other unit, so long as you are consistent.

Maximum binding of labeled ligand occurs when the concentration of cold ligand equals zero. This is not the same as B_{\max} , because the concentration of hot ligand will not saturate all the receptors. In the absence of cold ligand (set [cold] = 0), the binding equals

$$\text{Sp. Binding}_{[\text{Cold}] = 0} = \frac{B_{\max} \cdot [\text{Hot}]}{[\text{Hot}] + K_d}$$

The IC_{50} in a homologous binding experiment is the concentration of [Cold] that reduces specific binding of labeled ligand by 50%. So the IC_{50} is the concentration of cold that solves the equation below. The left side of the equation is half the maximum binding with no cold ligand. The right side is binding in the presence of a particular concentration of cold ligand. We want to solve for [Cold].

$$0.5 \cdot \frac{B_{\max} \cdot [\text{Hot}]}{[\text{Hot}] + K_d} = \frac{B_{\max} \cdot [\text{Hot}]}{[\text{Hot}] + K_d + [\text{Cold}]}$$

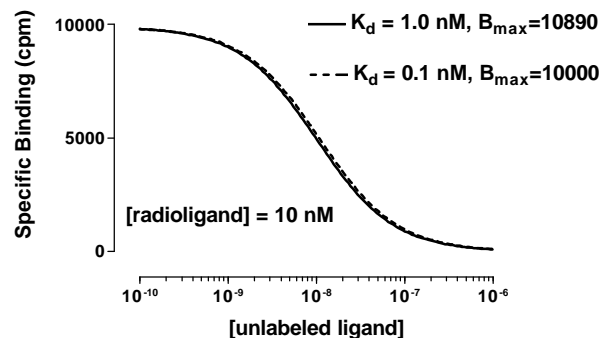
Solve this equation for [Cold], and you'll find that you achieve half-maximal binding when $[\text{Cold}] = [\text{Hot}] + K_d$. In other words,

$$IC_{50} = [\text{Hot}] + K_d$$

Why homologous binding data can be ambiguous

Since the IC_{50} equals $[\text{Hot}] + K_d$, the value of the K_d doesn't affect the IC_{50} very much when you use a high concentration of radioligand. This means

that you'll see the same IC_{50} with a large range of K_d values. For example if you use a Hot ligand concentration of 10 nM, the IC_{50} will equal 10.1 nM if the K_d is 0.1 nM (dashed curve below), and the IC_{50} will equal 11 nM if the K_d is 1 nM (solid curve below). These two IC_{50} values are almost identical, and cannot be distinguished in the presence of experimental error.



If the concentration of hot (radiolabeled) ligand greatly exceeds the K_d , the curve is ambiguous. There are an infinite number of curves, defined by different K_d and B_{max} values, which are almost identical. The data simply don't define the K_d and B_{max} . No curve-fitting program can determine the K_d and B_{max} from this type of experiment – the data are consistent with many K_d and B_{max} values.

Fitting homologous competition data (one site)

We recommend that you follow a two-step procedure for fitting homologous competition data, at least with new systems. Once you have a routine assay, you may want to skip step 1.

Step 1. Determine the IC_{50}

This first step is to check that you used a reasonable concentration of radioligand.

Fit your data to the built-in equation, One-site competition. If your competition curve doesn't have clearly defined top and bottom plateaus, you should set one or both of these to constant values based on control experiments.

Compare the best-fit value of the EC_{50} (same as IC_{50}) to the concentration of hot ligand you used. Homologous competition experiments only lead to useful results when the concentration of hot ligand is less than half the IC_{50} .

If	Then do this
IC_{50} is more than ten times [Hot]	The concentration of [Hot] is lower than it needs to be. If you have quality data, you can continue with step 2. If you have too few cpm bound or if more than 10% of the added radioactivity binds to tissue, rerun the experiment with more radioligand
IC_{50} is between two and ten times [Hot]	You've designed the experiment appropriately. Continue to step 2 to determine the K_d and B_{max} with confidence limits.
IC_{50} is greater than [Hot] but not twice as high	You'll get better results if you repeat the experiment using less hot ligand.
IC_{50} is less than [Hot]	If binding follows the assumptions of the analysis, it is impossible for the IC_{50} to be less than the concentration of hot ligand. Since $K_d = IC_{50} - [Hot]$, these data would lead to a negative value for K_d , which is impossible. One possibility is that your system does not follow the assumptions of the analysis. Perhaps the hot and cold ligands do not bind with identical affinities. Another possibility is that you simply used too high a concentration of hot ligand, so the true IC_{50} is very close to (but greater than) [Hot]. Experimental scatter may lead to a best-fit IC_{50} that is too low, leading to the impossible results. Repeat the experiment using a lot less hot ligand.

Step 2. Determine K_d and B_{max}

Once you've determined that the IC_{50} is quite a bit larger than the concentration of hot ligand, continue with this step to determine the B_{max} and K_d .

Total binding equals specific binding plus nonspecific binding. Nonspecific binding is the same for all tubes since it only depends on the concentration of hot ligand, which is constant. The equation for specific binding is derived in "Theory of homologous competition binding" on page 274. Add a nonspecific binding term to define total binding:

$$\text{Total Binding} = \frac{B_{max} \cdot [Hot]}{[Hot] + [Cold] + K_d} + NS$$

Since this equation is not built-in to Prism, you'll need to enter it as a user-defined equation as follows:

```

ColdNM=10^(x+9) ;Cold concentration in nM
KdNM=10^(logKD+9) ;Kd in nM
Y=(Bmax*HotnM)/(HotnM + ColdNM + KdNM) + NS

```

With any user-defined equation, you need to define initial values. The best way to do this is to enter rules, so Prism can compute initial values for any set of data. After entering the equation, press the button to define rules for initial values. Here are some suggested rules.

Variable	Units	Comments
X	Log(moles/liter)	The logarithm of the concentration of the unlabeled compound.
Y	CPM , 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 about one tenth of these receptors)
HotnM	nM	Concentration of labeled ligand in every tube. Set this to a constant value that you know from experimental design. Prism cannot fit this value. You must set it to a constant value.
LogKd	Log(moles/liter)	Initial value = 1.2*XMID
NS	Same as Y values	Initial value = 1.0*YMIN

This equation assumes that you have entered X values as the logarithm of the concentrations of the unlabeled ligand in molar, so 1nM (10^{-9} molar) is entered as -9. The first line in the equation adds 9 to make it the logarithm of the concentration in nM, and then takes the antilog to get concentration in nM.

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 K_d is fit as the log(K_d). The second line converts the log of the K_d in moles/liter to nM.

Set HotnM equal to the concentration of labeled ligand in nM, and set it to be a constant value. Prism cannot fit this value; you must make it a constant and enter its value.

Homologous competitive binding with ligand depletion

If a large fraction of the radioligand is bound to the receptors, homologous binding will be affected. Although you add the same amount of labeled ligand to each tube, the free concentration will not be the same. High concentrations of unlabeled drug compete for the binding of the labeled ligand, and thus increase the free concentration of the labeled ligand.

Nonspecific binding is also affected by ligand depletion. Since nonspecific binding is proportional to the free concentration of labeled ligand, the amount of nonspecific binding will not be the same in all tubes. The tubes with the highest concentration of unlabeled drug have the highest concentration of free radioligand, so will have the most nonspecific binding.

Because the free concentration varies among tubes, as does the nonspecific binding, there is no simple relationship between IC₅₀ and K_d. The IC₅₀ is nearly meaningless in a homologous binding curve with ligand depletion.

The equations for homologous binding with ligand depletion are quite a bit more complicated than for homologous binding without depletion. The math that follows is adapted from S. Swillens (Molecular Pharmacology, 47: 1197-1203, 1995).

Start with the equation for total binding in homologous competition as a function of the free concentration of radioligand.

$$\text{Specific} = \frac{B_{\text{max}} \cdot [\text{Free Radioligand, nM}]}{K_d + [\text{Free Radioligand, nM}] + [\text{Free Cold ligand, nM}]}$$

$$\text{Nonspecific} = [\text{Free Radioligand, cpm}] \cdot \text{NS}$$

$$Y = \text{Specific} + \text{Nonspecific}$$

This equation defines total binding as specific binding plus nonspecific binding. Nonspecific binding equals a constant fraction of free radioligand, and we define this fraction to be NS. 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).

The problem with this equation is that you don't know the concentrations of free radioligand or free cold ligand. What you know is the concentrations of labeled and unlabeled ligand you added. Since a high fraction of ligand binds to the receptors, you cannot assume that the concentration of free ligand equals the concentration of added ligand.

Defining the free concentration of hot ligand is easy. You added the same number of cpm of hot ligand to each tube, which we'll call HotCPM. The concentration of free radioligand equals the concentration added minus the

total concentration bound, or HotCPM-Y (both HotCPM and Y are expressed in cpm).

Defining the free concentration of cold ligand is harder, so it is done indirectly. The fraction of hot radioligand that is free equals (HotCPM - Y)/HotCPM. This fraction will be different in different tubes. Since the hot and cold ligands are chemically identical, the fraction of cold ligand that is free in each tube is identical to the fraction of hot ligand that is free. Define X to be the logarithm of the total concentration of cold ligand, the variable you vary in a homologous competitive binding experiment. Therefore, the total concentration of cold ligand is 10^X , and the free concentration of cold ligand is $10^X(\text{HotCPM} - Y)/\text{HotCPM}$.

Substitute these definitions of the free concentrations of hot and cold ligand into the equation above, and the equation is still unusable. The problem is that the variable Y appears on both sides of the equal sign. Some simple, but messy, algebra puts Y on the left side of a quadratic equation, 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)

```

Select this equation from the Advanced Radioligand Binding equation library.

Variable	Units	Comments
X	Log(Molar)	
Y	Cpm	
HotCPM	Cpm	Amount of labeled ligand added to each tube. Set to a 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 = $1.0 * X_{MID}$
Bmax	Units of Y axis, usually cpm	Initial value = $10 * Y_{MAX}$ (This assumes that you've used a concentration of radioligand that binds to one tenth of the receptors. You may wish to adjust this.)
NS	Unitless fraction	This is the fraction of free ligand that binds nonspecifically. 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.

Fitting homologous competition data (two sites)

With some systems it is possible to determine K_d and B_{max} values for two independent sites using homologous competition data. With most systems, however, you won't get reliable results.

You can only determine B_{max} and K_d from homologous binding data if you use a concentration of hot ligand that is much lower than the K_d value. If your system has two binding sites, you must choose a concentration much lower than the K_d of the high affinity site. Using such a low concentration of radioligand, you'll bind only a small fraction of low-affinity sites. You only be able to detect the presence of the second, low-affinity, site if they are far more abundant than the high-affinity sites.

For example, imagine that the low affinity site ($K_d = 10$ nM) is ten times as abundant as the high affinity site ($K_d = 0.1$ nM). You need to use a concentration of hot ligand less than 0.1 nM, say 0.05 nM. At this concentration you bind to 33.33% of the high affinity sites, but only to 0.0049% of the low affinity sites. Even though the low affinity sites are ten times as abundant, you won't find them in your assay (low affinity binding will be only 0.15% of the binding).

To attempt to determine the two K_d and B_{max} values from a homologous competition curve, fit the data to the equation below. Assuming no cooperativity and no ligand depletion, the binding to each site is independent and depends on the B_{max} and K_d values of each site. The

binding that you measure, Y, is the sum of the binding to the two receptor sites plus nonspecific binding.

$$\begin{aligned} \text{Site1} &= (\text{Bmax1} * \text{HotnM}) / (\text{HotnM} + 10^{-(X+9)} + 10^{-(\log Kd1+9)}) \\ \text{Site2} &= (\text{Bmax2} * \text{HotnM}) / (\text{HotnM} + 10^{-(X+9)} + 10^{-(\log Kd2+9)}) \\ Y &= \text{site1} + \text{site2} + \text{NS} \end{aligned}$$

Define rules for initial values using the suggestions in the table below. This is a difficult equation to fit, and you will almost certainly have to try many sets of initial values 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 high affinity 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 your experimental design.

Consider this approach for analyzing homologous competitive binding data to determine the characteristics of two sites. First use a very low concentration of radioligand and fit to a single site. This will determine the Bmax and Kd of the high affinity site. Then repeat the experiment with a higher concentration of radioligand. Fit these data to the two-site equation,

but set the Kd and Bmax of the high affinity site to constant values determined from the first experiment.

Advantages and disadvantages of homologous binding experiments

Determining receptor number with homologous binding has one clear advantage: You need far less radioligand than you would need if you performed a saturation binding experiment. This reason can be compelling for ligands that are particularly expensive or difficult to synthesize.

The disadvantage of determining receptor number and affinity from a homologous competitive binding experiment is that it can be hard to pick an appropriate concentration of radioligand. If you use too little radioligand, you'll observe little binding and will obtain poor quality data. If you use too much radioligand, the curve will be ambiguous and you won't be able to determine Bmax and Kd.

Using homologous binding to determine the Kd and Bmax of two binding sites with homologous binding is difficult. You are probably better off using a saturation binding experiment.

Competitive binding to two receptor types (different Kd for hot ligand)

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 situations where the labeled ligand binds differently to the two sites.

This is the standard equation for competitive binding to one site:

$$Y = \frac{[\text{Hot Ligand}] \cdot B_{\text{max}}}{[\text{Hot Ligand}] + K_D + \frac{K_D}{K_i} \cdot [\text{Cold Ligand}]} + \text{Nonspecific}$$

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 Bmax, Kd, and Ki:

```

;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

```

Select this equation from the Advanced Radioligand Binding equation library.

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 constant 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 this.)
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.

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.

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 be less than the added concentration of radioligand. The discrepancy between free and added radioligand concentration depends on the concentration of the unlabeled drug. The standard equation for competitive binding, shown below, needs two corrections to account for ligand depletion.

$$Y = \frac{[\text{Free Ligand}] \cdot B_{\max}}{[\text{Free Ligand}] + K_d \left(1 + \frac{[\text{Cold Ligand}]}{K_i} \right)} + \text{Nonspecific}$$

- 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. 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.

$$\text{Nonspecific binding} = \text{NS} \cdot [\text{Free ligand}]$$

Y, [Free ligand], and [Added ligand] are expressed in units of cpm. To be consistent, therefore the Kd also needs to be expressed in cpm units. [Cold

ligand] and K_i are expressed in the same units (molar), so the ratio is unitless.

Combine these 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)
```

Select this equation from the Advanced Radioligand Binding equation library.

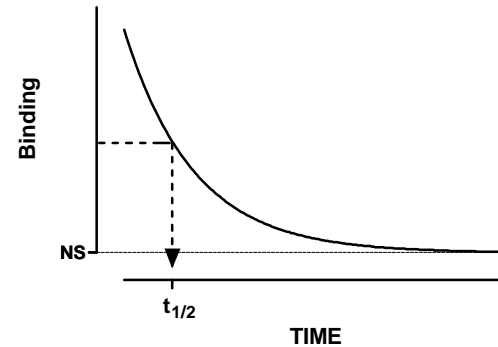
Variable	Units	Comments
X	log(Molar)	
Y	CPM	
Hot	CPM	Amount of labeled ligand added to each tube. Set to a constant value.
SpAct	cpm/fmol	Specific radioactivity of labeled ligand. Set to constant value.
Vol	ml	Incubation volume. Set to a constant value.
KdnM	nM	Kd of labeled ligand. Set to a constant 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 this.)
NS	Unitless fraction	Initial value =0.01

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 log K_i . It also fits two other variables which are of 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.

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.

Analyzing kinetic binding data



Dissociation ("off rate") experiments

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.
- Spin the suspension and resuspend in fresh buffer.
- Add a very high concentration of an unlabeled ligand. If this 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 to 100 fold dilution. This will reduce the concentration of radioligand by that factor. At such a low concentration, new binding of radioligand will be negligible. This method is only practical when you use a fairly low concentration of radioligand so its concentration after dilution is far below its K_d for binding.

You then measure binding at various times after that to determine how rapidly the ligand falls off the receptors.

Each ligand-receptor complex dissociates at a random time, so the amount of specific binding follows an exponential dissociation (see "Example model 2. Exponential " on page 158).

$$Y = \text{Span} \cdot e^{-K \cdot X} + \text{Plateau}$$

Variable	Meaning	Comment
X	Time	Usually expressed in units of seconds or minutes.
Y	Total binding	Usually expressed in units of cpm, fmol/mg, or sites/cell.
Span	Difference between binding at time zero and plateau.	Specific binding (same units as Y)
Plateau	Binding that doesn't dissociate.	Nonspecific binding (same units as Y).
K	Dissociation rate constant often called k_{off} .	Expressed In units of inverse time (inverse of units of X-axis)
$t_{1/2}$	Half-life	$0.69302/k_{off}$

Analyzing dissociation data with Prism

To analyze dissociation binding data:

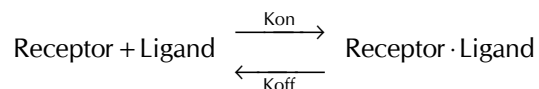
1. Enter the data with X equal to time after you initiated dissociation and Y equal to binding (usually total binding).
2. Perform nonlinear regression using the one phase exponential decay equation.

- 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 as a variable to fit.
- 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).

Association binding experiments

Association binding experiments are used to determine the association rate constant. You add radioligand and measure specific binding at various times thereafter.

Binding follows the law of mass action:



At any given time, the rate at which receptor-ligand complexes form is proportional to the radioligand concentration and the number of receptors still unoccupied. The rate of dissociation is proportional to the concentration of receptor-ligand complexes.

Binding increases over time until it plateaus when specific binding equals a value we call Y_{max} . This is not the same as B_{max} . Y_{max} is the amount of specific binding at equilibrium for a certain concentration of ligand used in an association experiment. B_{max} is the maximum amount of binding extrapolated to a very high concentration of ligand. The free concentration of receptors at any time equals Y_{max} minus the amount of specific binding at that time.

These principles let us define the model mathematically.

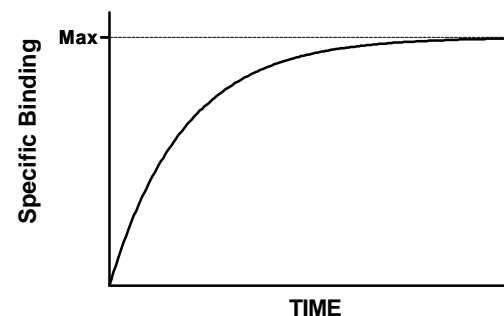
$$\text{Rate of association} = [\text{Receptor}] \cdot [\text{Ligand}] \cdot k_{on} = (Y_{max} - Y) \cdot [\text{Ligand}] \cdot k_{on}$$

$$\text{Rate of dissociation} = [\text{Receptor} \cdot \text{Ligand}] \cdot k_{off} = Y \cdot k_{off}$$

$$\begin{aligned} \text{Net rate of association} &= \frac{dY}{dX} = \text{Rate of association} - \text{rate of dissociation} \\ &= (Y_{max} - Y) \cdot [\text{Ligand}] \cdot k_{on} - Y \cdot k_{off} \\ &= Y_{max} \cdot [\text{Ligand}] \cdot k_{on} - Y([\text{Ligand}] \cdot k_{on} + k_{off}) \end{aligned}$$

Integrate that differential equation to obtain the equation defining the kinetics of association:

$$Y = Y_{max} \cdot (1 - e^{-K \cdot X})$$



The rate at which binding increases is determined by three factors (as well as experimental conditions such as pH and temperature):

- 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 both binds to and dissociates from receptors. The system reaches equilibrium when the two rates are equal. 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 (but with less binding).

Analyzing "on rate" experiments with Prism

To analyze association (on-rate) data:

- 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.)
- Fit the specific binding data to the one-phase exponential association equation.

$$Y = Y_{max} \cdot (1 - e^{-k \cdot X})$$

- 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.
- To calculate the association rate constant (k_{on} or k_1) usually expressed in units of $\text{Molar}^{-1} \text{min}^{-1}$, use this equation:

$$k_{on} = \frac{k_{ob} - k_{off}}{[\text{radioligand}]}$$

Variable	Units	Comment
k_{on}	$\text{Molar}^{-1} \text{min}^{-1}$	What you want to know.
k_{ob}	min^{-1}	The value of K determined by fitting an exponential association equation to your data.
k_{off}	min^{-1}	The dissociation rate constant. See the previous section.
[radioligand]	Molar	Set by the experimenter. Assumed to be constant during the experiment (only a small fraction binds).

Analysis checklist for kinetic binding experiments

Question	Comment
Did you go out to a long enough time point?	Dissociation and association data should plateau, so the data obtained at the last few time points should be indistinguishable.
Is the value of k_{on} reasonable?	The association rate constant, k_{on} , depends largely on diffusion so is similar for many ligands. Expect a result of about $10^8 \text{M}^{-1} \text{min}^{-1}$
Is the value of k_{off} reasonable?	If the k_{off} is greater than 1min^{-1} , the ligand has a low affinity for the receptor, dissociation will occur while you are separating bound and free ligands, and you'll have a hard time obtaining quality data. If k_{off} is less than 0.001min^{-1} , you'll have a difficult time obtaining equilibrium as the half-time of dissociation will be greater than 10 hours! Even if you wait that long, other reactions may occur that ruin the experiment.
Are the standard errors too large? Are the confidence intervals too wide.	Examine the SE and the confidence intervals to see how much confidence you have in the rate constants.
Does only a tiny fraction of radioligand bind to the receptors.	The standard analyses of association experiments assume that the concentration of free radioligand is constant during the experiment. This will be approximately true only if a tiny fraction of the added radioligand binds to the receptors. Compare the maximum total binding in cpm to the amount of added radioligand in cpm. If that ratio exceeds 10% or so, you should revise your experimental protocol.

Using kinetic data to test the law of mass action

Standard binding experiments are usually fit to equations derived from the law of mass action. Kinetic experiments provide a more sensitive test than equilibrium experiments to determine whether the law of mass action actually applies for your system. To test the law of mass action, ask these questions:

Does the K_d calculated from kinetic data match the K_d calculated from saturation binding?

According to the law of mass action, the ratio of k_{off} to k_{on} is the K_d of receptor binding:

$$K_d = \frac{k_{off}}{k_{on}}$$

The units are consistent: k_{off} is in units of min^{-1} ; k_{on} is in units of $\text{M}^{-1}\text{min}^{-1}$, so K_d is in units of M .

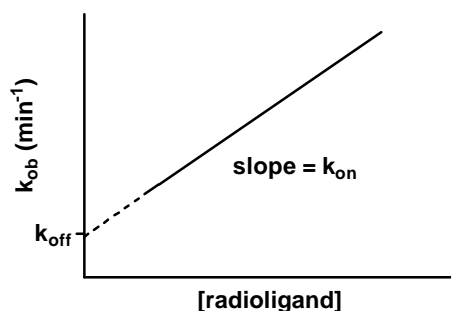
If binding follows the law of mass action, the K_d calculated this way should be the same as the K_d calculated from a saturation binding curve.

Does k_{ob} increase linearly with the concentration of radioligand?

The observed association rate constant, k_{ob} , is defined by this equation:

$$k_{ob} = k_{off} + k_{on} \cdot [\text{radioligand}]$$

Therefore, if you perform association rate experiments at various concentrations of radioligand, the results should look like the figure below. As you increase the concentration of radioligand, the observed rate constant increases linearly. If the binding is more complex than a simple mass action model (such as a binding step followed by a conformational change) the plot of k_{ob} vs. $[\text{radioligand}]$ may plateau at higher radioligand concentrations. Also, you should extrapolate the plot back to zero radioligand to determine the intercept which equals k_{off} . If the law of mass action applies to your system, the k_{off} determined this way should correspond to the k_{off} determined from a dissociation experiment. Finally, this kind of experiment provides a more rigorous determination of k_{on} than that obtained with a single concentration of radioligand.



Is specific binding 100% reversible, and is the dissociated ligand chemically intact?

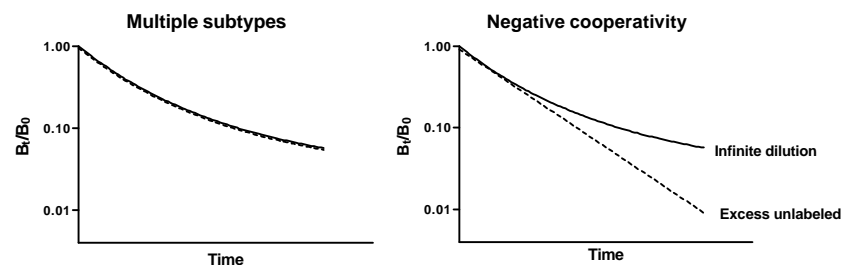
Nonspecific binding at "time zero" should equal total binding at the end (plateau) of the dissociation. In other words, all of the specific binding should dissociate if you wait long enough. Use chromatography to analyze the radioligand that dissociates to prove that it has not been altered.

Is the dissociation rate consistent with different experimental conditions?

Determine the dissociation constant after binding various concentrations of radioligand for various lengths of time. If your ligand binds to a single site and obeys the law of mass action, you'll obtain the same dissociation rate constant in all experiments.

Is there cooperativity?

If the law of mass action applies, binding of a ligand to one binding site does not alter the affinity of another binding site. This also means that dissociation of a ligand from one site should not change the dissociation of ligand from other sites. To test this assumption, compare the dissociation rate after initiating dissociation by infinite dilution with the dissociation rate when initiated by addition of a large concentration of unlabeled drug. If the radioligand is bound to multiple noninteracting binding sites, the dissociation will be identical in both experimental protocols as shown in the left figure. Note that the Y axis is shown using a log scale. If there were a single binding site, you'd expect the dissociation data to appear linear on this graph. With two binding sites, the graph is curved even on a log axis.



The right figure shows ideal dissociation data when radioligand is bound to interacting binding sites with negative cooperativity. The data are different depending on how dissociation was initiated. If dissociation is initiated by infinite dilution, the dissociation rate will change over time. The dissociation of some radioligand will leave the remaining ligand bound more tightly. When dissociation is initiated by addition of cold drug, all the

receptors are always occupied by ligand (some hot, some cold) and dissociation occurs at its maximal unchanging rate.

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 dissociation constant of the receptors for the unlabeled compound, the K_i . The law of mass action tells us that the K_i is the ratio of the off-rate to the on-rate of the unlabeled compound. You can determine these values in a kinetics experiment as follows. 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.

$$\begin{aligned}
 KA &= K1 * L * 1E-9 + k2 \\
 KB &= K3 * I * 1e-9 + K4 \\
 S &= \text{SQRT}((KA - KB)^2 + 4 * K1 * K3 * L * I * 1e-18) \\
 KF &= 0.5 * (Ka + KB + S) \\
 KS &= 0.5 * (KA + KB - S) \\
 DIFFF &= KF - KS \\
 Q &= Bmax * K1 * L * 1e-9 / DIFFF \\
 Y &= Q * (k4 * DIFFF / (KF * KS) + ((K4 - Kf) / KF) * \exp(-KF * X) - ((K4 - KS) / KS) * \exp(-KS * X))
 \end{aligned}$$

Select this equation from the Advanced Radioligand Binding equation library.

Variable	Units	Comments
X	Minutes	Time.
Y	cpm	Specific binding.
k1	M ⁻¹ min ⁻¹	Association rate constant of radioligand. Set to a constant value known from other experiments.
k2	min ⁻¹	Dissociation rate constant of radioligand. Set to a constant value known from other experiments.
k3	M ⁻¹ min ⁻¹	Association rate constant of unlabeled ligand. Variable to be fit. Try 1e8 as an initial value.
k4	min ⁻¹	Dissociation rate constant 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 number of receptors. 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 only 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.

Analyzing dose-response curves

Introducing dose-response curves

What is a dose-response curve?

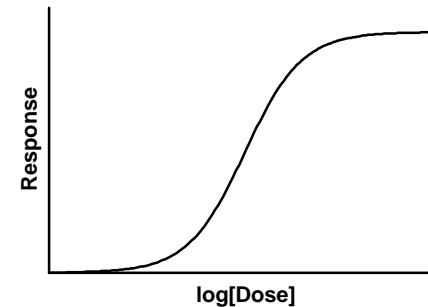
Dose-response curves can be used to plot the results of many kinds of experiments. The X-axis plots concentration of a drug or hormone. The Y-axis plots response, which could be almost anything. For example, the response might be enzyme activity, accumulation of an intracellular second messenger, membrane potential, secretion of a hormone, heart rate or contraction of a muscle.

The term "dose" is often used loosely. The term "dose" strictly only applies to experiments performed with animals or people, where you administer various doses of drug. You don't know the actual concentration of drug – you know the dose you administered. However, the term "dose-response curve" is also used more loosely to describe *in vitro* experiments where you apply known concentrations of drugs. The term "concentration-response curve" is a more precise label for the results of these experiments. The term "dose-response curve" is occasionally used even more loosely to refer to experiments where you vary levels of some other variable, such as temperature or voltage.

An *agonist* is a drug that causes a response. If you administer various concentrations of an agonist, the dose-response curve will go uphill as you go from left (low concentration) to right (high concentration). A *full agonist* is a drug that appears able to produce the full tissue response. A *partial agonist* is a drug that provokes a response, but the maximum response is less than the maximum response to a full agonist. An *antagonist* is a drug that does not provoke a response itself, but blocks agonist-mediated responses. If you vary the concentration of antagonist (in the presence of a fixed concentration of agonist), the dose-response curve will run downhill.

The shape of dose-response curves

Many steps can occur between the binding of the agonist to a receptor and the production of the response. So depending on which drug you use and which response you measure, dose-response curves can have almost any shape. However, in very many systems dose-response curves follow a standard shape, shown below.



Dose-response experiments typically use 10-20 doses of agonist, approximately equally spaced on a logarithmic scale. For example doses might be 1, 3, 10, 30, 100, 300, 1000, 3000, and 10000 nM. When converted to logarithms, these values are equally spaced: 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0.

Note: The logarithm of 3 is actually 0.4771, not 0.50. The antilog of 0.5 is 3.1623. So to make the doses truly equally spaced on a log scale, the concentrations ought to be 1.0, 3.1623, 10.0, 31.623 etc.

Since the linkage between agonist binding and response can be very complex, any shape is possible. It seems surprising, therefore, that so many dose-response curves have shapes identical to receptor binding curves. The simplest explanation is that the link between receptor binding and response is direct, so response is proportional to receptor binding. However, in most systems one or more second-messenger systems link receptor binding to response. For example, agonist binding activates adenylyl cyclase, which creates the second-messenger cAMP. The second messenger can then bind to an effector (such as a protein kinase) and initiate a response.

What do you expect a dose-response curve to look like if a second messenger mediates the response? If you assume that the production of second messenger is proportional to receptor occupancy, the graph of agonist concentration vs. second messenger concentration will have the same shape as receptor occupancy (a hyperbola if plotted on a linear scale, a sigmoid curve with a slope factor of 1.0 if plotted as a semilog graph). If the second messenger works by binding to an effector, and that binding

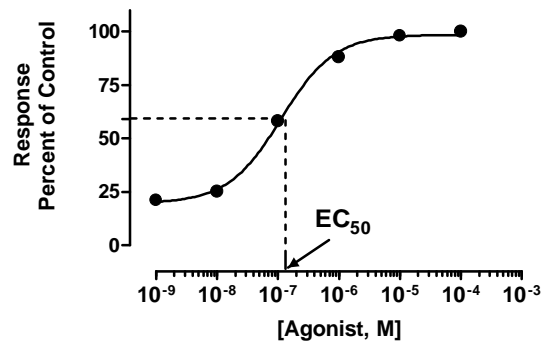
step follows the law of mass action, then the graph of second messenger concentration vs. response will also have that same standard shape. It isn't obvious, but Black and Leff (see "The operational model of agonist action" on page 305) have shown that the graph of agonist concentration vs. response will also have that standard shape (provided that both binding steps follow the law of mass action). In fact, it doesn't matter how many steps intervene between agonist binding and response. So long as each messenger binds to a single binding site according to the law of mass action, the dose-response curve will follow the same hyperbolic/sigmoid shape as a receptor binding curve.

The EC₅₀

A standard dose-response curve is defined by four parameters: the baseline response (Bottom), the maximum response (Top), the slope, and the drug concentration that provokes a response halfway between baseline and maximum (EC₅₀).

It is easy to misunderstand the definition of EC₅₀. It is defined quite simply as the concentration of agonist that provokes a response half way between the baseline (Bottom) and maximum response (Top). It is impossible to define the EC₅₀ until you first define the baseline and maximum response.

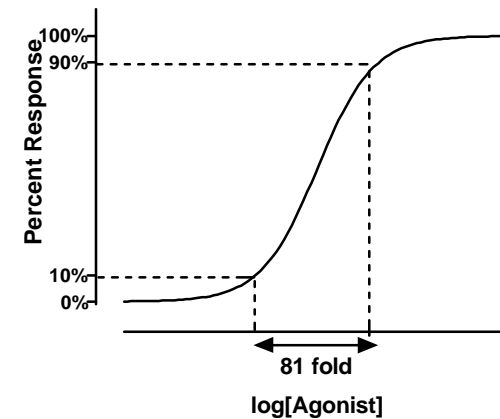
Depending on how you have normalized your data, this may not be the same as the concentration that provokes a response of Y = 50. For example, in the example below, the data are normalized to percent of maximum response, without subtracting a baseline. The baseline is about 20%, and the maximum is 100%, so the EC₅₀ is the concentration of agonist that evokes a response of about 60% (half way between 20% and 100%).



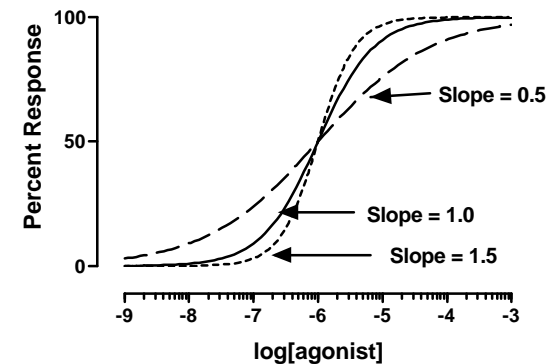
Don't over interpret the EC₅₀. It is simply the concentration of agonist required to provoke a response halfway between the baseline and maximum responses. It is usually not the same as the K_d for the binding of agonist to its receptor.

The steepness of a dose-response curve

Many dose-response curves follow exactly the shape of a receptor binding curve. As shown below, 81 times more agonist is needed to achieve 90% response than a 10% response.



Some dose-response curves however, are steeper or shallower than the standard curve. The steepness is quantified by the Hill slope, also called a slope factor. A dose-response curve with a standard slope has a Hill slope of 1.0. A steeper curve has a higher slope factor, and a shallower curve has a lower slope factor. If you use a single concentration of agonist and varying concentrations of antagonist, the curve goes downhill and the slope factor is negative.



Fitting sigmoid dose-response curves with Prism

To fit a sigmoidal curve with Prism, the X values must be logarithm of concentration. If you entered concentrations, Prism can transform those

values. Click Analyze, choose Transformations from the list of Data Manipulations, and choose to transform X to log(X). Also check the option box to create a new graph.

Since the log(0) is undefined, you cannot enter a concentration of zero as a logarithm. If you enter a concentration of zero and then transform to logarithms, the result will be blank. Instead of entering a dose of zero, enter a low concentration – one or two log units below your lowest concentration.

Beware of entering the zero value as a value much lower than two log units below your lowest concentration. When fitting a sigmoid curve, Prism sets the initial estimated value of the IC₅₀ to be halfway between the lowest and highest X values. If you enter a really low value for the zero concentration, the initial value will be much lower than the true IC₅₀, and Prism may not be able to complete the curve fit (unless you override the initial value).

Before fitting a dose-response curve, make these decisions:

Decision	Discussion
Choose standard slope or variable slope?	Prism offers two equations. One uses a standard slope factor of 1.0; the other fits the slope factor. If you have plenty of data points, choose the equation with variable slope. If data are scanty, use the standard slope.
Set Top to a constant value?	Ideally, the top part of the curve is defined by at least several data points. In this case, Prism will be able to fit the top plateau of the curve. If the top plateau is not well defined by data, then you'll need to make the top plateau be a constant based on controls.
Set Bottom to a constant value?	Ideally, the bottom part of the curve is defined by at least several data points. In this case, Prism will be able to fit the bottom plateau of the curve. If the bottom plateau is not well defined by data, then you'll need to make the bottom plateau be a constant based on controls. If you have subtracted a background value, then the bottom plateau of the curve must be 0. Prism doesn't know this unless you tell it. Make Bottom a constant equal to zero in this case.

Absolute or relative weighting? See "Weighting method" on page 200.

Fit each replicate or averages? See "Replicates" on page 202.

To fit the curve: From the data table or a graph, click Analyze. Choose nonlinear regression from the list of regressions. Choose the list of classic equations and choose either "Sigmoid dose-response" or "sigmoid dose-response (variable slope)". Then click the Method and/or Constants option buttons to configure the fit according to your choices above.

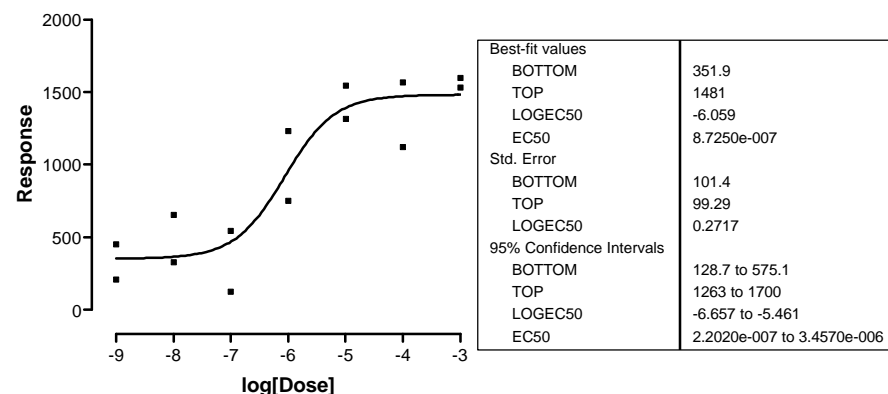
Prism fits the data to the equation below. If you choose the equation with a standard slope, HillSlope is fixed to 1.0.

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{-(\text{LogEC50} - X) * \text{HillSlope}})$$

Why Prism fits the logEC₅₀ rather than EC₅₀

You can write an equation for a dose response curve either in terms of EC₅₀ or log(EC₅₀). Curve fitting finds the curve that minimizes the sum-of-squares of the vertical distance from the points. Rewriting the equation to change between EC₅₀ and log(EC₅₀) isn't going to make a different curve fit better. All it does is change the way that the best-fit EC₅₀ is reported.

However, rewriting the equation to change between EC₅₀ and log(EC₅₀) has a major effect on standard error and confidence interval of the best-fit values. Consider these sample results:



These data were fit to a dose-response curve with a standard slope. The best-fit value for logEC₅₀ is -6.059. Converting to the EC₅₀ is no problem – simply take the antilog. The EC₅₀ is 10^{-6.059} M, about 0.87 μM.

The standard error of the logEC₅₀ is 0.2717. It is used as an intermediate result to calculate a confidence interval, which ranges from -6.657 to -5.461. This means that the 95%CI of the EC₅₀ extends from 10^{-6.657} to 10^{-5.461} — from 0.2207 to 3.458 μM. Expressed as concentrations (rather than log of concentration) the interval is not centered on the best-fit value (0.87 μM). Switching from linear to log scale turned a symmetrical confidence interval into a very asymmetrical interval, which you can report.

If you fit the same data to an equation describing a dose-response curve in terms of the EC₅₀ rather than the logEC₅₀, the EC₅₀ remains 0.87 μM. But now the program computes the SE of the EC₅₀ (0.5459 μM), and uses this to compute the 95% confidence interval of the EC₅₀, which ranges from -0.3290 to +2.074 μM. Note that the lower limit of the confidence interval is negative! Even setting aside the negative portion of the confidence interval, it includes all values from zero on up, which isn't terribly useful. The uncertainty of the EC₅₀ really isn't symmetrical, so the confidence intervals are not useful if you fit to a model written in terms of the EC₅₀.

When some people see the SE of the logEC₅₀, they are tempted to convert this to the standard error of the EC₅₀ by taking the antilog. In the example, the SE of the logEC₅₀ is 0.2717. The antilog of 0.2717 equals 10^{0.2717} or 1.896. What does this mean? It certainly is NOT the SE of the EC₅₀. The SE does *not* represent a point on the axis; rather it represents a distance along the axis. A distance along a log axis does not represent a consistent distance along a linear (standard) axis. For example, increasing the logEC₅₀ 1 unit from -9 to -8 increases the EC₅₀ 9nM; increasing the logEC₅₀ 1 unit from -3 to -2 increases the EC₅₀ by 9 mM (which equals 9,000,000 nM). So you cannot interpret the number 1.896 as a concentration. You can interpret it as a multiplier – a factor you multiply by or divide into the EC₅₀. To calculate the 95% CI, first multiply 1.896 by a constant from the t distribution for 95% confidence and the appropriate number of degrees of freedom (2.2201 for this example). The result is 3.963. Then compute the 95% CI of the EC₅₀. It extends from the best-fit EC₅₀ divided by 3.963 to the best-fit EC₅₀ times 3.963, from 0.22 μM to 3.45 μM.

Other measures of potency: pEC₅₀, EC₈₀, EC₉₀, etc.

The pEC₅₀

The pEC₅₀ is defined as the negative logarithm of the EC₅₀. If the EC₅₀ equals 1 micromolar (10⁻⁶ molar), the log(EC₅₀) is -6 and the pEC₅₀ is 6. There is no particular advantage to expressing potency this way, but it is the custom in some fields.

If you fit a dose response curve using one of the classic equations, Prism will report the logEC₅₀. Multiply by -1 to obtain the pEC₅₀.

If you want to fit the pEC₅₀ directly, perhaps to embed into a table on a Prism graph, use the following equation.

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{-(X - \text{pEC}_{50}) * \text{HillSlope}})$$

Calculating any EC value from the EC₅₀ and Hill slope

The potency of a drug is commonly quantified as the EC₅₀, the concentration that leads to 50% maximal response (or the logarithm of the EC₅₀). But in some systems you might be more interested in the EC₈₀ or the EC₉₀ or some other value. You can compute the EC₈₀ or EC₉₀ (or any other EC value) from the EC₅₀ and Hill slope. Or you can fit data to determine any EC value directly. If you express response as a percentage, a standard dose response curve is described by this equation:

$$F = 100 \frac{L^H}{L^H + EC_{50}^H}$$

L is the ligand concentration, EC₅₀ is the concentration that gives half-maximal effect, and H is the Hill constant or slope factor that defines the steepness of the curve. L and EC₅₀ are expressed in the same units of concentration, so the units cancel out. F is the fractional response (or fractional occupancy for binding data), expressed as a percentage.

Set F to any percentage you want (80 if you want to obtain the EC₈₀) and then solve for L. We call that value of L the EC_F, as that is the quantity of drug needed to elicit an F percentage response (or with binding data, F is the concentration of ligand needed to occupy fraction F of the receptors). A bit of algebra yields this equation:

$$EC_F = \left(\frac{F}{100 - F} \right)^{1/H} \cdot EC_{50}$$

If you know the EC₅₀ and Hill slope (H), you can easily compute the EC₈₀ or EC₁₀ or any other value you want. For example, if the Hill slope equals 1, the EC₉₀ equals the EC₅₀ times nine. If H equals 0.5, the curve is shallower and the EC₉₀ equals the EC₅₀ times 81.

Determining any EC value directly

You can also fit data directly to an equation written in terms of the EC_F. The advantage of this approach is that Prism will report the 95% confidence

value for EC₅₀. Use the equation below, where X is the log of concentration and Y is response, which ranges from Bottom to Top.

$$F=80$$

$$\log EC_{50} = \log ECF - (1/\text{HillSlope}) * \log(F/(100-F))$$

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{((\log EC_{50} - X) * \text{HillSlope})})$$

To fit data to this equation, you'll need to enter rules for computing initial values. Set Top equal to 1*Y_{max}, Bottom equal to 1*Y_{min}. For the Hill slope, simply pick a value, probably +1.0 or -1.0. For the log EC value, enter 1*X_{MID} as a crude initial value, or enter a value based on the range of concentrations you use.

Here is a simplified equation, for fitting the EC₉₀. Here the response is expressed as a percent ranging from zero to one hundred, so we dispense with the variables Top and Bottom.

$$\log EC_{50} = \log EC_{90} - (1/\text{HillSlope}) * \log(9)$$

$$Y = 100 / (1 + 10^{((\log EC_{50} - X) * \text{HillSlope})})$$

Checklist. Interpreting a dose-response curve.

After fitting a dose-response model to your data, ask yourself these questions:

Question	Comment
Is the logEC ₅₀ reasonable?	The EC ₅₀ should be near the middle of the curve, with at least several data points on either side of it.
Are the standard errors too large? Are the confidence intervals too wide.	The SE of the logEC ₅₀ should be less than 0.5 log unit (ideally a lot less).
Is the value of BOTTOM reasonable?	BOTTOM should be near the response you observed with zero drug. If the best-fit value of BOTTOM is negative, consider fixing it to a constant value equal to baseline response. If you know where the bottom of the curve should be, then set BOTTOM to that constant value.

Is the value of TOP reasonable?

TOP should be near the response you observed with maximal concentration drug. If the best-fit value of TOP is not reasonable, consider fixing it to a constant value. If you know where the top of the curve should be, then set TOP that constant value.

If you used a variable slope model, are there enough data to define the slope?

If you asked Prism to find a best-fit value for slope, make sure there at least a few data points between 10 and 90% . If not, your data don't accurately define the slope factor. Consider fixing the slope to its standard value of 1.0

If you used a model with a standard slope, does the data appear to be steeper or shallower?

If the data appear to form a curve much steeper or shallower than the standard dose-response curve, consider fitting to a model with a variable slope.

Does the curve appear to be biphasic?

The standard dose-response models assume that the curve is monotonic. If the curve goes up, and then down, you'll need a more complicated model (beyond the scope of this manual).

The operational model of agonist action

Limitations of dose-response curves

Fitting a sigmoidal (logistic) equation to a dose-response curve to determine EC₅₀ (and perhaps slope factor) doesn't tell you everything you want to know about an agonist. The EC₅₀ reflects both the ability of the drug to bind to its receptor (the agonist's *affinity*) and the ability of the drug to cause a response once it is bound (the agonist's *efficacy*). Thus, the EC₅₀ you observe could have different causes. The agonist could bind with very high affinity, but have low efficacy once bound. Or it could bind weakly with low affinity, but have very high efficacy. Two very different drugs could have the *same* EC₅₀s and maximal responses (in the same tissue). One drug could have high affinity and low efficacy, while the other has low affinity and high efficacy. Since efficacy reflects properties of both agonist and tissue, a single drug acting on one kind of receptor could have different EC₅₀ values in different tissues.

Derivation of the operational model

Black and Leff (Proc. R. Soc. Lond. B, 220:141-162, 1983) developed the *operational model of agonism* to help understand the action of agonists and partial agonists, and to develop experimental methods to determine the affinity of agonist binding and a systematic way to measure relative agonist efficacy based on an examination of the dose-response curves.

Start with a simple assumption: Agonists bind to receptors according to the law of mass action. At equilibrium, the relationship between agonist concentration ([A]) and agonist-occupied receptor ([AR]) is described by the following hyperbolic equation:

$$[AR] = \frac{[R_T] \cdot [A]}{[A] + K_A}$$

[R_T] represents total receptor concentration and K_A represents the agonist-receptor equilibrium dissociation constant (see "Example model 3. Equilibrium binding" on page 160).

What is the relationship between agonist occupied receptor (AR) and receptor action? We know biochemical details in some cases, but not in others. This lack of knowledge about all the steps between binding and final response prevents the formulation of explicit, mechanistic equations that completely describe a dose-response curve. However, Black and Leff derived a "practical" or "operational" equation that encompasses the behavior of all of these unknown biochemical cascades. They began with the observation that dose-response curves often have a sigmoidal shape with a Hill Slope of 1.0, (the curves are hyperbolas when response is plotted against agonist concentration, sigmoidal when response is plotted against the log of concentration). They then proved mathematically that if agonist binding is hyperbolic and the dose-response curve has a Hill slope of 1.0, the equation linking the concentration of agonist occupied receptors to response must also be hyperbolic. This second equation, shown below, has been termed the "transducer function", because it is a mathematical representation of the transduction of receptor occupation into a response:

$$\text{Effect} = \frac{\text{Effect}_{\max} \cdot [AR]}{[AR] + K_E}$$

The parameter, Effect_{max}, is the maximum response possible in the system. This may not be the same as the maximum response that a particular agonist actually produces. The parameter K_E is the concentration of [AR] that elicits half the maximal tissue response. The efficacy of an agonist is determined by both K_E and the total receptor density of the tissue ([R_T]). Black and Leff combined those two parameters into a ratio ([R_T]/K_E) and

called this parameter tau (τ), the "transducer constant". Combining the hyperbolic occupancy equation with the hyperbolic transducer function yields an explicit equation describing the effect at any concentration of agonist:

$$\text{Effect} = \frac{\text{Effect}_{\max} \cdot \tau \cdot [A]}{(K_A + [A]) + \tau \cdot [A]}$$

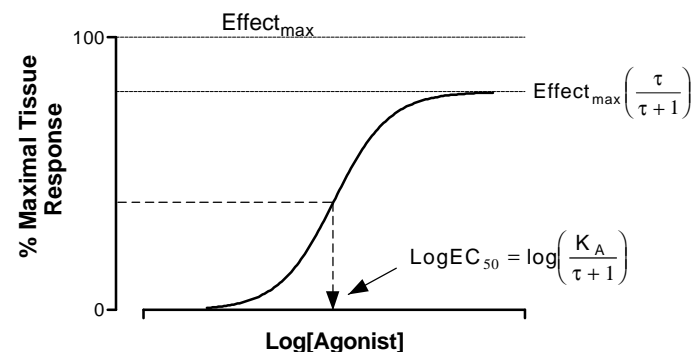
This equation can be rewritten as follows, to make it easier to compare the operational model with the equation of agonist occupancy of receptors.

$$\text{Effect} = \frac{\text{Effect}_{\max} \cdot \tau \cdot [A]}{K_A + [A] \cdot (\tau + 1)} = \frac{[A] \cdot \left(\text{Effect}_{\max} \cdot \frac{\tau}{\tau + 1} \right)}{\frac{K_A}{\tau + 1} + [A]}$$

This form of the equation makes it clear that the maximum effect seen with a particular agonist is not Effect_{max}, but rather is Effect_{max} multiplied by tau/(tau + 1). Only a full agonist in a tissue with plenty of receptors (high values of τ) will yield a maximum response that approaches Effect_{max}.

The EC₅₀ does not equal K_A (the equilibrium dissociation constant for agonist binding to the receptors) but rather K_A/(1 + τ). With a strong agonist, you'll get half-maximal response by binding fewer than half the receptors, so the EC₅₀ will be much less than K_A.

This figure shows a dose-response curve for a partial agonist, and shows the relationship between EC₅₀ and maximum response to terms in the operational model.



The parameter, tau, is a practical measure of efficacy. It equals the total concentration of receptors in the system divided by the concentration of receptors that need to be occupied by agonist to provoke a half-maximal tissue response. The tau value is the inverse of the fraction of receptors that

must be occupied to obtain half-maximal response. If tau equals 10, that means that occupation of only 10% of the receptors leads to a half-maximal response. If tau equals 1.0, that means that it requires occupation of all the receptors to give a half-maximal response. This would happen with a partial agonist or with a full agonist in a tissue where the receptors had been significantly depleted. Because tau is a property of *both* the tissue and receptor system, it is not a direct measure of *intrinsic efficacy*, which is commonly defined as a property belonging only to an agonist-receptor pair, irrespective of the assay system in which it is measured.

The equations here show agonist stimulated response, so the curves all begin at zero. It is easy to add a basal term to model observed response, so the response with no agonist equals basal rather than zero.

Shallower and steeper dose-response curves

Some dose-response curves are steeper or shallower than a sigmoid curve with standard slope. The operational model can be extended to analyze these curves.

If you assume the initial binding of the agonist to the receptor follows the law of mass-action (hill slope equals 1 for the binding step), then transduction step(s) between occupancy and final response must follow an equation that allows for variable slope. If the dose-response curve is still symmetrical and sigmoid, then the operational model can be extended fairly simply, by including a slope parameter, n. The extended form of the operational model is:

$$\text{Effect} = \frac{\text{Effect}_{\max} \cdot \tau^n \cdot [A]^n}{(K_A + [A])^n + \tau^n \cdot [A]^n}$$

The relationship between this operational model and the variable slope sigmoid equation are as follows:

$$EC_{50} = \frac{K_A}{(2 + \tau^n)^{1/n} - 1}$$

$$E_{\max} = \frac{\text{Effect}_{\max} \cdot \tau^n}{\tau^n + 1}$$

When n equals 1, the equation is the same as those shown earlier, describing dose-response curves with Hill slopes of 1.0. However, n is *not* always the same as the Hill Slope (but the two values will be close for full agonists).

Designing experiments to fit to the operational model

If you try to fit the operational model equation indiscriminately to dose-response data, you'll run into a problem. Either the curve-fitting program will report an error message, or it will report best-fit values with enormously wide confidence intervals.

Any symmetrical dose-response curve is defined by four parameters: Bottom (response with no agonist), Top (response at very high concentrations), EC₅₀ (concentration of agonist needed to provoke a response halfway between Bottom and Top) and the Hill Slope. However, the operational model equation has five parameters: Basal (response with no agonist), K_A (dissociation constant of agonist binding), Effect_{max} (maximum possible effect with a full agonist and plenty of receptors), τ (a measure of agonist efficacy) and n.

Since the operational model has more parameters than are needed to describe a sigmoid dose-response curve, any curve can be defined by an infinite combination of operational model parameters. Even if a curve fitting program could find best-fit values (rather than report an error message), the best-fit parameter estimates may not be correct.

To fit the operational model to data, you must compare curves. The most common approach has been to reduce the receptor number, usually with an irreversible alkylating agent, to such an extent that a full agonist can no longer produce the maximal tissue response, no matter how high a concentration is used. The agonist curve before alkylation is then compared to the curve after alkylation. Alkylation is not required with partial agonists. Instead, the dose-response curve of the partial agonist curve is compared to the dose-response curve of the full agonist.

All the dose-response curves should be obtained in the same tissue, in order to reduce the variability in Effect_{max} estimates that can occur between tissues.

Fitting the operational model with Prism

To fit the operational model, the analysis must simultaneously account for the dose-response curve of a full agonist as well as one or more curves from a partial agonist or a full agonist in receptor-depleted tissue.

Some investigators have fit the two (or more) curves at once, using a nonlinear regression program that can fit several data sets at once (Leff et al., J. Pharmacol. Meth., 23: 225-237, 1990). However, Prism can only fit one curve at a time (it will fit a whole family of curves as part of one analysis, but each fit is mathematically independent). Use the following approach (devised by A. Christopoulos).

The first step is to fit the dose-response curve for the full agonist. Use the sigmoidal dose-response (variable slope) equation using nonlinear

regression. Record the best-fit values for Top, Bottom and HillSlope. (The EC₅₀ is not relevant.)

The second step is to fit the dose-response curve for each partial agonist (or full agonist in receptor depleted tissue) to the operational model written as a user-defined Prism equation. When writing the equation for Prism, consider these three points.

- When writing any model for data analysis, you should arrange the parameters so that the uncertainty is symmetrical and Gaussian. See "Why Prism fits the logEC₅₀ rather than EC₅₀" on page 301. If you fit to the logarithm of K_a and tau, the uncertainty is more symmetrical (and more Gaussian) than it would be if you fit to K_a and tau (see A. Christopoulos, Trends Pharmacol. Sci, 19:351-357, 1998).
- Since concentrations are equally spaced on a log scale, enter data with X equal to the logarithm of the agonist concentration.
- You may measure a "response" even in the absence of agonist. So we'll include a Basal parameter in the model. Basal is the measured response in the absence of agonist. If there is no basal activity, then fix Basal to a constant value of zero.

Here is the operational model written as a Prism equation.

```
operate= (((10^logKA)+(10^X))/(10^(logtau+X)))^n
Y=Basal + (Effectmax-Basal)/(1+10^operate)
```

Variable	Comment
Effect _{max}	Set constant to value of Top fit from the sigmoid curve of the full agonist (because the top plateau in a sigmoid dose-response curve is very close to Effect _{max} for a full agonist).
n	Set constant to the value of the Hill slope of the full agonist (which is very close to the value of n in the operational model).
Basal	Set constant to the value of Bottom determined by fitting the sigmoid curve.
logKA	Make the initial value equal to equal to XMID (half way between the highest and lowest X values). An alternative approach is to make the initial value equal to the logEC ₅₀ value obtained by fitting the data to a standard dose-response curve.

logTau Make the initial value zero, so tau equals 1. This value of tau corresponds to a dose-response curve that plateaus at half Effect_{max}. An alternative is to calculate the initial value using the equation below. Top is the maximum response of the partial agonist (the top plateau from the dose-response curve of the partial agonist). Effect_{max} is the top plateau of the full agonist. Hill Slope also comes from the fit of the full agonist to the variable slope equation.

$$\log\tau = \log \left[\left(\frac{\text{Top}}{\text{Effect}_{\max} - \text{Top}} \right)^{1/\text{HillSlope}} \right]$$

Since you fix the first three parameters to constant values, nonlinear regression will find values for logK_A and logTau.

Dose-response curves in the presence of antagonists

Competitive antagonists

The term *antagonist* refers to any drug that will block, or partially block, a response. When investigating an antagonist the first thing to check is whether the antagonism is surmountable by increasing the concentration of agonist. The next thing to ask is whether the antagonism is reversible. After washing away antagonist, does agonist regain response? If an antagonist is surmountable and reversible, it is likely to be competitive (see next paragraph). Investigations of antagonists that are not surmountable or reversible are beyond the scope of this manual.

A competitive antagonist binds reversibly to the same receptor as the agonist. A dose-response curve performed in the presence of a fixed concentration of antagonist will be shifted to the right, with the same maximum response and (generally) the same shape.

Gaddum derived the equation that describes receptor occupancy by agonist in the presence of a competitive antagonist. The agonist is drug A. Its concentration is [A] and its dissociation constant is K_a. The antagonist is called drug B, so its concentration is [B] and dissociation constant is K_b. If the two drugs compete for the same receptors, fractional occupancy by agonist (f) equals:

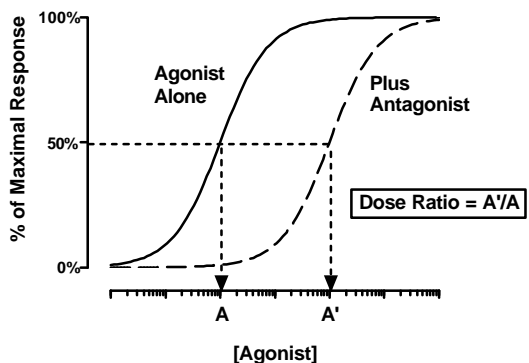
$$f = \frac{[A]}{[A] + K_a \left(1 + \frac{[B]}{K_b} \right)}$$

The presence of antagonist increases the EC₅₀ by a factor equal to 1 + [B]/K_b. This is called the dose-ratio.

You don't have to know the relationship between agonist occupancy and response for the equation above to be useful in analyzing dose response curves. You don't have to know what fraction of the receptors is occupied at the EC₅₀ (and it doesn't have to be 50%). Whatever that occupancy, you'll get the same occupancy (and thus the same response) in the presence of antagonist when the agonist concentration is multiplied by the dose-ratio.

The graph below illustrates this point. If concentration A of agonist gives a certain response in the absence of antagonist, but concentration A' is needed to achieve the same response in the presence of a certain concentration of antagonist, then the dose-ratio equals A'/A. You'll get a different dose ratio if you use a different concentration of antagonist.

If the two curves are parallel, you can assess the dose-ratio at any point. However, you'll get the most accurate results by calculating the dose-ratio as the EC₅₀ in the presence of antagonist divided by the EC₅₀ in the absence of antagonist. The figure below shows the calculation of dose ratio.



Schild plot

If the antagonist is competitive, the dose ratio equals one plus the ratio of the concentration of antagonist divided by its K_d for the receptor. (The dissociation constant of the antagonist is sometimes called K_b and sometimes called K_d)

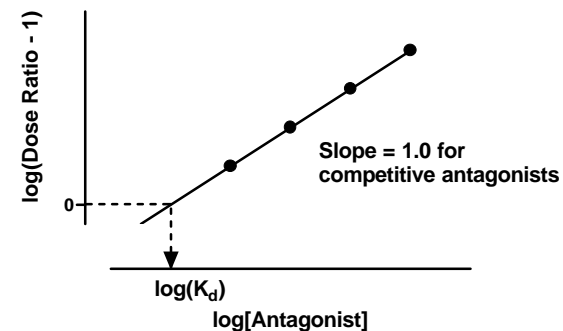
$$\text{Dose ratio} = 1 + \frac{[B]}{K_b} = 1 + \frac{[\text{Antagonist}]}{K_d}$$

A simple rearrangement gives:

$$\text{Dose ratio} - 1 = \frac{[\text{Antagonist}]}{K_d}$$

$$\log(\text{dose ratio} - 1) = \log([\text{Antagonist}]) - \log(K_d)$$

If you perform experiments with several concentrations of antagonist, you can create a graph with log(antagonist) on the X-axis and log(dose ratio - 1) on the Y-axis. If the antagonist is competitive, you expect a slope of 1.0 and the X-intercept and Y-intercept will both equal the K_d of the antagonist.



If the agonist and antagonist are competitive, the Schild plot will have a slope of 1.0 and the X intercept will equal the logarithm of the K_d of the antagonist. If the X-axis of a Schild plot is plotted as log(molar), then minus one times the intercept is called the pA₂ (p for logarithm, like pH; A for antagonist; 2 for the dose ratio when the concentration of antagonist equals the pA₂). The pA₂ (derived from functional experiments) will equal the K_d from binding experiments if antagonist and agonist compete for binding to a single class of receptor sites.

Creating and analyzing Schild plots with Prism

Enter your dose-response data with X as log of the agonist concentration, and Y as response. (If you enter your data with X as concentration, do a transform to create a table where X is log of agonist concentration). Label each Y column with a heading (title) that is the log of antagonist concentration. The first column should be the control, with agonist only (no antagonist). Label this column "control".

Use nonlinear regression to fit a sigmoid dose-response curve. Choose a standard slope or variable slope, depending on your data. From the nonlinear regression dialog, check the option to calculate dose-ratios for Schild plots.

The values of the dose ratio can only be interpreted if all the dose-response curves are parallel. If you selected the sigmoid curve with a standard slope,

this will be true by definition. If you let Prism determine the slope factor for each curve, look at these (and their standard errors) to see if they differ significantly. If the slope factors differ, then the interaction is probably not strictly competitive, and Schild analysis won't be useful. If the slope factors are indistinguishable, consider holding all the slope factors constant to a single value.

The curve fit results include a results view called Summary table which tabulates the log(DR-1) for each data set (except the first, which is the control). To graph these data, go to the graph section and click the button New graph. Choose a new graph from the summary table of the nonlinear regression results.

First fit to linear regression to determine slope and intercept. If the antagonist is competitive, the Schild plot ought to have a slope that is indistinguishable from 1.0. You can check this assumption by seeing whether the confidence interval for the slope includes 1.0.

If the confidence interval for the slope does not include 1.0, your antagonist is probably not a simple competitive antagonist. For suggestions of further analyses, see T. Kenakin, Pharmacologic Analysis of Drug-Receptor Interaction, 3rd Ed. Lippincott-Raven Press, 1997.

If the confidence interval does include 1.0, refit the line constraining the slope to equal 1.0. You cannot do this with Prism's linear regression analysis. However, you can use Prism's nonlinear regression to fit a line with a constant slope. Use this equation:

$$Y = X - pA2$$

When $X = pA2$, $Y = 0$. As X increases above $pA2$, Y increases as well the same amount. Fit this equation to determine the $pA2$ of the antagonist.

An alternative to Schild plots

Analyzing a Schild plot with linear regression of dose-ratios is not the best way to determine the K_d of the antagonist from functional data. The problem is that the EC_{50} of the control curve is used to compute dose ratios for all other curves. Any error in that control value shows up in all the data points. The Schild plot was developed in an era when nonlinear regression was unavailable, so it was necessary to transform data to a linear form. This is no longer an advantage, and Schild plots can be thought of in the same category as Scatchard plots. See "Avoid Scatchard, Lineweaver-Burke and similar transforms" on page 142.

Lew and Angus (Trends Pharmacol. Sci., 16:328-337, 1995) have presented an alternative method for analyzing Schild experiments using nonlinear regression instead of the linear regression method of standard Schild

analysis. This alternative method also avoids the need to calculate dose ratios.

Start with the Gaddum equation for occupancy as a function of agonist and antagonist concentrations:

$$f = \frac{[A]}{[A] + K_a \left[1 + \frac{[B]}{K_b} \right]}$$

Simple algebra expresses the equation this way:

$$f = \frac{1}{1 + \frac{K_a}{K_b} \cdot \frac{[B] + K_b}{[A]}}$$

Thus you can obtain any particular occupancy f , with any concentration of antagonist ($[B]$) so long as you adjust A to keep the quantity in the parentheses constant (C).

$$\frac{[B] + K_b}{[A]} = C$$

Rearrange to show how you must change the agonist concentration to have the same response in the presence of an antagonist.

$$[A] = \frac{[B] + K_b}{C}$$

The EC_{50} is the concentration needed to obtain 50% of the maximal response. You don't know the fraction of receptors occupied at that concentration of agonist, but can assume that the same fractional occupancy by agonist leads to the same response, regardless of the presence of antagonist. So you can express the equation above to define EC_{50} as a function of the antagonist concentration $[B]$.

$$EC_{50} = \frac{[B] + K_b}{C}$$

You determined the EC_{50} at several concentrations of antagonist (including 0), so you could fit this equation to your data to determine a best-fit value of K_b (and C , which you don't really care about). But it is better to write the equation in terms of the logarithm of EC_{50} , because the uncertainty is more symmetrical on a log scale. See "Why Prism fits the log EC_{50} rather than EC_{50} " on page 301. By tradition, we use the negative logarithm of EC_{50} , called the pEC_{50} . For similar reason, you want to determine the best-fit

value of $\log(K_b)$ (logarithm of the dissociation constant of the antagonist) rather than K_b itself.

$$pEC_{50} = -\log([B] + 10^{\log(K_b)}) - \log(C)$$

Define Y to be the pEC_{50} , X to be the antagonist concentration [B], and a new constant P to be $\log(c)$. Now you have an equation you can use to fit data:

$$Y = -\log(X + 10^{\log(K_b)}) - P$$

To determine the K_b using Prism:

1. Determine the EC_{50} of the antagonist in the presence of several concentrations of antagonist, including zero concentration. Enter these values into a Prism data table. Into the X column, enter the antagonist concentrations in micromolar. Into the Y columns, enter the negative log of the EC_{50} .
2. Use nonlinear regression to fit this equation.

$$Y = -1 * \log((X * 1e-6) + (10^{\log K_b})) - P$$

The above equation may also be extended to allow for an assessment of the *conformity* of the data to a model of simple competition. This is analogous to testing the slope of a Schild plot for deviation from a value of one, and simply requires the inclusion of a Schild slope parameter in the equation, as follows:

$$pEC_{50} = -\log([B]^{\text{slope}} + 10^{\log K}) - \log(c)$$

Note that the parameter $\log K_b$ in the first equation has been replaced with $\log K$ in the above equation. This is because, strictly speaking, if the value for the slope is significantly different than one, then the antagonist fitting parameter is *not* the log of the K_b !

Enter the following user-defined equation into Prism:

$$Y = -1 * \log((X * 1e-6)^{\text{slope}} + (10^{\log K})) - P$$

When performing this analysis, it is a good idea to use Prism to fit the data to both equations at the same time and allow the program to decide, via the F-test, which one is the more appropriate equation. If the simpler equation is the better equation, then the $\log K_b$ estimate may be quoted. Otherwise, you must conclude that your data are not consistent with a model of simple competition.

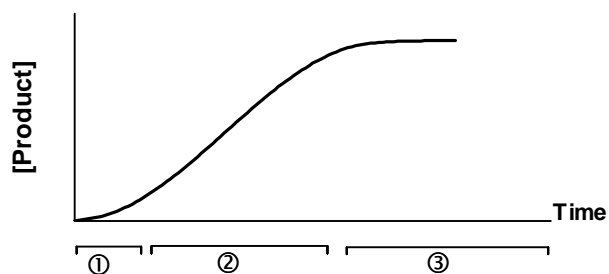
Analyzing enzyme kinetic data

Introduction to enzyme kinetics

Living systems depend on chemical reactions which, on their own, would occur at extremely slow rates. Enzymes are catalysts which reduce the needed activation energy so these reactions proceed at rates that are useful to the cell.

Product accumulation is often linear with time

In most cases, an enzyme converts one chemical (the *substrate*), into another (the *product*). A graph of product concentration vs. time follows three phases as shown in the following graph.



1. At very early time points, the rate of product accumulation increases over time. Special techniques are needed to study the early kinetics of enzyme action, since this transient phase usually lasts less than a second (the figure greatly exaggerates the first phase).
2. For an extended period of time, the product concentration increases linearly with time.
3. At later times, the substrate is depleted, so the curve starts to level off. Eventually the concentration of product reaches a plateau and doesn't change with time.

It is difficult to fit a curve to a graph of product as a function of time, even if you use a simplified model that ignores the transient phase and assumes

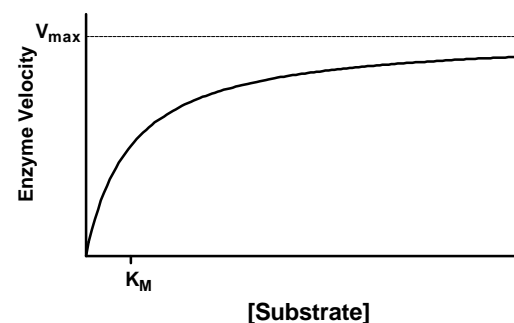
that the reaction is irreversible. The model simply cannot be solved to an equation that expresses product concentration as a function of time. To fit these kind of data (called an *enzyme progress curve*) you need to use a program that can fit data to a model defined by differential equations or by an implicit equation. Prism cannot do this. For more details, see RG Duggleby, "Analysis of Enzyme Reaction Progress Curves by Nonlinear Regression", *Methods in Enzymology*, 249: 61-60, 1995.

Rather than fit the enzyme progress curve, most analyses of enzyme kinetics fit the initial velocity of the enzyme reaction as a function of substrate concentration. The velocity of the enzyme reaction is the slope of the linear phase, expressed as amount of product formed per time. If the initial transient phase is very short, you can simply measure product formed at a single time, and define the velocity to be the concentration divided by the time interval.

This chapter considers data collected only in the second phase. The terminology describing these phases can be confusing. The second phase is often called the "initial rate", ignoring the short transient phase that precedes it. It is also called "steady state", because the concentration of enzyme-substrate complex doesn't change. However, the concentration of product accumulates, so the system is not truly at steady state until, much later, the concentration of product truly doesn't change over time.

Enzyme velocity as a function of substrate concentration

If you measure enzyme velocity at many different concentrations of substrate, the graph generally looks like this:



Enzyme velocity as a function of substrate concentration often follows the Michaelis-Menten equation:

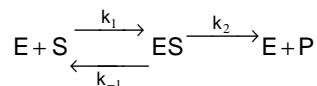
$$\text{Velocity} = V = \frac{V_{\max}[S]}{[S] + K_M}$$

V_{\max} is the limiting velocity as substrate concentrations get very large. V_{\max} (and V) are expressed in units of product formed per time. If you know the molar concentration of enzyme, you can divide the observed velocity by the concentration of enzyme sites in the assay, and express V_{\max} as units of moles of product formed per second per mole of enzyme sites. This is the *turnover number*, the number of molecules of substrate converted to product by one enzyme site per second. In defining enzyme concentration, distinguish the concentration of enzyme molecules and concentration of enzyme sites (if the enzyme is a dimer with two active sites, the molar concentration of sites is twice the molar concentration of enzyme).

K_M is expressed in units of concentration, usually in Molar units. K_M is the concentration of substrate that leads to half-maximal velocity. To prove this, set $[S]$ equal to K_M in the equation above. Cancel terms and you'll see that $V = V_{\max}/2$.

The meaning of K_M

To understand the meaning of K_M , you need to have a model of enzyme action. The simplest model is the classic model of Michaelis and Menten, which has proven useful with many kinds of enzymes.



The substrate (S) binds reversibly to the enzyme (E) in the first reaction. In most cases, you can't measure this step. What you measure is production of product (P), created by the second reaction.

From the model, we want to derive an equation that describes the rate of enzyme activity (amount of product formed per time interval) as a function of substrate concentration.

The rate of product formation equals the rate at which ES turns into E + P, which equals k_2 times $[ES]$. This equation isn't helpful, because we don't know ES. We need to solve for ES in terms of the other quantities. This calculation can be greatly simplified by making two reasonable assumptions. First, we assume that the concentration of ES is steady during the time intervals used for enzyme kinetic work. That means that the rate of ES formation, equals the rate of ES dissolution (either back to E + S or forward to E + P). Second, we assume that the reverse reaction (formation of ES from E + P) is negligible, because we are working at early time points where the concentration of product is very low.

Rate of ES formation = Rate of ES dissolution

$$k_1 \cdot [S] \cdot [E_{\text{free}}] = k_{-1} \cdot [ES] + k_2 \cdot [ES]$$

We also know that the total concentration of enzyme, E_{total} , equals ES plus E. So the equation can be rewritten.

$$k_1 \cdot [S] \cdot ([E_{\text{total}}] - [ES]) = k_{-1} \cdot [ES] + k_2 \cdot [ES]$$

Solving for ES:

$$[ES] = \frac{k_1 \cdot [E_{\text{total}}][S]}{k_1 \cdot [S] + k_2 + k_{-1}} = \frac{[E_{\text{total}}][S]}{[S] + \frac{k_2 + k_{-1}}{k_1}}$$

The velocity of the enzyme reaction therefore is:

$$\text{Velocity} = k_2 \cdot [ES] = \frac{k_2 \cdot [E_{\text{total}}][S]}{[S] + \frac{k_2 + k_{-1}}{k_1}}$$

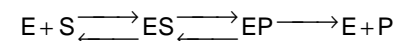
Finally, define V_{\max} (the velocity at maximal concentrations of substrate) as k_2 times E_{total} , and K_M , the Michaelis-Menten constant, as $(k_2 + k_{-1})/k_1$.

Substituting:

$$\text{Velocity} = V = \frac{V_{\max}[S]}{[S] + K_M}$$

Note that K_M is not a binding constant that measures the strength of binding between the enzyme and substrate. Its value includes the affinity of substrate for enzyme, but also the rate at which the substrate bound to the enzyme is converted to product. Only if k_2 is much smaller than k_{-1} will K_M equal a binding affinity.

The Michaelis-Menten model is too simple for many purposes. The Briggs-Haldane model has proven more useful:



Under the Briggs-Haldane model, the graph of enzyme velocity vs. substrate looks the same as under the Michaelis-Menten model, but K_M is defined as a combination of all five of the rate constants in the model.

Assumptions of enzyme kinetic analyses

Standard analyses of enzyme kinetics (the only kind discussed here) assume:

- The production of product is linear with time during the time interval used.
- The concentration of substrate vastly exceeds the concentration of enzyme. This means that the free concentration of substrate is very close to the concentration you added, and that substrate concentration is constant throughout the assay.
- A single enzyme forms the product.
- There is negligible spontaneous creation of product without enzyme
- No cooperativity. Binding of substrate to one enzyme binding site doesn't influence the affinity or activity of an adjacent site.
- Neither substrate nor product acts as an allosteric modulator to alter the enzyme velocity.

How to determine V_{\max} and K_M

To determine V_{\max} and K_M with Prism:

1. Enter substrate concentrations into the X column and velocity into the Y column (entering replicates if you have them).
2. Click Analyze and choose built-in analyses. Pick nonlinear regression from the list of curves and regressions.
3. Choose more equations. Enter this equation as a new equation, or choose from the enzyme kinetics equation library.

$$Y = (V_{\max} * X) / (K_M + X)$$

Variable	Comment
X	Substrate concentration. Usually expressed in μM or mM .
Y	Enzyme velocity in units of concentration of product per time. It is sometimes normalized to enzyme concentration, so concentration of product per time per concentration of enzyme.
V_{\max}	The maximum enzyme velocity. A reasonable rule for choosing an initial value might be that V_{\max} equals 1.0 times Y_{MAX} . V_{\max} is expressed in the same units as the Y values.
K_M	The Michaelis-Menten constant. A reasonable rule for choosing an initial value might be $0.2 * X_{\text{MAX}}$

Checklist for enzyme kinetics

When evaluating results of enzyme kinetics, ask yourself these questions:

Question	Comment
Was only a small fraction of the substrate converted to product?	The analysis assumes that the free concentration of substrate is almost identical to the concentration you added during the time course of the assay. You can test this by comparing the lowest concentration of substrate used in the assay with the concentration of product created at that concentration.
Is the production of product linear with time?	Check the concentration of product at several times to test this assumption.
Did you use high enough concentrations of substrate?	Calculate the ratio of the highest substrate concentration you used divided by the best-fit value of K_M (both in the same concentration units). Ideally, the highest concentration should be at least 10 times the K_M .
Are the standard errors too large? Are the confidence intervals too wide?	Divide the SE of the V_{\max} by the V_{\max} , and divide the SE of the K_M by the K_M . If either ratio is much larger than about 20%, look further to try to find out why.
Is product produced in the absence of enzyme?	The analysis assumes that all product formation is due to the enzyme. If some product is produced spontaneously, you'll need to do a fancier analysis to account for this.
Did you pick a time point at which enzyme velocity is constant.	Measure product formation at several time points straddling the time used for the assay. The graph of product concentration vs. time should be linear.
Is there any evidence of cooperativity?	The standard analysis assumes no cooperativity. This means that binding of substrate to one binding site does not alter binding of substrate to another binding pocket. Since many enzymes are multimeric, this assumption is often not true. If the graph of V vs. [S] looks sigmoidal, see "Allosteric enzymes" on page 325.

Comparison of enzyme kinetics with radioligand binding

The Michaelis-Menten equation for enzyme activity has a form similar to the equation describing equilibrium binding.

$$\text{Enzyme Velocity} = V = \frac{V_{\max}[S]}{[S] + K_M}$$

$$\text{Specific Binding} = B = \frac{B_{\max}[L]}{[L] + K_D}$$

Note these differences between binding experiments and enzyme kinetics.

- It usually takes many minutes or hours for a receptor incubation to equilibrate. It is common (and informative) to measure the kinetics prior to equilibrium. Enzyme assays reach steady state (defined as constant rate of product accumulation) typically in a few seconds. It is uncommon to measure the kinetics of the transient phase before that, although you can learn a lot by studying those transient kinetics (see an advanced text of enzyme kinetics for details).
- The equation used to analyze binding data is valid at equilibrium - when the rate of receptor-ligand complex formation equals the rate of dissociation. The equation used to analyze enzyme kinetic data is valid when the rate of product formation is constant, so product accumulates at a constant rate. But the overall system is not at equilibrium in enzyme reactions, as the concentration of product is continually increasing.
- K_D is a dissociation constant that measures the strength of binding between receptor and ligand. K_M is not a binding constant. Its value includes the affinity of substrate for enzyme, but also the kinetics by which the substrate bound to the enzyme is converted to product
- B_{\max} is measured as the number of binding sites normalized to amount of tissue, often fmol per milligram, or sites/cell. V_{\max} is measured as moles of product produced per minute.

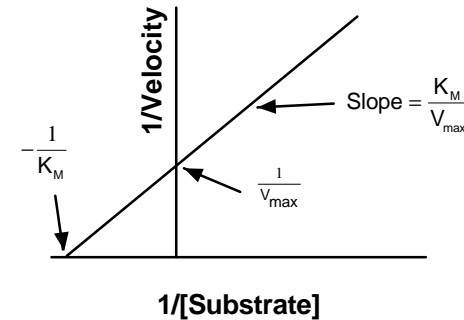
Displaying enzyme kinetic data on a Lineweaver-Burke plot

The best way to analyze enzyme kinetic data is to fit the data directly to the Michaelis-Menten equation using nonlinear regression. Before nonlinear regression was available, investigators had to transform curved data into straight lines, so they could analyze with linear regression.

One way to do this is with a Lineweaver-Burke plot. Take the inverse of the Michaelis-Menten equation and simplify:

$$\frac{1}{V} = \frac{[S] + K_M}{V_{\max}[S]} = \frac{[S]}{V_{\max}[S]} + \frac{K_M}{V_{\max}[S]} = \frac{1}{V_{\max}} + \frac{K_M}{V_{\max}} \cdot \frac{1}{[S]}$$

Ignoring experimental error, a plot of $1/V$ vs. $1/S$ will be linear, with a Y-intercept of $1/V_{\max}$ and a slope equal to K_M/V_{\max} . The X-intercept equals $-1/K_M$.



Use the Lineweaver-Burke plot only to display your data. Don't use the slope and intercept of a linear regression line to determine values for V_{\max} and K_M . If you do this, you won't get the most accurate values for V_{\max} and K_M . The problem is that the transformations (reciprocals) distort the experimental error, so the double-reciprocal plot does not obey the assumptions of linear regression. Use nonlinear regression to obtain the most accurate values of K_M and V_{\max} (see "Avoid Scatchard, Lineweaver-Burke and similar transforms" on page 142).

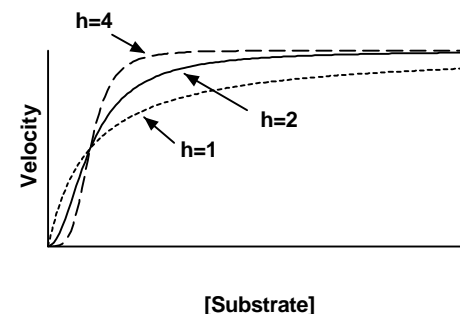
Tip. You should analyze enzyme kinetic data with nonlinear regression, **not** with Lineweaver-Burke plots. Use Lineweaver-Burke plots to display data, not to analyze data.

To create a Lineweaver-Burke plot with Prism, start from a table where X is substrate concentration and Y is velocity. Click Analyze, and choose a built-in analysis. Then choose Transformations from the list of data manipulations. Check the option boxes to transform both X to be $1/X$, and Y to be $1/Y$. Be sure to check the option to create a new graph of the results.

From that graph, click Analyze and choose linear regression to superimpose the regression line. This linear regression line should NOT be used to obtain values for V_{\max} and K_M . The X-intercept of the regression line will be near $-1/K_M$, and the negative inverse of the slope will be near the V_{\max} . However, the V_{\max} and K_M values determined directly with nonlinear regression will be more accurate. It is better to draw the line that corresponds to the nonlinear regression fit.

To create a Lineweaver-Burke line corresponding to the nonlinear regression fit, follow these steps:

1. Create a new data table, with numerical X values and single Y values.
2. Into row 1 enter $X=0$, $Y=-1/K_M$ (previously determined by nonlinear regression).
3. Into row 2 enter $X=1/S_{max}$ (S_{max} is the largest value of [substrate] you want to include on the graph) and $Y=(1/V_{max})(1.0 + K_M/S_{max})$.
4. Note the name of this data table. Perhaps rename it to something appropriate.
5. Go to the Lineweaver-Burke graph.
6. Press Change, then Data on graph.
7. Add the new data table to the graph.
8. Press Change, then Symbols and lines.
9. Drop down the list of data sets, and select the one you noted in step 4.
10. Choose to plot no symbols, but to connect with a line.



To fit data to the equation for enzyme activity with positive cooperativity, use the equation below. For initial values, try these rules: $V_{max}=Y_{max}$, $K=.5*X_{mid}$, and $h=1.0$

$$Y = V_{max} * X^h / (K^h + X^h)$$

The variable h does not always equal the number of interacting binding sites (although h can not exceed the number of interacting sites). Think of h as an empirical measure of the steepness of the curve and the presence of cooperativity.

Allosteric enzymes

One of the assumptions of Michaelis-Menten kinetics is that there is no cooperativity. If the enzyme is multimeric, then binding of a substrate to one binding site should have no effect on the activity of neighboring sites. This assumption is often not true.

If binding of substrate to one binding site increases the activity of neighboring sites, the term *positive cooperativity* is used. Activity is related to substrate concentration by this equation:

$$\text{Velocity} = V = \frac{V_{max}[S]^h}{[S]^h + K_{0.5}^h}$$

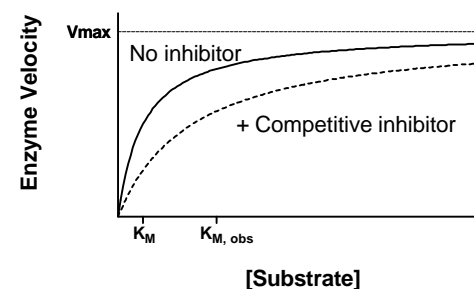
When the variable h equals 1.0, this equation is the same as the Michaelis-Menten equation. With positive cooperativity, h will have a value greater than 1.0. If there are two interacting binding sites, h will equal two (or less, depending on the strength of the cooperativity). If there are three interacting binding sites, h will equal 3 (or less). Note that the denominator has the new variable $K_{0.5}$ instead of K_M .

Enzyme kinetics in the presence of an inhibitor

Competitive inhibitors

If an inhibitor binds reversibly to the same site as the substrate, the inhibition will be competitive. Competitive inhibitors are common in nature.

One way to measure the effect of an inhibitor is to measure enzyme velocity at a variety of substrate concentrations in the presence and absence of an inhibitor. As the graph below shows, the inhibitor substantially reduces enzyme velocity at low concentrations of substrate, but doesn't alter velocity very much at very high concentrations of substrate.



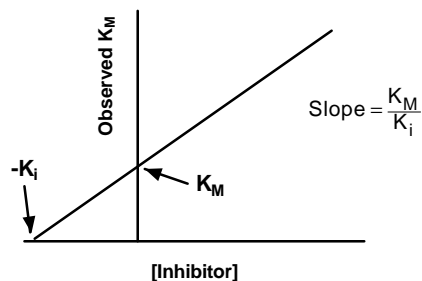
As the graph above shows, the inhibitor does not alter V_{max} , but increases the observed K_M (concentration of substrate that produces half-maximal velocity, in the presence of a competitive inhibitor). The observed K_M is defined by the following equation, where K_i is the dissociation constant for inhibitor binding (in the same concentration units as [Inhibitor]):

$$K_{M,obs} = K_M \cdot \left[1 + \frac{[Inhibitor]}{K_i} \right]$$

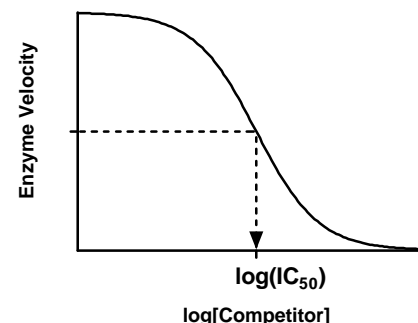
If you have determined the K_m plus and minus a single concentration of inhibitor, you can rearrange that equation to determine the K_i .

$$K_i = \frac{[Inhibitor]}{\frac{K_{M,obs}}{K_M} - 1.0}$$

You'll get a more reliable determination of K_i if you determine the observed K_m at a variety of concentrations of inhibitor. Fit each curve to determine the observed K_m . Enter the results onto a new table, where X is the concentration of inhibitor, and Y is the observed K_m . If the inhibitor is competitive, the graph will be linear. Use linear regression to determine the X- and Y-intercepts. The Y-axis intercept equals the K_M and the X-axis intercept equals the negative K_i .



Another experimental design is to measure enzyme velocity at a single concentration of substrate with varying concentrations of a competitive inhibitor. The results will look like this.



The concentration of competitor that reduces enzyme velocity by half is called the EC_{50} or IC_{50} . Its value is determined by three factors:

- The dissociation constant for binding of inhibitor to enzyme, the K_i . If the K_i is low (the affinity is high), the EC_{50} will be low. The subscript i is used to indicate that the competitor inhibited enzyme activity. It is the concentration of the competitor that will bind to half the enzyme sites at equilibrium in the absence of substrate or other competitors.
- The concentration of the substrate. If you choose to use a higher concentration of substrate, it will take a larger concentration of inhibitor to compete for 50% of the activity.
- The K_M . It takes more inhibitor to compete for a substrate with a low K_M than for a substrate with a high K_M .

Prism calculates the K_i , using the equation of Cheng and Prusoff (Cheng Y., Prusoff W. H., *Biochem. Pharmacol.* 22: 3099-3108, 1973).

$$K_i = \frac{IC_{50}}{1 + \frac{[Substrate]}{K_M}}$$

To determine the K_i with Prism (from data collected with a single concentration of substrate):

1. Enter data with X equal to logarithm of inhibitor concentration and Y equal to velocity.
2. Press Analyze, and choose built-in analyses. From the curves section, choose nonlinear regression.
3. Choose the one-site competitive binding equation.
4. If you want to fix the top and bottom plateaus to constant values, click the Constants button and enter the values. Most often, you will fix the bottom plateau to zero, assuming that there is no enzyme velocity at maximum concentrations of inhibitor.

5. On the nonlinear regression dialog, choose the option " K_i from IC_{50} ". This option is usually used for radioligand binding studies, so enter the K_M where it asks for the K_d , and enter substrate concentration where it asks for radioligand concentration. Enter both in mM (or any concentration units; only the ratio matters). Enter concentrations, not the logarithm of concentrations.

Inhibitors that are not competitive

If an inhibitor binds to a site on the enzyme distinct from the site that binds substrate, the inhibition cannot be overcome by increasing the concentration of substrate. The inhibition is not competitive, and the inhibitor decreases the observed V_{max} (and may also increase the observed K_M). Consult an advanced text on enzyme kinetics for information about *non-competitive*, *uncompetitive*, and mixed inhibition.

Competitive and non-competitive inhibitors bind reversibly. An inhibitor that binds covalently to irreversibly inactivate the enzyme is called an *irreversible inhibitor* or *inactivator*.

Reading unknowns from standard curves

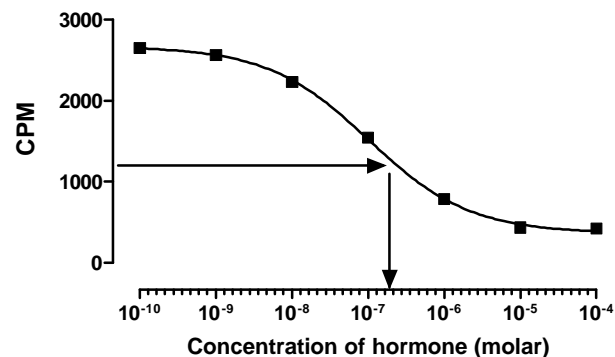
Introduction to standard curves

Standard curves are used to determine the concentration of substances. First you perform an assay with various known concentrations of a substance you are trying to measure. The response might be optical density, luminescence, fluorescence, radioactivity or something else. Graph these data to make a standard curve – concentration on the X axis, and assay measurement on the Y axis.

Also perform the same assay with your unknown samples. You want to know the concentration of the substance in each of these unknown samples.

To analyze the data, fit a line or curve through the standards. 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. Read down the graph until you intersect the X-axis. The concentration of substance in the unknown sample is the value on the X-axis.

In the example below, the unknown sample had 1208 counts per minute, so the concentration of the hormone is 0.236 micromolar.



Prism makes it very easy to fit your standard curve, and to read (interpolate) the concentration of unknown samples.

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:

Creating a standard curve with linear regression

Standard curves are often nearly linear, at least within a certain range of concentrations. If you restrict your standard curve values to a linear region, you can analyze the curve with linear regression. This will be a useful analysis, even if the overall standard curve is not quite straight, so long as you choose a reasonable range. The standard curve should start a little below your lowest unknown value and extend to a little beyond your highest unknown value. There is no benefit to continuing the standard curve far above or below the range of your unknowns.

See "Linear regression" on page 141.

Creating a standard curve with cubic spline (or lowess)

The easiest way to fit a curve is to create a cubic spline or lowess curve. They are easier than nonlinear regression, because you don't have to choose an equation. Spline and lowess curves tend to wiggle too much, so are not often used as standard curves.

See "Fitting a curve without choosing a model" on page 167.

Creating a standard curve with polynomial regression

Polynomial regression is a convenient method to create a smooth curve. With Prism, you perform polynomial regression by choosing a polynomial

equation from the nonlinear regression dialog. Try a second, third or fourth order polynomial equation. The higher order polynomial equations generate standard curves with more inflection points.

See "Polynomial regression" on page 166.

Creating a standard curve with nonlinear regression

Nonlinear regression is often used to fit standard curves generated by radioimmunoassay (RIA) or similar assays (ELISA). 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 sigmoid dose response curve with variable slope. When fitting sigmoid curves, enter the X values as the logarithms of concentrations, not concentrations.

Ordinarily, the choice of an equation is very important when using nonlinear regression. If the equation does not describe a model that makes scientific sense, the results of nonlinear regression won't make sense either. With standard curve calculations, the choice of an equation is less 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.

See "Fitting curves with nonlinear regression" on page 195.

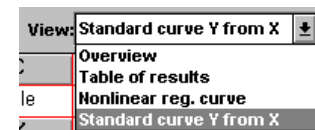
Determining unknown concentrations from standard curves

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 option to perform standard curve calculations.
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, transform the results to anitlogs. Click the Analyze button. On the New Analysis dialog, choose Transforms and choose to analyze the data you are looking at.

Standard curves with replicate unknown values

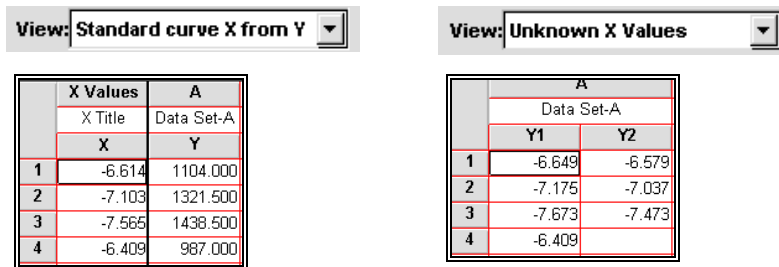
Prism's nonlinear regression analysis can interpolate from a standard curve, even if you have replicate unknown values.

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.



The Y column in “standard curve X from Y” shows the average of the replicate unknown Y values you entered. The X values are the concentrations that correspond to the mean Y values.

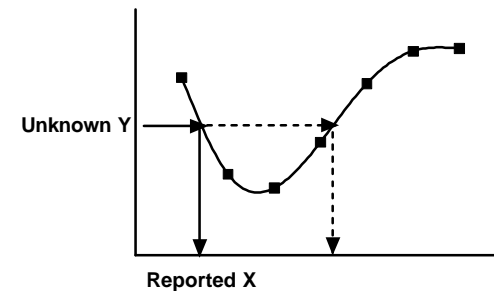
Each value in “unknown X values” is a concentration corresponding to one of the replicate values you entered, and is expressed in the same units as the X axis of your standard curve. Because Prism cannot deal with replicate X values, Prism places these unknown X values in a Y column on the results table. Think of them as X values on your standard curve. But think of them as Y values when you want to do further analyses (such as a transform).

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. With nonlinear regression, Prism only will perform standard curve calculations within the limits that the curve is defined. But you can extend the curve in both directions, to include lower and higher X values. Do this from the Output options dialog – press Output from the nonlinear regression parameters dialog.
- 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 increase the number of line segments, click Output from the nonlinear regression parameters dialog.

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 (Tools menu).

Transforming data

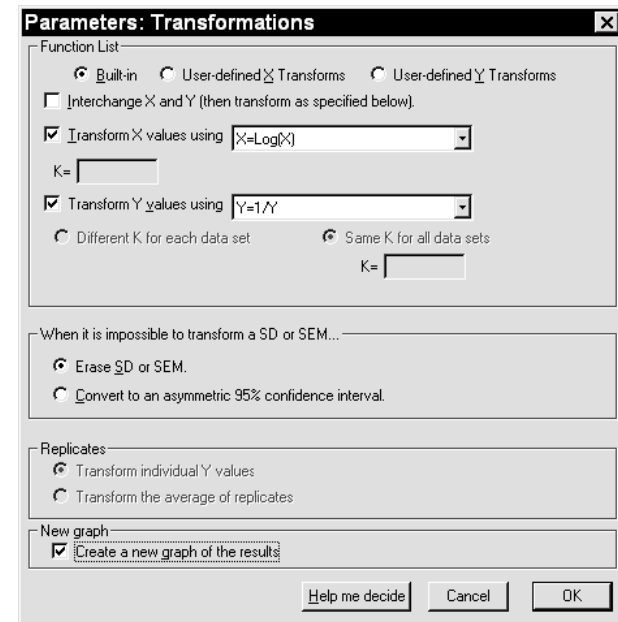
The advantages of transforming data

You'll often find it useful to transform data before performing analyses for the following reasons:

- To change units.
- To make scatter more Gaussian (see "Advantages of transforming the data first" on page 40).
- To avoid using huge or tiny numbers, that can cause numerical problems with some analyses. In general you should avoid analyzing numbers less than 0.000001 or so, or numbers greater than 1000000 or so.
- To convert a curve into a line. This can be useful for displaying data, but should usually be avoided when analyzing data. See "Avoid Scatchard, Lineweaver-Burke and similar transforms" on page 142.

How to transform data

Click Analyze and choose Transform from the list of data manipulations to bring up the dialog.



Prism can interchange X and Y values, and can transform either X or Y values. Prism's transform analysis cannot combine columns, for example creating a new column from the sum of two other columns. Prism has limited ability to combine columns. See "Subtracting (or dividing by) baseline values" on page 341. For more choices, enter your data into Excel or another spreadsheet, use Excel to combine columns, and then link or embed the results into Prism.

Interchange X and Y values

If you choose to interchange X and Y values, Prism can only handle one data set, since 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. If you also selected X or Y transforms, Prism applies these to the data after interchanging. – 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

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.
$Y = Y + K$	"
$Y = Y - K$	"
$Y = Y / K$	"
$Y = Y \text{ squared}$	
$Y = Y ^ K$	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 (inverse of log).
$Y = \exp(Y)$	e^Y (inverse of ln)
$Y = 1/Y$	
$Y = \text{SQRT}(Y)$	Square root.
$Y = \text{LOGIT}(y)$	$\ln(Y/1-Y)$
$Y = \sin(Y)$	Y is in radians.
$Y = \cos(Y)$	"
$Y = \tan(Y)$	"
$Y = \arcsin(Y)$	Result is in radians.
$Y = \text{ABS}(Y)$	Absolute value. If Y is negative, multiply by -1.
$Y = Y + \text{Random}$	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". 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. To enter different K values for each data set, choose a data set, enter K, choose another data set, enter its K, and so on.



Use this option to transform different data sets with different values of K. If you don't want to transform some data sets at all, don't select them for the analysis. See "Changing an analysis" on page 353.

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. Select a user-defined transform you have used before or enter a new one.

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

When you enter your equations, you can use any of the functions listed in "Available functions" on page 184.

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 a 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, Prism can 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

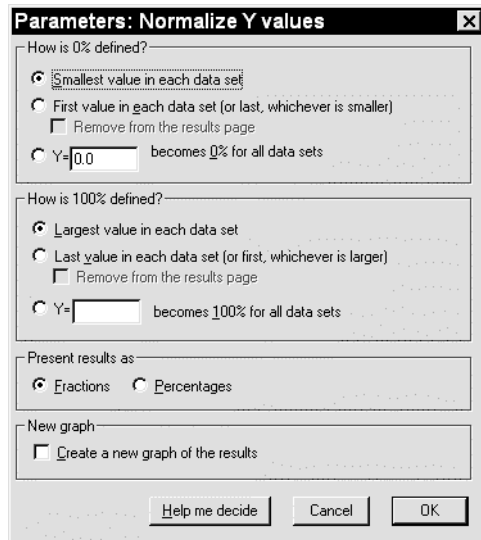
asymmetrical (i.e. logarithms), it is mathematically impossible to transform a SD and end up with a SD. You have two choices. You may either transform the mean only or erase the error bars. Or you may convert the error bars to 95% confidence intervals, and then transform both ends of the confidence interval.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

Normalizing data

Normalize the data to convert Y values from different data sets to a common scale. This is useful when the want to compare the shape or position (EC_{50}) of two or more curves, and don't want to be distracted by different maximum and minimum values.

Investigators who analyze dose-response curves commonly normalize the data so all curves begin at 0% and plateau at 100%. If you then fit a sigmoid dose-response curve to the normalized data, be sure to set the top and bottom plateaus to constant values. If you've defined the top and bottom of the curves by normalizing, you shouldn't ask Prism to fit those parameters.

To normalize, click Analyze and choose built-in analyses. Then select Normalize from the list of data manipulations to bring up this dialog. .



To normalize between 0 and 100%, you must define these baselines. Define zero as the smallest value in each data set, the value in the first row in each data set, or to a value you enter. Define one hundred as the largest value in each data set, the value in the last row in each data set, or a value you enter. Prism can express the results as fractions or percentages.

Notes:

- If you have entered replicate values, zero and one hundred percent are defined by the mean of the replicates. It is not possible to normalize each subcolumn separately.
- 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 become 0.000 and 1.000) from the results.

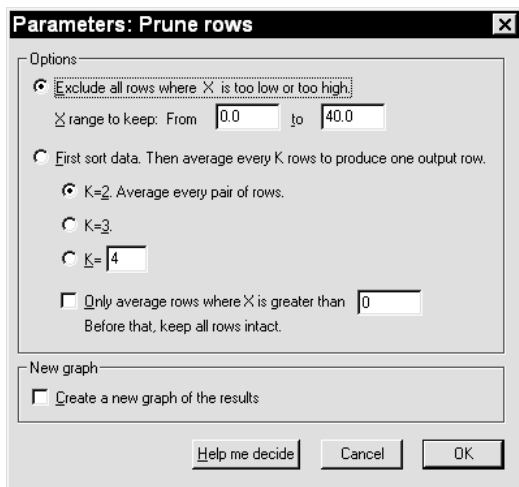
Tip: The Remove Baseline analysis lets you subtract (or divide) all values by the mean of the first few rows. With some experimental designs, this is the best way to normalize.

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. The pruning analysis starts with a large data table, and generates a shorter results table. Another way to deal with large data sets is to decimate data while importing, so Prism only reads every tenth (or some other number) row. See the chapter in importing data in the Prism User's Guide.

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 Delete Sheet command (on the Tools menu) to remove it.

To prune, click Analyze and choose built-in analyses. Then choose Prune from the list of data manipulations to bring up this dialog.

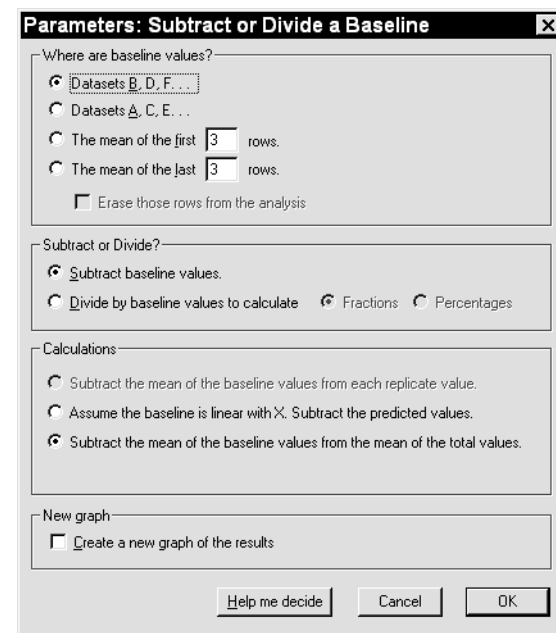


One way to prune data is to remove all rows where X is too low or too high, and keep only rows where X is between limits you enter. The other way to prune is to average every K rows to produce one output row (you enter K). First Prism sorts the table by X (if not already sorted). Then it averages every 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. Optionally average only rows after a threshold X value. When X is lower than 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.

Subtracting (or dividing by) baseline values

Many kinds of data combine a measurement (signal) you care about with a baseline or background (noise) you don't care about. You can analyze these data using two approaches. One approach is to perform analysis on the total signal. The other approach is to subtract or divide by a baseline or nonspecific value and then analyze and graph the results.

Click Analyze and choose built-in analyses. Then choose Remove Baseline from the list of data manipulations to bring up this dialog.



Where are the baseline values?

Pick one of the first two choices when you have measured a baseline value at every value of X, and have placed baseline or nonspecific values in datasets adjacent to the main data. Prism can then subtract column B from column A, and column D from C, etc. The results table will have half as many data sets (Y columns) as the original table, but the same number of rows.

Pick 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. You enter the number of rows to average to define the baseline. If you want to use just the first or last row, tell Prism to "average" the first 1 row.

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.

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 replicates and subtract (divide) this value from each total replicate. If you entered

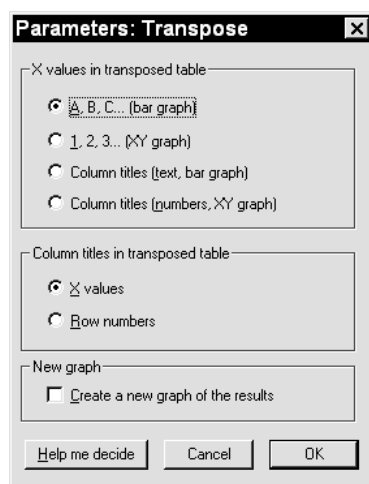
triplicate values, Prism subtracts the mean of the three baseline values from each of the three total values.

- Assume that the baseline is linear with X. Prism performs linear regression with the background (nonspecific) values and then subtracts (or divides) the Y value predicted from that line. This method is appropriate when you know that the nonspecific or background values must be linear with the X values (for example nonspecific binding is often linear with ligand concentration). 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. When Prism fits linear regression, it does not assume that the line goes through the origin and does not display its results.
- 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).

Each row of Y values becomes one column (data set) in the results table. The first row becomes the first data set, the second row becomes the second data set, 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 in the results table are determined by your choices in the dialog.

Transposing rows and columns

Transpose rows and columns to change the appearance of bar graphs. Click Analyze and choose Transpose from the list of data manipulations to bring up this dialog. See “Fitting a curve without choosing a model” on page 167.

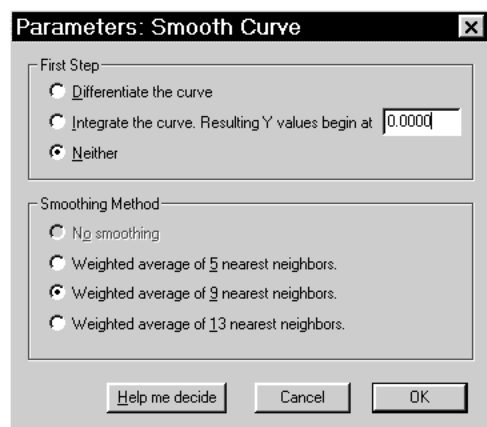


Analyzing continuous data

Smooth, differentiate or integrate a curve

Prism can compute numerical integrals or derivatives, and smooth the results. It can also smooth a curve directly, without computing its derivative or integral.

From a graph of the curve or a data table, click Analyze and choose built-in analyses. Then select "Smooth, differentiate or integrate a curve" to bring up this dialog.



Prism can only smooth data sets (and compute derivatives and integrals) 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).

Computing derivatives and integrals is new to Prism 3. If you open a file in Prism 2, these analyses will be lost unless you freeze the results sheets (Tools menu).

The derivative of a curve

The derivative is the steepness of the curve at every X value. The derivative is positive when the curve heads uphill and is negative when the curve heads downhill. The derivative equals zero at peaks and troughs in the curve. Prism first calculates a simple numerical derivative. For every row i , the resulting X and Y values are:

$$\text{ResultX}(i) = \frac{X(i+1) + X(i)}{2}$$
$$\text{ResultY}(i) = \frac{\Delta Y}{\Delta X} = \frac{Y(i+1) - Y(i)}{X(i+1) - X(i)}$$

After calculating the numerical derivative, Prism can smooth the results, if you choose.

Prism does not do any symbolic algebra, and cannot compute analytical derivatives. If you give Prism a series of XY points that define a curve, it can compute the derivative of that curve. If you give Prism an equation, it cannot compute a new equation that defines the derivative.

The integral of a curve

The integral is the cumulative area under the curve. The integral at any value X equals the area of the curve for all values less than X.

Prism uses the trapezoid rule to integrate curves (see "How Prism calculates the area under a curve" on page 349). The X values of the results are the same as the X values of the data you are analyzing. The first Y value of the results equals a value you specify (usually 0.0). For other rows, the resulting Y value equals the previous result plus the area added to the curve by adding this point. This area equals the difference between X values times the average of the previous and this Y value. Prism uses this equation (i refers to row number):

$$\text{ResultY}(i) = \text{ResultY}(i-1) + \frac{Y(i-1) + Y(i)}{2} \cdot [X(i) - X(i-1)]$$

After doing a simple integration, Prism can then smooth the results, if you choose.

Prism does not do any symbolic calculus, and cannot compute analytical integrals.

Smoothing a curve

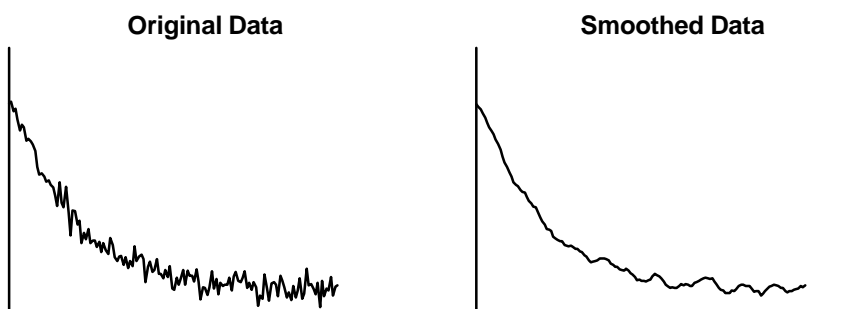
If you import a curve from an instrument, smooth the data to improve the appearance of a graph. The purpose of smoothing is solely to improve the

appearance of a graph. Since you lose data when you smooth a curve, you should not smooth a curve prior to nonlinear regression or other analyses.

There is no point trying to smooth curves created by nonlinear regression, since they are already smooth. It only makes sense to smooth curves collected from an instrument. It can make sense to compute the derivative or integral of a perfect (smooth) when X is time.

Smoothing is not a method of curve fitting. If your goal is to draw a smooth curve through your data, see "Fitting a curve without choosing a model" on page 167.

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. The results table has a few less rows than the original data.



Area under the curve

The area under the curve is an integrated measure of an effect. It is used as a cumulative measure of drug effect in pharmacokinetics, and as a way to compare peaks in chromatography.

Note the difference between integrating a curve and computing area under the curve. When you integrate a curve, the result is another curve showing cumulative area. When you ask Prism to compute area under the curve, it gives you one value for the area under the entire curve as well as the area under well-defined peaks.

If your data come from chromatography or spectroscopy, Prism can break the data into separate regions and find the highest point (the peak) of each.

It can only do this if the regions are clearly defined, with the signal going below the baseline between regions. It cannot separate overlapping peaks.

Click Analyze and choose built-in analyses. Then choose Area under the curve from the list of curve analyses to bring up the parameters dialog.

Parameters for area under curve

- Define the baseline by entering a Y value (usually $Y=0$) or calculate the baseline as the mean of the first and last few values (you define how many). If you want to use the first value as the baseline, tell Prism you want to use the "mean of the first 1 rows". Prism always assumes the baseline is horizontal (parallel to the X axis).
- If your data are noisy, Prism may find too many regions, with some being tiny and irrelevant. 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 distance between the minimum and maximum Y values.
- Optionally define the minimum width of a region worth considering. Give the width as the number of adjacent rows in the data table. Do not use X units.
- Choose whether values below the baseline should be considered to be "negative peaks" or ignored.

If all values are above the baseline, then the dialog choices (except for definition of baseline) are irrelevant. Prism finds one peak, which is the area under the entire curve. This is useful in analysis of pharmacokinetic data.

Interpreting area under the curve

For each region, Prism shows its area in units of the X axis times the units of the Y axis and as a fraction of the area under all regions. Prism also identifies the peak of each region by reporting the X and Y coordinates of the highest point and the X coordinates of the beginning and end of the region.

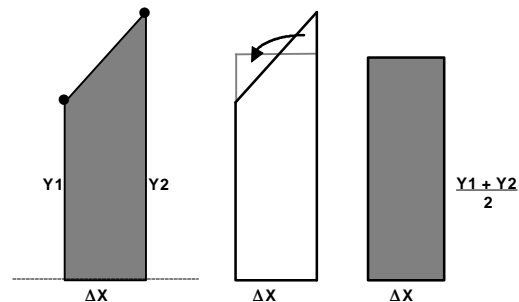
Note the limitations of this analysis:

- 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.

If Prism reports finding more regions than you care about, go back to the parameters dialog and define a larger value for the minimum width or height of a peak.

How Prism calculates the area under a curve

Prism computes the area under the curve using the trapezoid rule, illustrated in the figure below.



In Prism, a curve is simply a series of connected XY points, with equally spaced X values. The left part of the figure below shows two of these points and the baseline as a dotted line. The area under that portion of the curve, a trapezoid, is shaded. The middle portion of the figure shows how Prism computes the area. The two triangles in the middle panel have the same area, so the area of the trapezoid on the left is the same as the area of the rectangle on the right (whose area is easier to calculate). The area, therefore, is $\Delta X * (Y1 + Y2) / 2$. Prism uses this formula repeatedly for each adjacent pair of points defining the curve.

Appendix. Using GraphPad Prism to analyze data

	X Values	A		B		
	X Title	Title		Title		
	X	Y1	Y2	Y1	Y2	
1						
2						
3						
4						

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

The Prism User's Guide

This book explains each analysis that Prism can perform. A separate volume, [The Prism User's Guide](#), explains how to use Prism. This appendix summarizes the basic operation of Prism, so you can use it to analyze and graph your data.

The full text of both volumes is present in the online Help system.

Projects, sections and sheets

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

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.

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.

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.

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.

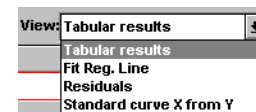
How to analyze data with Prism

To analyze a data table, click the Analyze button. You can use a built-in analysis, apply a stored analysis method (a sequence of Prism analyses), or analyze and graph this data table exactly as you have already analyzed and graphed another table.

You can perform linear and nonlinear regression directly from a graph. While looking at the graph, click Analyze. You can also click Analyze from a results table to chain two analyses. 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.

The analyze button isn't just for statistics and regression, but also for data manipulations such as transforming and normalizing. Prism doesn't alter the original data, but creates a new results table that you can graph or analyze further.

If you analyze from a table, Prism will take you right to the 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. Drop down the list box in the bottom row of the toolbar to select a view.



If you analyze from a graph, Prism will superimpose the regression line or curve on that graph. Click on the yellow results folder tab to see the numerical results.

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. While looking at the results sheet, press the **Change Params.** button to bring up the Parameters dialog with your previous choices. Change settings as needed (for example, pick a different equation for nonlinear regression). Press ok to recalculate the analysis.

You can also change parameters directly from a graph. Press the Change button and choose Analysis parameters.

You can also change which data table and data sets are analyzed. From the results table, press Change and select Data Analyzed.

Analyzing repeated experiments

Science is repetitive by its very nature. Prism provides several approaches to analyzing and graphing repetitive data. You'll use a different approach depending on whether you simply want to repeat analysis and graphing steps within the same project, repeat those analyses and graphs in different projects, or automate analysis of a large number of data files at once.

Repetition within a project

You can easily instruct Prism to analyze and graph data from a repeat experiment exactly as you analyzed and graphed data from the first experiment. From the data table of the second experiment, press Analyze, then choose **Method by Example** and select the data set to use as an example (which must be in the same project).

An alternative approach is to start from the data table of a representative experiment, and duplicate the data along with all of its linked graphs and results. To do this, go to the data table and click the New Table button. Choose to duplicate the current table and its analyses and graphs. Then go to the new table, and replace the data. Because the new data table is linked to new results and analyses, simply go to these new sheets to see the analyses and graphs of the second experiment.

There are advantages to each approach. If you apply a method by example, you don't have to format all the data tables the same way. For example, you could format the first experiment with triplicate values, but the second experiment with quadruplicates. One advantage of duplicating is that you don't have to choose column formats for the second experiment – you simply get a copy of the first data table. Another advantage is that the duplicated table will have column titles (and perhaps X values) that you can keep for the second experiment.

Repetition in ongoing experiments, involving several projects

If you frequently analyze and graph data using the same steps, Prism offers two approaches: **methods** and **templates**. Methods and templates both let you analyze and graph a new data table using steps stored in a file. The difference is in how you use them. You apply a method to an existing data table, but insert a template (which includes an empty data table) into an existing project.

To save a **method file** go to a data table of a representative experiment (already linked to analyses and graphs), pull down the file menu and choose Save Method. To apply the method, go to data table (which can be in a different file) and click Analyze. Choose to apply a method, and then choose a category and method. Prism will analyze and graph the data using the steps stored in the method file.

To create a **template file**, make a project containing nothing but data for a representative experiment, and all necessary analyses and graphs. Then select Save Template from the File menu. You can choose a template to begin a new project (from the Welcome dialog), and can insert a template into an existing project (from the New table dialog). Enter the new data onto the data table from the template, then go to other new sheets to see the new graphs and results.

There are advantages to each approach. If you apply a method, you don't have to format all the data tables the same way. For example, you could define the method using data formatted with triplicate values, but apply it to a table formatted with quadruplicates. The advantage of inserting a template is that you don't have to choose column formats, and don't have to reenter column titles and X values. Another advantage of templates is that they can contain more than one data table. You can create a template with two (or more) tables to analyze and graph an experiment with several parts.

Analyzing a large series of experiments

Methods and templates automate analyses and graphing. If you want to analyze a large series of experiments at once, you will also want to automate data importing. To do this, run a Prism script. Prism scripts tend to be very short, and easy to write since you only need to write script commands to import data, and export results and graphs. The analysis and graphing steps are performed automatically by your Prism project, so don't require scripting commands. See the chapter on scripts in the Prism User's Guide.

Intuitive Biostatistics (book)

If you like the style of this book, 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:

"*Intuitive 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 can see the table of contents and read five complete chapters at www.graphpad.com. 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. *Intuitive Biostatistics* is also available from the online bookstore www.amazon.com.

Technical support

If you need help using Prism, and can't find the answers in this manual, please visit our web site first at www.graphpad.com. We will post updated versions of the program there, along with frequently asked questions and technical notes about using Prism.

If you need personal help, GraphPad Software offers free technical support. While we reserve the right to charge for support in the future, we do promise that you'll receive free support for at least one year. Don't hesitate to contact us.

Email: support@graphpad.com
Web Site: www.graphpad.com
Telephone: 858-457-3909 (619-457-3909 before June 12, 1999)
Fax: 858-457-8141 (619-457-8141 before June 12, 1999)
Mail: 5755 Oberlin Drive #110
San Diego CA 92121

While GraphPad's technical support staff cannot provide any statistical advice, you may email simple questions about data analysis to me. Your Prism license does not include free statistical consulting, but I'll try to answer straightforward questions.

Dr. Harvey Motulsky, President GraphPad Software
hmotulsky@graphpad.com

Index

A

Allosteric enzymes	335
Analyses, changing data sets to analyze	365
Analysis of Covariance	150
Analyzing repeated experiments	365
ANCOVA	150
ANOVA table, from two-way ANOVA	93
ANOVA, missing values	94
ANOVA, one-way, ANOVA table	70
ANOVA, one-way, Bartlett's test	71
ANOVA, one-way, checklist	76
ANOVA, one-way, choosing	65
ANOVA, one-way, entering data	63
ANOVA, one-way, how it works	70
ANOVA, one-way, introduction	63
ANOVA, one-way, results	73
ANOVA, repeated measures, checklist	81
ANOVA, repeated measures, results	79, 80
ANOVA, two way, introduction	89
ANOVA, two-way, ANOVA table	93
ANOVA, two-way, checklist	101
ANOVA, two-way, choosing	90
ANOVA, two-way, entering data	89
ANOVA, two-way, how it works	94
ANOVA, two-way, interaction	96
ANOVA, two-way, interpreting	96
ANOVA, two-way, post tests	98
ANOVA, used to compare curves	96
Area under a curve, integrate	358
Area under the curve	360
Association binding experiments	297
Association curve	178
Averaging replicates	37

B

Bartlett's test for equal variances	71
Baselines, subtracting or dividing	353

Bayesian approach	15
Becquerel	249
Binding equation, one site	171
Binding equation, two sites	172
Binding isotherm (hyperbola)	171
Boltzmann sigmoid curve	176
Bonferroni posttest	69
Bootstrapping methods	6

C

Case-control study	119
Censored survival data	105
Central limit theorem	7
Changing which data sets are analyzed	365
Checklist, for enzyme kinetics	331
Cheng and Prusoff equation	274, 338
Chi-square test for trend	124
Chi-square test, checklist	127
Chi-square test, how it works	123
Chi-square vs. Fisher's test	121
Choosing which data sets to analyze	365
Ci, curie defined	249
Circularity	80, 103
Coefficient of determination, defined	135
Coefficient of variation	26
Comparing slopes of linear regression	149
Competitive binding	292
Competitive binding, checklist	272
Competitive binding, introduction	269
Competitive binding, one site	174
Competitive binding, two sites	175
Competitive inhibitors	336
Compound symmetry	80
Confidence interval of a mean	8
Confidence interval of best fit values	213
Confidence interval of linear regression line	147
Confidence interval, definition of	27
Constants for nonlinear regression	203
Constrained linear regression, when to choose	142
Constraining variables	189
Constraining variables with MAX and MIN	190
Contingency table	119
Contingency table analyses, checklist	127
Contingency table, entering data	120
Contingency tables, choosing analyses	121
Convergence criteria for nonlinear regression	207
Correlation	133
Correlation coefficient, defined	135
Correlation vs. regression	133
Correlation, checklist	137

Counting error of radioactive decay	251
Covariance, analysis of.....	150
Cross-sectional study.....	119
Cubic spline curve	169
Curie	249
Curve fitting	<i>See</i> Nonlinear regression
Curves, generating	196
CV, Coefficient of variation	26

—D—

Dallal-Wilkinson method.....	29
Data sets, definition of.....	363
Default options for nonlinear regression.....	209
Depletion of radioligands	247
Derivative of a curve.....	358
Differential equations, can't enter.....	184
Dissociation binding curve	177, 295
Divide by baseline	353
Dose-response curve	173, 174
Dose-response curves, shape of	306
Dunn's post test, following Friedman test	87
Dunnett's test.....	68
Dunn's post test, following Kruskal-Wallis test	83

—E—

EC50	174, 307
Efficacy	314
Efficiency of detecting radioactivity	248
ELISA	343
Enzyme kinetics, checklist	331
Enzyme kinetics, compared with radioligand binding	332
Enzyme progress curve	328
Enzymes	327
Equation library, saving	189
Equation syntax	185
Equation, generating a curve from	196
Equilibrium binding, model of.....	161
Equilibrium dissociation constant	244
ESD method to detect outliers	19
Exponential association, one phase	178
Exponential association, two phases	178
Exponential decay curve, one phase	177
Exponential decay curve, two phases	178
Exponential decay model.....	159
Exponential growth.....	179
Extended ternary complex model.....	279
Extremely significant	12

—F—

F ratio, one-way ANOVA.....	70
F ratio, two-way ANOVA	93
F test to compare the fits of two equations	225
F test to compare two data sets	231
F test to compare two models	227
F test to compare variance, from unpaired t test.....	45
False minimum, following nonlinear regression	221
Fisher's test vs. Chi-square	121
Fisher's test, checklist.....	127
Forcing a regression line through the origin	142
Four-parameter logistic equation	174
Frequency distributions	24
Friedman test, checklist.....	87
Friedman test, how it works.....	86
Friedman test, posttests	87
Friedman test, results	87
Functions. available for entering user-defined equations	186

—G—

Gaussian distribution, origin of.....	7
Gaussian distribution, testing for	29
Gauss-Newton method	166
Generate a theoretical curve	196
Geometric mean.....	28
Grubbs' method to detect outliers.....	19

—H—

Half-life of radioactive decay.....	250
Half-life, equation for.....	160
Hazard ratio to compare survival curves	110
Hill slope.....	174, 275, 308
Histograms. Generating frequency distributions.	24
Homologous competitive binding curve	280
Hyperbola.....	171
Hypothesis testing, defined.....	11

—I—

IC50	174
IF-THEN relationships in equations.....	187
Implicit equations, can't enter	184
Independence, statistical use of the term	5
Independent samples, need for.....	4
Indexed format.....	40, 64

Inhibitors, competitive	336
Inhibitors, not competitive.....	339
Initial rate, enzyme kinetics.....	328
Initial values for nonlinear regression	201
Initial values, rules for.....	188
Integrate a curve	358
Interaction, in ANOVA	96
Interchange X and Y values.....	348
Interpolating from a standard curve	341
Intuitive Biostatistics, book	367

—K—

Kaplan-Meier method for survival data.....	109
Kinetic experiments, analyzing.....	296
Kinetics of competitive binding.....	303
Kinetics of radioligand binding, association	297
Kinetics of radioligand binding, dissociation.....	295
K_M	329
K_m , of enzyme reaction	329
Kolmogorov-Smirnov test	29
Kruskal-Wallis test, checklist	84
Kruskal-Wallis test, how it works	82
Kruskal-Wallis test, posttests.....	83
Kruskal-Wallis test, results.....	84

—L—

Law of mass action	244
Law of mass action, testing with kinetics	300
Levenberg-Marquardt method	167
Library for equations, saving	189
Ligand depletion, competitive binding	292
Ligand depletion, general	247
Ligand depletion, homologous competitive binding.....	286
Ligand depletion, saturation binding	264
Lilliefors method	29
Limitations of statistics	6
Linear descent method	166
Linear regression vs. correlation	133
Linear regression, checklist	150
Linear regression, comparing slopes and intercepts	149
Linear regression, constrained	142
Linear regression, graphing 95% CI.....	152
Linear regression, introduction	139
Linear regression, residuals	148
Linear regression, results.....	150
Linear regression, runs test.....	149
Linear regression, theory	139

Linear regression, two segments	194
Lineweaver-Burke plot.....	333
Lineweaver-Burke plots, avoiding	140
Logistic equation	173, 174
Logistic regression	169
Logrank test for trend.....	110
Logrank test to compare survival curves	109
Lowess curve.....	169

—M—

Mann-Whitney test, checklist	57
Mann-Whitney test, how it works	56
Mann-Whitney test, results of.....	57
Mantel-Haenszel test for comparing survival curves.....	109
Marquardt method.....	167
Mass action, law of	244
McNemar's test	121
Mean, geometric	28
Means, by rows.....	37
Median survival, definition.....	110
Method by example.....	365
Method file.....	366
Method of linear descent	166
Method options for nonlinear regression	204
Michaelis-Menten constant.....	172
Michaelis-Menten equation:	328
Missing values in ANOVA.....	94
Mixed ANOVA model.....	91
Model I (fixed effects) vs. Model II (random effects) ANOVA	94
Model I vs. Model II ANOVA.....	94
Multiple comparison post tests, choosing	68
Multiple comparisons.....	17
Multiple regression	168

—N—

Newman-Keuls vs. Tukey post test.....	69
Non-competitive inhibitors.....	339
Nonlinear regression, advantages over linear regression	140
Nonlinear regression, approach to interpreting	211
Nonlinear regression, comparing fits to two data sets	229
Nonlinear regression, comparing two equations	225
Nonlinear regression, constants	203
Nonlinear regression, convergence criteria.....	207
Nonlinear regression, default options.....	209
Nonlinear regression, fixing problems	223
Nonlinear regression, how it works.....	165
Nonlinear regression, initial values.....	201
Nonlinear regression, method options	204

Nonlinear regression, output options.....	208
Nonlinear regression, steps.....	199
Nonlinear regression, summary table.....	209
Nonparametric posttests.....	83
Nonparametric tests, choosing.....	42, 66
Nonspecific binding.....	246, 247
Nonspecific, subtracting or dividing.....	353
Normality test.....	29
Normalizing data.....	351
Normalizing transformations.....	40, 64
Null hypothesis, defined.....	9, 11
Numerical derivative of a curve.....	358

O

Odds ratio.....	122, 123, 124
On rates of radioligand binding.....	297
One sample t test, checklist.....	33
One site binding equation.....	171
One-sample t test, choosing.....	25
One-sample t test, results of.....	30
One-tail P value, defined.....	10
One-way ANOVA.....	<i>See</i> ANOVA
Operational model, derivation.....	315
Outliers.....	18
Output options for nonlinear regression.....	208

P

P value, one- vs. two-tailed defined.....	10
P values, common misinterpretation.....	9
P values, general.....	9
Paired t test, checklist.....	54
Paired t test, how it works.....	51
Paired t test, results.....	52
Paired t test, testing for adequate pairing.....	51
Paired tests, choosing.....	41
Pairing, testing for adequate pairing.....	51
Parallel regression lines, testing for.....	149
Peaks, area under.....	360
Pearson correlation.....	133
Percentiles.....	27
Poisson distribution.....	251
Polynomial equations.....	180
Polynomial regression.....	167, 168
Population, in statistics.....	4
Post test for linear trend, results.....	72
Post tests following ANOVA, results.....	72
Post tests, choosing.....	68

Post tests, following two-way ANOVA, choosing.....	93
Posttests following two-way ANOVA, interpreting.....	98
Power calculations, example.....	13
Power series.....	179
Power, defined.....	13
Prediction interval of the regression line.....	147
Prior probability.....	15
Progress curve, of enzyme.....	328
Proportional hazards regression.....	169
Prospective study.....	119

Q

Quartiles.....	27
----------------	----

R

R squared from correlation.....	135
R squared, from one-way ANOVA.....	70
R squared, linear regression.....	145
R squared, nonlinear regression.....	217
r, correlation coefficient, defined.....	135
Radioactive decay.....	250
Radioactivity, efficiency of detecting.....	248
Radioimmunoassay.....	343
Radioimmunoassay, using standard curves for.....	341
Radioligand binding, compared with enzyme kinetics.....	332
Radioligand binding, competition experiments.....	272
Radioligand binding, introduction.....	243
Radioligand binding, kinetic experiments.....	296
Radioligand binding, law of mass action.....	244
Radioligand binding, ligand depletion.....	247
Radioligand binding, nonspecific binding.....	246
Radioligand binding, saturation experiments.....	255
Radioligand binding, Scatchard plots.....	259
Random numbers, how Prism calculates.....	197
Randomized block design.....	66, 92
Rank sum test, choosing.....	25
Ratio t tests.....	56
Receptor binding experiments.....	243
Rectangular hyperbola.....	171
Regression vs. correlation.....	133
Regression, linear.....	<i>See</i> Linear Regression
Regression, logistic.....	169
Regression, multiple.....	168
Regression, nonlinear.....	<i>See</i> Nonlinear regression
Regression, polynomial.....	167
Regression, proportional hazards.....	169
Relative risk.....	124
Relative risk, equation for.....	122

Remove baseline.....	353
Repeated experiments, analyzing.....	365
Repeated measures ANOVA, checklist.....	81
Repeated measures ANOVA, results.....	79, 80
Repeated measures test, choosing.....	65
Repeated measures two-way ANOVA, choosing.....	91
Replicates in nonlinear regression.....	206
Resampling approach to statistics.....	6
Residuals following linear regression.....	148
Residuals following nonlinear regression.....	219
Results of analyses, viewing.....	364
Retrospective case-control study.....	119
RIA.....	343
Row means or totals analysis.....	37
Rules for entering user-defined equations.....	185
Rules for initial values in nonlinear regression.....	188
Runs test following linear regression.....	149
Runs test following nonlinear regression.....	220

Standard curves.....	341
Standard deviation, definition of.....	26
Standard error of best-fit values.....	212
Standard error of the mean, definition of.....	26
Statistical hypothesis testing.....	11
Statistical power.....	13
Statistical significance, defined.....	11
Statistics, limitations of.....	6
Steady state enzyme kinetics.....	328
Subtract baseline.....	353
Subtracting (or dividing by) baseline values.....	353
Sum of the squares, nonlinear regression.....	164
Sum of the squares, why minimize?.....	144
Sum-of-squares from nonlinear regression.....	216
Support, technical.....	367
Survival analyses, choosing.....	108
Survival analysis, how it works.....	109
Survival curves, comparing.....	109
Survival curves, entering data.....	106
Survival curves, interpreting.....	109
$s_{y,x}$	146, 217
Syntax of user-defined equations.....	185

S

Sampling from a population.....	4
Saturation (Scatchard) radioligand binding.....	255
Saturation binding curve.....	171
Saturation binding data, checklist.....	258, 331
Save method.....	366
Scatchard plots.....	259
Scatchard plots, avoiding.....	140
SD of dataset (column).....	26
SD of replicates, calculating.....	37
SD of rows.....	37
SD vs. SEM.....	28
SD, definition of.....	26
SE of best-fit values.....	212
Segmental linear regression.....	194
SEM of dataset (column).....	26
SEM of replicates, calculating.....	37
SEM of rows.....	37
SEM, definition of.....	26
Semilog graphs, straight lines on.....	154
Sigmoidal dose-response.....	173
Significance, defined.....	11
Simulate theoretical curve.....	196
Sine wave.....	180
Slope factor.....	174, 275, 308
Smooth a curve.....	358
Spearman correlation.....	134
Specific radioactivity.....	249
Sphericity.....	80
Spline curve.....	169

T

t ratio, from one-sample t test.....	30
t ratio, from unpaired t test.....	45
t test, one sample, checklist.....	33
t test, one sample, choosing.....	25
t test, one sample, results of.....	30
t test, paired, checklist.....	54
t test, paired, how it works.....	51
t test, paired, results.....	52
t test, ratio.....	56
t test, unpaired, how it works.....	45
t test, unpaired, results of.....	46
t test, Welch's.....	44
t tests, choosing.....	41
t tests, entering data for.....	39
t tests, paired or unpaired?.....	41
Technical support.....	367
Ternary complex model.....	279
Test for linear trend, choosing.....	69
Test for trend, chi-square.....	124
Theoretical curve, generating.....	196
Three-parameter logistic equation.....	173
Totals, by row.....	37
Transforming data before t tests.....	40
Transforming data before t tests and ANOVA.....	64
Transposing rows and columns.....	355

Tukey vs. Newman-Keuls post test.....	69
Two linear regression segments.....	194
Two site binding equation	172
Two-tail P value, defined	10
Two-way ANOVA.....	<i>See ANOVA</i>
Type I error, defined	12
Type II error, defined.....	13

—U—

Unpaired t test, how it works.....	45
Unpaired t test, results of.....	46
Unpaired tests, choosing.....	41
User-defined functions	350

—V—

Velocity of an enzyme reaction.....	328
Viewing results.....	364
V_{max}	329

—W—

Weighting in nonlinear regression	204
Welch's t test.....	44
Wilcoxon matched pairs test, checklist.....	61
Wilcoxon matched pairs test, results of	60
Wilcoxon rank sum test, choosing.....	25
Wilcoxon signed rank test, checklist.....	36
Wilcoxon signed rank test, how it works.....	34
Wilcoxon signed rank test, results of.....	35

—Y—

Yates' continuity correction, choosing.....	121
---	-----