A statistical companion

GraphPad StatMate

Version 1.0

User s Guide

GraphPad Software, Inc.

© 1995, GraphPad Software, Inc. All rights reserved.

Program design and programming: Dr. Harvey J. Motulsky
Manual and help screens: Dr. Harvey J. Motulsky and Paige Searle

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

All Rights Reserved. Manufactured in the United States of America. Except as permitted under the United States copyright law of 1976, no part of this publication may be reproduced or distributed in any form or by any means without the prior written permission of the publisher. Use of the software is subject to the restrictions contained in the accompanying software license agreement.

GraphPad Software, Inc.
10855 Sorrento Valley Road #203
San Diego, CA 92121 USA

PHONE:	619-457-3909
FAX:	(USA) 619-457-8141
EMAIL:	support@graphpad.com or sales@graphpad.com
WORLD WIDE WEB:	http://www.graphpad.com

Important

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

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

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

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

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

SOFTWARE LICENSE AGREEMENT

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

1. GraphPad Software Inc. (GraphPad) grants you a limited license to use the accompanying disks and manual (the Licensed Software). Title to the Licensed Software remains with GraphPad at all times. You may not sublicense, rent, lease, convey, or modify the software. You may transfer this license to another individual who agrees to these terms, provided that you keep no copies of the manual or disks.
2. You may use the Licensed Software on more than one computer, provided there is no chance it will be running simultaneously on more than one machine. You may make backup copies of the disks, provided they are labeled with the copyright notice. Unauthorized copying is an infringement of GraphPad's copyright.
3. You may make backup copies of the disks, provided they are labeled with the copyright notice. Unauthorized copying is an infringement of GraphPad's copyright.
4. We offer a 90 day money-back guarantee. If you find that the Licensed Software does not suit your needs, return the manual and disks and a copy of the invoice for a full refund.
5. GraphPad provides the Licensed Software as is , and does not warrant that the Licensed Software will meet customer requirements or be error-free. GraphPad shall not be liable for any consequential, incidental, special, punitive, exemplary or other damages, even if advised of the possibility of such damages. GraphPad expressly disclaims any and all warranties, expressed or implied. GraphPad's entire liability, and your exclusive remedy, shall be to replace the Licensed Software or refund the license fee.
6. The Licensed Software is designed for research purposes only, and should not be used for the diagnosis or treatment of patients.
7. This agreement shall be governed by the laws of the State of California, USA If you have any questions, please contact GraphPad Software: 10855 Sorrento Valley Road #203, San Diego, CA 92121 USA; FAX 619-457-8141. Phone: 619-457-3909.

Contents

OVERVIEW 7
Introducing GraphPad StatMate 7
Installing GraphPad StatMate 8
Starting and quitting StatMate 9
The StatMate Approach 9

TEST GUIDE 11
Using StatMate's test guide 11
Example of StatMate's test guide 12

DETERMINING SAMPLE SIZE 13
Introduction to determining sample size 13
Review of statistical hypothesis testing 14
Determining necessary sample size when comparing two means 15
Calculating sample size when comparing two proportions 19
Do you have to make the sample sizes equal? 21
Don't interpret sample size results too rigidly 22

RANDOMIZING SUBJECTS 25
Randomizing with StatMate 25
Example 26

P VALUES 27
Calculating a P value from t, z, r, F or chi-square 27
Example. P from an F statistic. 28
Calculating a P value from the runs test, McNemars test or the sign test 28
Meaning of a P value 31

CONFIDENCE INTERVALS 33
Meaning of a confidence interval 33
Calculating confidence intervals with StatMate 33
Example 35

POWER 37
Why calculate power? 37
Example of power calculations 38

BINOMIAL AND POISSON DISTRIBUTIONS 41
What are the binomial and Poisson distributions? 41
Using StatMate to calculate the binomial and Poisson distribution 42
Example of binomial calculations 42

BAYESIAN CALCULATIONS 45
Introduction to Bayesian logic 45
Bayesian analysis of statistically significant results 45
Bayesian analysis of clinical lab tests 49
Bayesian analysis of linkage data 52
The likelihood ratio 53
Why are Bayesian calculations controversial? 54

THE RESULTS LOG 57
Adding results to the log 57
Printing 58
Exporting the results log 59
Copying results to the clipboard 60

ALGORITHMS USED 63
Algorithm for sample size and power. Comparing two means. 63
Algorithm for sample size and power. Comparing two proportions. 64
Randomizing method 66
P value computations 66
Confidence intervals 66
Bayesian equations 66

GRAPHPAD SOFTWARE 67
Software license agreement 67
Technical support 67
Citing GraphPad StatMate 68
Uninstalling Statmate 68
Intuitive Biostatistics (the book) 68

INDEX 70

Overview

Introducing GraphPad StatMate

You probably already have one or more statistics programs (such as GraphPad InStat and GraphPad Prism) to help you analyze your data. StatMate is designed to complement these programs. StatMate helps you plan statistically sound experiments, and helps you interpret results from other programs and publications.

Planning an experiment

- StatMate helps you pick the right statistical test to use. Tell it the kind of data you are collecting, and the kind of comparison you want to make. StatMate suggests a test, and tells you whether it can be performed with other GraphPad programs.
- How many subjects (data points) do you need? Naturally, the answer is it depends . It depends on how large a difference you are looking for, how much your data vary, and on how willing you are to risk mistakenly finding a difference by chance or mistakenly missing a real difference. StatMate helps you see the tradeoffs, so you can pick an appropriate sample size.
- To avoid subtle biases, it is best to randomly assign subjects to treatments. StatMate will do this for you. Simply tell it how many treatment groups you need, and how many subjects you want in each group. StatMate assigns each subject to a group.

Interpreting an experiment

- Many publications (and some programs) just tell you that $P<0.05$ or $P>0.05$ but don t tell you the exact value of P. StatMate will calculate P (to four decimal places) from t, z, F, r, or chi-square. It will also calculate a P value from the results of the runs test, sign test, and McNemar s test.

- While many publications focus on P values, results are often easier to interpret when expressed as a confidence interval. StatMate calculates a confidence interval of a proportion, count, or mean. You specify whether you want 90%, 95% or 99% confidence.
- Just because a study reaches a conclusion that the results are not statistically significant doesn t mean that the treatment was ineffective. It is possible that the study missed a small effect due to small sample size and/or large scatter. StatMate calculates the power of a test to detect various hypothetical differences.
- When reviewing the results of an experiment (or a laboratory test) you need to interpret the statistical information in the context of what you knew before you did the experiment (based on theory and previous data). StatMate performs the Bayesian calculations required to do this.

Although this manual reviews many statistical principles, it is not a statistics text book. See page 68.

Installing GraphPad StatMate

GraphPad StatMate runs on any PC running Microsoft Windows version 3.1, or Windows 95. We recommend that you use a computer that has a fast 386 processor or better, at least 4 megabytes of RAM, and a mouse (or other pointing device).

If you purchased GraphPad Prism together with GraphPad StatMate, install Prism first.

To install GraphPad StatMate, follow these steps:

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

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

6. Follow the instructions on screen. The installation program copies program files to your hard drive, but does not make any changes to your system files (AUTOEXEC.BAT, CONFIG.SYS, WIN.INI, SYSTEM.INI).

It is not possible to install StatMate manually. You must use the installation program which decompresses and copies files from the original disk to your hard drive. The installation program also places the StatMate icon onto your Program Manager desktop.

Starting and quitting StatMate

To start StatMate

- Double click on its icon:

  or

- From GraphPad Prism (version 2), pull down the Windows menu and choose StatMate. This works only if StatMate is installed in the Prism2 directory.

To quit StatMate:

- Pull down the File menu and select Exit.

The StatMate Approach

When you start StatMate, the top of your screen will look like this. The tabs at the top of the Window show the names of StatMate s eight calculators.

To use a calculator:

1. Click on one of the tabs to choose a calculator and bring it to the front.
2. Fill out the form on the top half of the calculator, labeled Input.
3. Click the button labeled Calculate (sometimes renamed Advise or Assign). Alternatively, can click anywhere in the results panel at the bottom of the StatMate window.
4. Read the results on the bottom of the same screen, labeled Output.

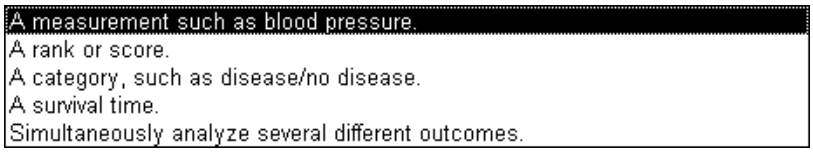
5. Press the button labeled Add to log. You may accumulate several sets of results in the log, and then print or export the entire log. See The results log on page 57.

Test guide

Using StatMate's test guide

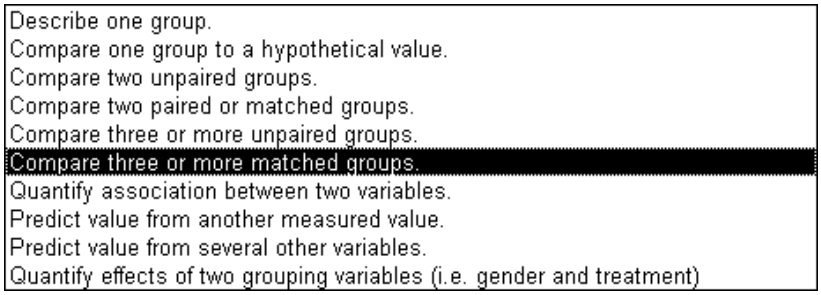
StatMate s test guide can help you find the right statistical test to analyze your data. Follow these steps:

1. Tell StatMate what kind of data you are collecting.



2. If your data are measurements, decide whether you are willing to assume that the values are sampled from a Gaussian distribution. Uncheck the box if you are not willing to make that assumption.

3. Choose the kind of comparison you want to make. Choose from this list.



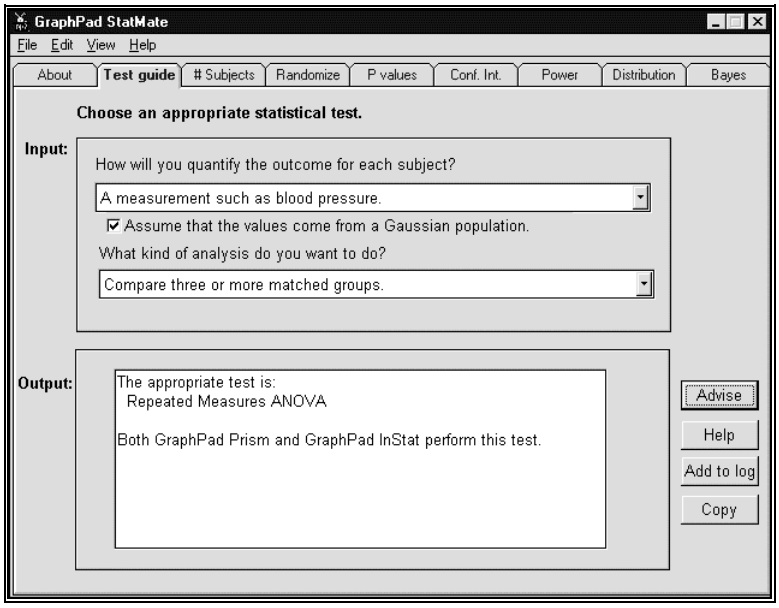
4. Click the Advise button to see what test StatMate recommends.

Important: Before using the test that StatMate advised, review a statistics book to make sure that the recommendation is sound. StatMate can usually point you in the right direction for choosing a test, but you should be sure you understand the test and agree that it is appropriate for your data.

Example.

Example of StatMate's test guide

What test should you use to compare a measurement in three matched groups, assuming Gaussian populations?



StatMate suggests repeated measures analysis of variance. Before performing this test, you should review a textbook or manual to make sure you agree with StatMate s recommendation.

Determining sample size

Introduction to determining sample size

How many subjects (data points) do you need? Naturally, the answer is it depends . It depends on the answers to these questions:

- How large a difference are you are looking for?
- How variable are your data?
- How willing are you to risk mistakenly finding a difference by chance?
- How willing are you to mistakenly miss a real difference?
- How much time and money do you have to spend on this experiment?

StatMate helps you see the tradeoffs, so you can pick an appropriate sample size.

To understand how StatMate determines sample size, you first need to review the principles of statistical hypothesis testing and the meaning of statistical power.

See also:

Review of statistical hypothesis testing

Determining necessary sample size when comparing two means

Calculating sample size when comparing two proportions

Do you have to make the sample sizes equal?

Don t interpret sample size results too rigidly

How StatMate calculates sample size for comparing means

How StatMate calculates sample size for comparing proportions

Review of statistical hypothesis testing

α and hypothesis testing

You do an experiment because you want to know whether the treatment affects the outcome. You need statistical analyses because the answer can be ambiguous.

Most statistical programs present the results as a P value that answers this question: If the treatment really does not affect the outcome, what is the chance that random variability would result in a difference as large or larger than you observed in your experiment? The P value is a fraction between zero to one. If the P value is small, you are inclined to believe that the difference you observed is due to the treatment.

In many cases, you want to make a decision from the data. You want to conclude either that the difference is statistically significant, or that it is not statistically significant. This is done in a simple manner. Before running the experiment you set a threshold P value, called α, usually set equal to 0.05. After running the experiment and calculating the P value, use this logic:

- If the P value is less than or equal to α, then you conclude that the treatment had a statistically significant affect, and you reject the null hypothesis that the treatment was ineffective.
- If the P value is greater than α, then the data do not persuade you that the treatment has any effect. You conclude that the treatment did not have a statistically significant effect.

Type I vs. Type II error

The conclusion that a result is statistically significant or not statistically significant sounds stronger than it is. You can make two kinds of mistakes.

- A Type I error occurs when you conclude that there is a statistically significant effect, but the difference was actually due to random variability. The treatment had no effect, but the random variation between the groups made you conclude that the difference existed.
- A Type II error occurs when you conclude that there is no statistically significant effect, but the treatment was really effective. The treatment had an effect, but the random variation between the groups obscured the difference.

β and power

Even if the treatment affects the outcome, you might not obtain a statistically significant difference in your experiment. Just by chance, your data may yield a P value greater than α. In other words, you may make a Type II error, also called a β error.

Assume that the two means truly differ by a particular amount, and that you perform many experiments with the same sample size. For each experiment, there are two possibilities:

- The P value is less than α, so you call the results statistically significant. The probability of this happening is called power.
- You obtain a P value greater than α, so call the difference not statistically significant. You have missed finding the true difference, and thus made a Type II error. The probability of this happening is called β.

Since the two possibilities are mutually exclusive, power and β add up to 100%.

Determining necessary sample size when comparing two means

Choosing an appropriate sample size depends on balancing five factors:

- The anticipated SD within the two groups. If the SD is large, the random variation within groups will tend to obscure differences between groups. To overcome this problem, you ll need many subjects. If the SD is small, you can get by with fewer subjects.
- α, the threshold P value for calling a difference significant. If you choose a small value for α, you need a large sample size. If you choose a higher value for α, and thus accept a higher chance of a Type I error, you can get by with fewer subjects. Most often, α is set to 0.05, but this is arbitrary.
- Δ_H. The hypothetical difference you are looking for. It takes many subjects to find a small difference. You can get by with fewer subjects when you are looking only for large differences.
- Power. If you want to have high power (high chance of finding a real difference), you ll need many subjects. If you are willing to have less power (and accept the increased risk of a Type II error) then you need fewer subjects.

- Expense. Thinking purely of the statistical aspects of the problem, you will always want huge samples. But you need to balance the statistical considerations against the cost and effort required to run huge experiments.

Most programs require you to enter values for SD, α, Δ_H, and power. The program then calculates N. There are two problems with this approach. One problem is that many scientists find it difficult to set values for Δ_H and power. The other problem is that the value of N is likely to be very high. You then will want to go back and adjust the values to get a lower value of N.

StatMate uses a different approach. Rather than calculating a single sample size, StatMate shows you a table of the tradeoffs between power and Δ_H for various values of N.

Using StatMate to calculate sample size for comparing two means

Recall that sample size calculations are based on knowing the anticipated SD, and setting values for α, power and Δ_H.

There is no way to get around the need to estimate an anticipated SD. Base the estimate on previous similar experiments. If you have no idea what SD to expect, you should run a pilot experiment to find out.

Traditionally, α is set to 0.05, and most scientists are willing to stick with that custom. Ideally, you should set a value for α based on the consequence of making a Type I error (concluding that a difference is statistically significant when in fact the treatment is ineffective and the difference was due to random variation). If the consequences of making a Type I error are serious, set α to a smaller value. If the consequences of a Type I error are minor, set α to a higher value.

- Example 1. You are performing a screening test for new drugs. The consequence of a Type I error is minor you ll waste an afternoon repeating the experiment. So it makes sense to set α to a high value, perhaps 0.10. Setting α to a high value lets you use a smaller sample size.
- Example 2. You are performing a clinical trial of a new drug for a condition for which standard treatments work pretty well. The consequences of a Type I error would be serious recommending a worthless drug to patients who otherwise would get an effective standard therapy. So it makes sense to set α to a low value, perhaps 0.01.

StatMate does not require you to enter values for power or Δ_H. After you enter values for SD and α, StatMate calculates a table of results. Each row in the

table represents a different sample size. Read across the table to see the relationship between power and Δ_H for that sample size. By making all the trade-offs explicit, StatMate helps you choose an appropriate sample size.

Follow these steps:

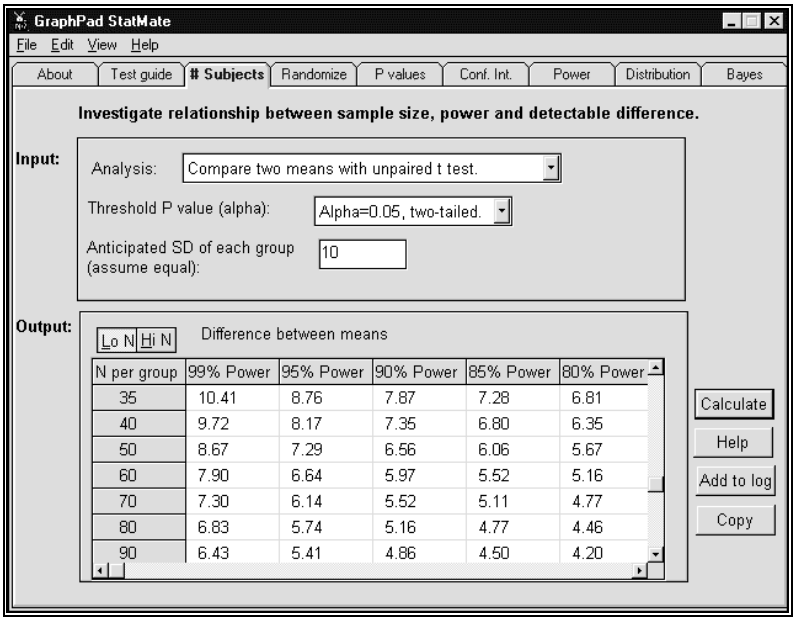
1. On top of the dialog, state that you want to compare two means with a t test. If you want to compare two proportions with a chi-square test, see Calculating sample size when comparing two proportions on page 19. StatMate cannot calculate sample sizes for other experimental designs.
2. Choose a value for α. The traditional value is $\alpha = 0.05$, two-tailed.
3. Enter an estimated value for the SD. Enter the estimated value in the units you will use for the outcome variable. If you are comparing blood pressures measured in mm Hg, enter the estimated SD as mm Hg. If you are comparing enzyme activities measured in pmol cAMP produced per milligram membrane protein per minute, enter the estimated SD in those units.
4. Press Calculate.
5. View the table. For various sample sizes and powers, StatMate shows you the difference between means that you can detect.
6. StatMate initially shows results for sample size from 3 to 300. If you plan to use larger sample sizes (N = 300 to 10,000), press the button labeled Hi N. Press Lo N to return to the initial range.

7. To learn how to interpret the results, see the example below.

Example of sample size calculations for comparing two means

You plan to test whether a treatment alters blood pressure. From previous experiments (or published data) you anticipate that the SD of blood pressures will be about 10 mmHg in the population you study. You want to figure out what sample size you need to find a statistically significant difference with $P < 0.05$ (two-tailed).

Choose that you are comparing two means and that $\alpha = 0.05$. Enter the anticipated SD = 10. Then press Calculate. StatMate presents the results in a table.



To decide on a sample size, you need to think about the tradeoff between sample size, power, and the detectable difference between means. You also need to think about the expense, effort and time of large sample sizes. There are two approaches you can use.

Approach 1. Start with a proposed sample size

Pick a sample size that seems reasonable in terms of cost and effort in this example lets start with N = 50 in each group. Look across the row at N = 50, and you ll see the differences that can be detected with various powers. With 50 subjects in each group, you ll have 99% power to detect a mean difference of 8.67, 95% power to detect a difference of 7.29, but only 80% power to detect a difference of 5.67. Those differences are expressed in the same units as you used to enter the SD, mmHg in this example. This means that if you performed many studies with N = 50 in each group, and if the real difference was 7.29 mmHg, you d find a statistically significant difference in 95% of those hypothetical studies but would find no significant difference in the other 5%. More simply, if the real difference were 7.29 mmHg, there is a 95% chance that you will obtain a statistically significant result in a study with 50 subjects in each group.

At this point, you need to think about the difference between means that you would find scientifically interesting, and balance that against the expense of increasing the number of subjects. You might conclude that N = 50 is overkill, and that you are willing to accept lower power so that you can use fewer

subjects. Or you might conclude that these powers are too low. Perhaps you want a high power to detect a difference as small as 4 mmHg. In this case, you d need to use a higher sample size. Pick a higher sample size, perhaps N = 100, and look across that row in the table to see how powerful such a study would be.

Approach 2. Start with the smallest difference you want to detect.

Pick a difference between means that you think would be scientifically important or interesting. In this example, you might state that a mean difference of 6 mmHg would be scientifically or clinically important. Look through the table to find values near 6 mmHg.

In the figure above, you can see that a sample size of 70 subjects in each group gives you 95% power to detect a difference of 6 mmHg, and that a sample size of about 45 in each group gives you 80% power to detect that difference. You need to think about the tradeoff between power and sample size to pick a sample size that makes sense in the context of your experiment.

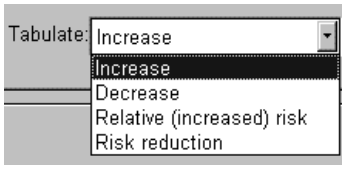
Calculating sample size when comparing two proportions

Calculating sample size to compare two proportions follows the same principles as used when comparing two means (see Introduction to determining sample size on page 13) with these differences:

- You don t need to estimate the standard deviation, as this would be meaningless for outcomes quantified with proportions.
- You need to estimate the proportion you will observe in the control group and whether you expect an increase or a decrease in the treated group.

To calculate sample size:

1. On top of the dialog, state that you want to compare two proportions with a chi-square test.
2. Choose a value for α. The traditional value is $\alpha = 0.05$, two-tailed.
3. Enter an estimated value for the proportion success in the controls.
4. Select whether you expect to see an increase or a decrease, and whether you want to see the results presented as a ratio of proportions or a difference.



5. Press Calculate.
6. View the table. For various sample sizes and powers, StatMate shows you the difference (or ratio) between proportions that you can detect.
7. StatMate initially shows results for sample size from 3 to 300. If you plan to use larger sample sizes (N = 300 to 10,000), press the button labeled Hi N. Press Lo N to return to the initial range.

If you ask StatMate to tabulate an increase or decrease, the table shows the difference between the control proportion (p_c) and the experimental proportion (p_e) for each combination of sample size and power. If you ask to see the relative risk, it tabulates p_e/p_c. If you ask to see the risk reduction, it tabulates $(p_c-p_e)/p_c$.

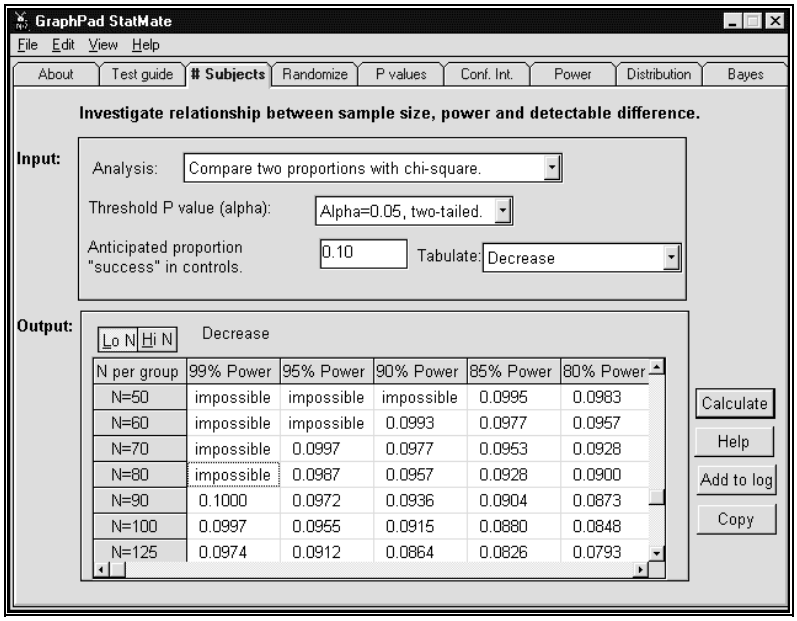
You may be surprised that you have to tell StatMate whether you expect to see an increase or a decrease. This is necessary because the power to detect a difference between two proportions depends not only on the difference (or ratio) of the two proportions, but also on their actual values. The power to find a difference between 0.2 and 0.4 is not the same as the power to find a difference between 0.4 and 0.6 (same difference) or between 0.4 and 0.8 (same ratio). So it is not enough to tell StatMate that you expect a control proportion of 0.4 you also need to tell it whether you expect to see an increase or decrease.

The issue is not whether the treatment increases or decreases the proportion the issue is knowing the proposed values for the two proportions. The power to find a doubling from 0.2 to 0.4 is exactly the same as the power to find a decrease from 0.4 to 0.2. In one case you enter the control proportion as 0.2 and tabulate an increase. In the other case you enter the control proportion as 0.4 and tabulate a decrease.

Example of calculating sample size for comparing two proportions

You know that the incidence of a complication of a disease is 10% in the population of patients you study. You are giving a treatment that you hope

will reduce the complication rate, and want to tabulate the results as the difference in complication rates.



Read the results table using either approach detailed in Example of sample size calculations for comparing two means on page 17.

- Start with a proposed sample size and read across to see the tradeoff between power and difference. If the powers seem too low, try a higher sample size. If the powers seem too high, try a lower sample size.
- Look for a difference between proportions that you think would be scientifically or clinically important. Read up to see the power and over to see the sample size.

Do you have to make the sample sizes equal?

StatMate calculations are based on the assumption that the two groups have the same number of subjects. This gives you the most statistical power for any given total number of subjects. If the two treatments require equal expense, effort and risk, you should plan to have the same number of subjects in each group.

In some situations it makes sense to have different numbers of subjects in the two groups, even though this arrangement has less statistical power. For example, you might want to do this in initial comparisons of a very risky treatment with placebo. It would also make sense to use unequal sample sizes when comparing diseased patients with controls, when the disease is very rare (making it hard to have a large sample size).

Here are some approximate guidelines:

- You can reduce the number of subjects in one group by 25% if you double the number in the other group.
- You can reduce the number of subjects in one group by half if you quadruple the number in the other group.
- No matter how many subjects you add to the larger group, you can t reduce the number in the smaller group to less than half the number needed for equal sample sizes.

If you plan to have different sample sizes in each group, you can still use StatMate with a bit of extra effort. For each proposed sample sizes N_1 and N_2, calculate the harmonic mean:

$$N = \text{Harmonic Mean} = \frac{2 \cdot N_1 \cdot N_2}{N_1 + N_2}$$

Find the value of N that corresponds to the harmonic mean and read across the results table to see the tradeoff between detectable difference and power.

Don’t interpret sample size results too rigidly

The table of sample size and power can be useful in designing a study, but don t interpret the values too rigorously. Consider these problems:

- The value of SD (or the control proportion) is only an estimate, and your estimate might not be accurate.
- The values of α, power, and Δ_H are arbitrary.
- The calculations are based on some simplifications, so are not completely accurate. Different texts give different equations that give slightly different results. The methods used by StatMate are documented in Algorithm for sample size and power. Comparing two means. on page 63 and Algorithm for sample size and power. Comparing two proportions. on page 64. The calculations are the least accurate when sample sizes are small (less than 10).

- The calculations tell you how many subjects you need at the end of the study. Because you may lose some values along the way, you need to start with more subjects than indicated.

These uncertainties don t affect the usefulness of the calculations as long as you use StatMate to estimate an approximate number of subjects.

Randomizing subjects

Randomizing with StatMate

Ideally, when you compare two groups in an experiment, the two groups are identical in all ways but the one factor or treatment you are investigating. In practice, it is hard to reach this ideal.

Since you can rarely be sure that the two groups are identical in all ways, you should at least ensure that there is no systematic bias, where certain kinds of subjects tend to be assigned to certain groups. The best way to do this is to randomly assign subjects to treatment groups.

To randomly assign subjects:

1. Click on the Randomize tab.
2. Choose how many treatment groups you wish to use.
3. Choose how many subjects you want in each group.
4. Press the Assign button. StatMate assigns each subject to a group.

How StatMate randomizes

There are two approaches to randomizing. To illustrate the difference, consider the simple case of randomly assigning subjects to two groups without the aid of a computer. You can flip a coin, or you can pull names out of a hat.

StatMate does not follow the coin flipping approach. The problem with this approach is that the groups may not end up with the same number of subjects. In fact, all the subjects may be assigned to the same group if the coin happens to land on heads many times in a row, or on tails many times in a row.

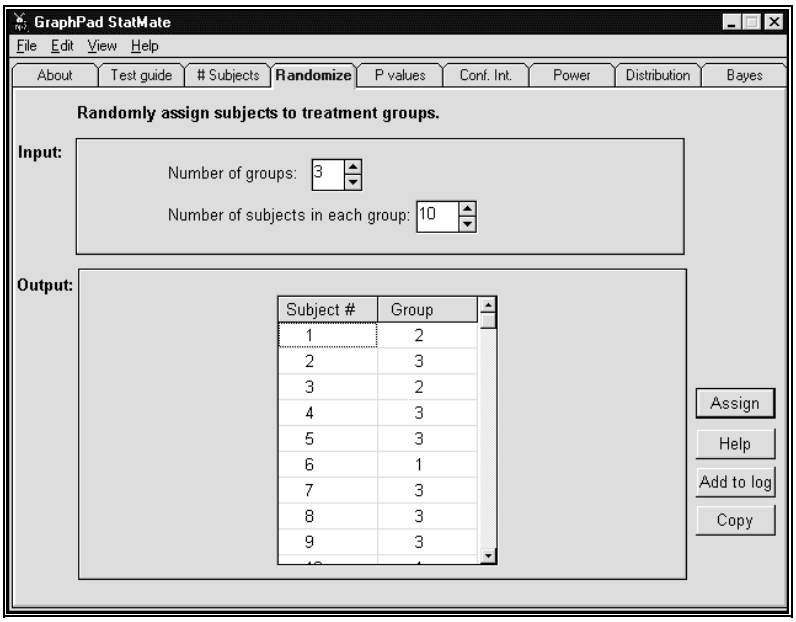
StatMate uses an approach similar to drawing names out of a hat. Make a slip of paper for each subject, and write down their number (or name). Put the slips into a hat, and mix. Pull out the first slip of paper, and assign that subject to group 1. Pull out the next slip of paper and assign that subject to

group 2. Pull out the third slip, and assign to group 1. Continue until all subjects are assigned.

The only way to get truly random numbers is to randomize based on a physical process such as flipping a coin, spinning a roulette wheel or measuring radioactive decay. StatMate does not use this type of method, so its random numbers are not strictly random. The numbers are generated using a mathematical algorithm, and are called pseudo-random numbers. The choice of the first number is influenced by the time of day that you run the program. All the remaining numbers are generated in sequence using a series of calculations that are completely reproducible, and thus not truly random. But the sequence is only predictable to someone who knows the algorithm. To anyone else, the sequence appears to be completely random. The numbers are certainly random enough to be used to assign subjects to treatment groups. This is not unique to StatMate, as all computer programs generate pseudo-random numbers.

Example

You want to randomly divide 30 subjects into three groups. This screen shows the input and the results. Since the assignment is random, subjects will be assigned to different groups when you try this example.



P values

Calculating a P value from t, z, r, F or chi-square

StatMate doesn t analyze raw data to generate P values. StatMate only calculates a P value from statistical parameters: z, r, F, t, and chi-square. Calculating a P value with StatMate is useful when you have results from another program (or publication) that only says $P<0.05$ (or $P>0.05$) but doesn t give the exact P value.

Follow these steps:

1. Choose whether you want to calculate P from t, z, F, r, chi-square, or the number of runs (runs test).
2. Enter the value of the statistical parameter.
3. Enter the degrees of freedom (not needed for z). If you are calculating the P value from F, enter the df for the numerator and denominator. Make sure you don t mix up the two, as the reversing the two values will affect the result.
4. Press the Calculate button.

Usually the number of degrees of freedom will be reported by the other program or publication. If not, follow these rules:

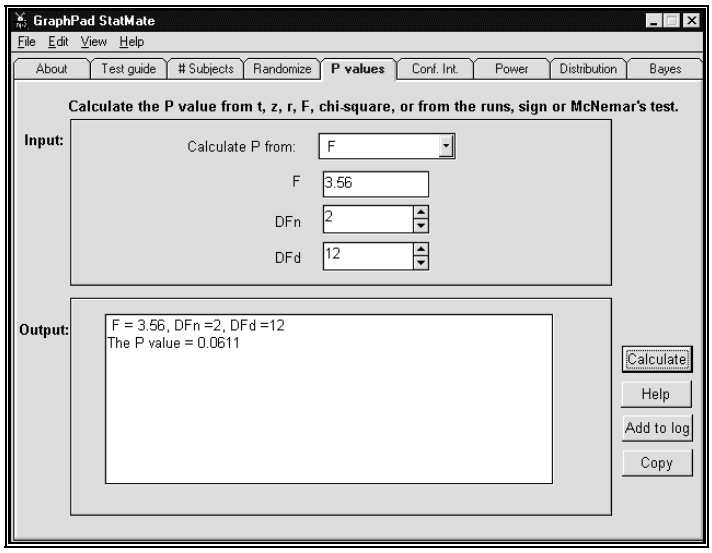
Test	Degrees of freedom
Unpaired t test	Total number of subjects minus two
Paired t test	Number of pairs minus one
Chi-square of contingency table	(R-1)*(C-1), where R is number of rows and C is number of columns
Chi-square comparison of observed and expected counts	Number of categories minus 1
Linear regression	Number of XY pairs minus 2
Nonlinear regression	Number of XY pairs minus the number

	of variables fit by the regression
One-way ANOVA	Dfn = Number of groups minus 1
	Dfd = Number of subjects minus number of groups minus 1

Calculating a P value from the runs test, McNemars test or the sign test

Example. P from an F statistic.

You read in a paper that F = 3.56 with 2 df in the numerator and 12 df in the denominator. The paper says that the P value was greater than 0.05. You want to know how much greater.



Calculating a P value from the runs test, McNemars test or the sign test

The runs test

The runs test is most often used to test whether a linear or nonlinear regression model adequately fits the data.

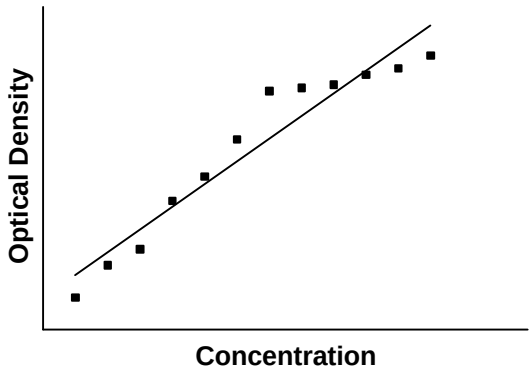
A run is a series of consecutive points that are all above the line or curve or all below the line or curve. In other words, a run is a series of consecutive points whose residuals (distance from the line or curve) are all positive or are all negative.

If the data points are randomly distributed above and below the regression line or curve, it is possible to calculate the expected number of runs. If the number of runs are fewer than expected, it may mean that the regression model is wrong. The P value from the runs test answers this question: If the data really follow the linear or nonlinear equation used to create the line or curve, what is the chance of obtaining as few (or fewer) runs as observed in this experiment? If the P value is small, you d be inclined to conclude that the data really don t follow the model.

The P values are always one-tailed, 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.

To calculate the runs test with StatMate, enter the number of points above the line or curve and the number of points below the line or curve. If a point is exactly on the line or curve, don t count it. Also enter the number of runs.

Here is an example of data analyzed by linear regression:



There are 12 data points. Six are below the line and six above. There are only three runs (the first run is the first three points below the line, the next run consists of six points above the line, and the last run are the three points on the right below the line). It looks like these data really follow a curve rather than a straight line. One way to test this is with the runs test.

Select the runs test, and enter 3 runs, 6 points above the line and 6 below. Press Calculate and StatMate tells you that the P value is 0.0130.

If the data really followed a linear model, there is only a 1.3% chance of observing so few runs. Therefore it is reasonable to conclude that the data don t follow a linear model, but follow some sort of nonlinear (curved) model instead.

The sign test

The sign test is used to analyze data where there are two alternative outcomes that, in theory, are equally likely. The sign test computes a P value that answers this question: If the two outcomes are truly equally likely, what is the chance of obtaining such a large discrepancy as observed in the current experiment.

Example. Someone has 12 grandchildren, of which 1 is a boy and 11 are girls. What is the chance of this happening?

The sign test starts with the assumption that both outcomes (boys and girls) are equally likely. This is not exactly true, but is nearly so. Select the Sign test, and enter 11 positive signs and 1 negative sign . (Or enter 1 positive and 11 negatives you ll get the same answer.)

The two-tail P value (0.0063) is the probability of seeing such a large deviation from 50:50 by chance. This includes the possibility of 1 boy and 11 girls and the more extreme possibility of 0 boys and 12 girls. Since that is a two-sided P value, it also includes the opposite possibilities (1 girl and 11 boys, and 0 girls and 12 boys).

It doesn t matter which of the outcomes you label positive and which negative . You ll get identical results if you swap the two values.

McNemar's test

McNemar s test is used to analyze paired case-control (retrospective) studies. You want to investigate the relationship between a possible risk factor and a disease. For each case that you study, you find an age-matched control. You may also match for other variables that might influence the results, such as ethnic group or sex. Then you look back to see whether the case and control were exposed to the risk factor. Because each case is matched to a particular control, you don t want to simply compare the numbers of cases and controls exposed to the risk factor. You want to do an analysis that takes into account the pairing.

For each pair of case and control, there are four possibilities:

1. Neither case nor control were exposed.
2. Both case and control were exposed.
3. The case was exposed, but not the control.
4. The control was exposed, but not the case.

The pairs that fall into categories 1 or 2 give you no information about the association between the risk factor and the disease. Only the pairs that fall into categories 3 and 4 give information about a possible association.

If there is no association between the risk factor and the disease, then you d expect to see about the same number of pairs in category 3 and category 4. If the risk factor is associated with the disease, then you d expect more pairs in category 3 than in category 4. If the risk factor protects against the disease, then you d expect more pairs in category 4 than in category 3.

McNemar s test calculates a P value that answers this question: if there really is no association between risk factor and disease, what is the chance of randomly selecting subjects and observing such a large discrepancy (or larger) between the number of pairs in categories 3 and 4.

StatMate calculates this P value using two methods and reports both. The first method is to use the binomial distribution. This method gives an exact P value. The other method is to use McNemar s method. This method calculates a value for chi-square and uses that to determine P. This method is an approximation. With large sample sizes it is a good approximation, with smaller sample sizes it is less accurate. We include it in StatMate only because other programs calculate it, and you may be used to seeing it. We recommend that you report the value from the binomial calculations rather than from McNemar's test.

McNemar s test can also be used to analyze experimental studies. You match two subjects for age, sex and other relevant variables. One subjects gets one treatment, and the other subject gets the other treatment. The outcome is scored as success or failure . The results for each pair fall into four categories.

1. Both subjects outcome is success.
2. Both subjects outcome is failure.
3. The subject who got treatment A had success but the subject who got treatment B had failure.
4. The subject who got treatment A had failure but the subject who got treatment B had success.

Enter the number of pairs in the last two categories into StatMate, and it will calculate McNemar s test.

Meaning of a P value

Many statistical tests report their result as a P value. Here is what a P value means.

1. Tentatively assume that both populations have the same mean overall, and that the difference you observed is entirely due to random variation. This is the null hypothesis.
2. Given the size of the experiment and the amount of scatter you observed, calculate (in theory) the results of all possible experiments under the null hypothesis.

Calculate the fraction of those experiments that would result in a difference as large, or larger, than you observed. That is the P value.

Confidence intervals

Meaning of a confidence interval

The whole point of statistical calculations is to extrapolate from the data you have collected to make general conclusions. Statisticians say that you extrapolate from your sample to make inferences about the overall population.

One way to make an inference about the overall population is to calculate a confidence interval. Let s review the meaning of the confidence interval of a mean. Calculating a mean is an exact calculation. But you know that the mean you calculate is unlikely to be the true mean overall. If you repeat the experiment, the mean will probably be different. To quantify this uncertainty, you want to combine the mean you calculate, with the SD and sample size to calculate a range of values called the 95% confidence interval.

The 95% CI is centered on the mean you calculated, and you can be 95% sure that the interval contains the true overall population mean. This means that if you calculate many 95% confidence intervals from many experiments, you expect that 95% of the time your interval will include the true population mean and that 5% of the time your interval will not include the true population mean. When analyzing any particular experiment, you won t know whether the interval you calculate includes the true population mean, but you can say that you are 95% sure that the interval contains the true mean.

You ll often see means or proportions published without confidence intervals. Use StatMate to calculate the confidence interval so you can get a feel for how strong the data really are.

Calculating confidence intervals with StatMate

Follow these steps:

1. Indicate whether you wish to calculate the confidence interval of a mean, count or proportion.
2. To calculate the CI of a mean, you need to also enter either the SD or SEM.

To calculate the CI of a count, you need only enter the counted value, i.e. number of radioactive counts in a certain time interval or number of cells observed in a certain volume. Make sure you enter the number of events of objects you actually observed don t normalize to another scale.

To calculate the CI of a proportion, enter the numerator and denominator.

3. Choose the degree of confidence. 95% confidence intervals are most common.
4. Press the Calculate button.

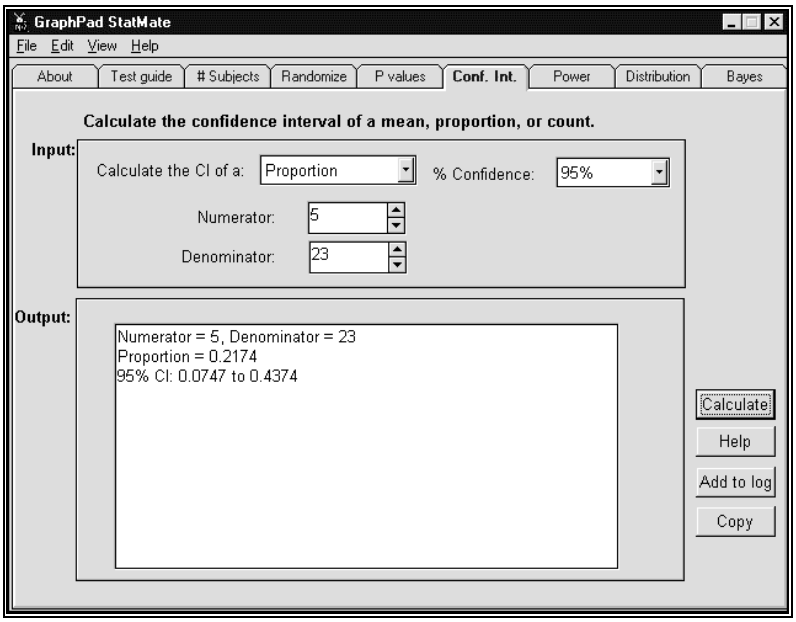
Keep these points in mind:

- Confidence intervals are calculated from a single experiment, and extrapolate the knowledge gained from that one experiment to the more general situation.
- The calculated confidence interval is often optimistic. The interpretation of the CI is based on the assumption that your data are representative of the larger population. In many cases, this isn t strictly true. If the experiment were to be repeated in many other labs, the variation would likely be larger than the variation observed in your lab.
- If you want to be more confident that a range of values includes the true population value, you need to make the range wider. So 99% confidence intervals are wider than 95% confidence intervals.

Review the meaning of a confidence interval.

Example

Out of 23 animals given a treatment, 5 died. What is the 95% CI for the proportion who die?



The 95% CI ranges from 7.5% to 43.7%. Many people are surprised to see how wide the intervals are.

Power

Why calculate power?

Just because a study reaches a conclusion of no statistically significant difference doesn t mean that the treatment was ineffective. It is possible that the study missed a small effect due to small sample size and/or large scatter.

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, as quantified by the 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 have a P value less than α and thus are deemed statistically significant.

In other words, power is the chance of finding the difference you are looking for.

Calculating power with StatMate

Follow these steps:

1. Choose whether you have compared two means with a t test or two proportions with a chi-square test. StatMate cannot calculate power for other experimental designs.
2. Enter the sample size of each group.
3. If you are calculating the power of a t test, enter the SD of each group.

   or

If you are calculating the power of a chi-square test, enter both proportions. Also choose whether you want to see the results displayed as a difference between proportions or a ratio of proportions.

4. Press the Calculate button.

StatMate displays a table. For various hypothetical differences between population means (or proportions), StatMate shows the power of your study to detect a significant difference.

Look in the table to find the smallest difference that you think would have been scientifically interesting or important. If the power of the study to find that difference is high, your negative results are solid. If the power of the study to find that difference was low, you should conclude that the study needs to be repeated with more subjects.

Example of power calculations.

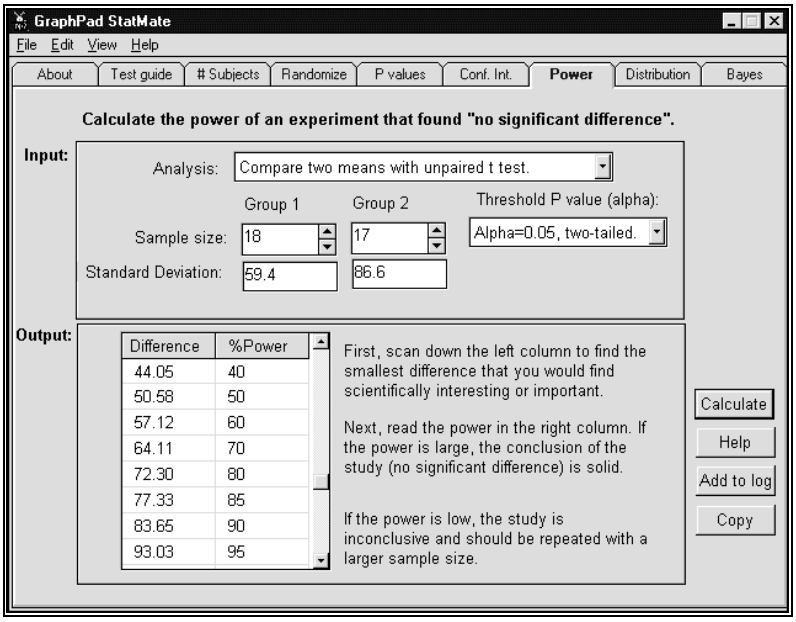
Example of power calculations

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 with 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 hypertensives do not have altered number of α_2 receptors on their platelets.

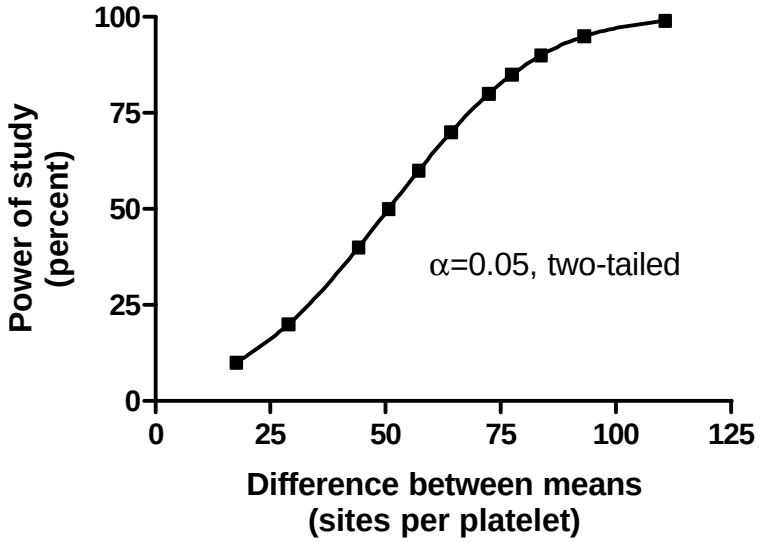
What was the power of this study to find a difference if there was one?



If the true difference between means was 50.58, then this study had only 50% power to find a statistically significant difference. If hypertensives really averaged 51 more receptors per cell, you d find a statistically significant difference in 50% of studies of this size. This is about a 20% change, 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.

The relationship between hypothetical differences and power is easiest to view as a graph.



(To make this graph, copy the results table from StatMate to GraphPad Prism, and fit the data with a cubic spline curve).

All studies have low power to find small differences and high power to find large differences. 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.

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.

Binomial and Poisson distributions

What are the binomial and Poisson distributions?

The binomial distribution predicts the distribution of results when each trial or subject can have one of two alternative outcomes traditionally labeled success and failure . You know the overall probability of success for each trial or subject, and want to calculate the probability of observing any particular number of successes and failures.

The binomial distribution answers questions like these:

- When two heterozygotes (carriers) for a disease have a child, there is a 25% chance that the child will have the disease. If they have three children, what is the chance of having zero, one, two or three affected children?
- What is the chance that five children will all be boys?

The Poisson distribution predicts the distribution of results when you are counting the number of events per time or the number of objects per volume. You know the average number of events per time or objects per volume, and want to predict the probability of observing a certain number of events or objects.

The Poisson distribution answers questions like these:

- If the average number of radioactive disintegrations is 100 per minute, what is the chance of observing 110 counts in a particular minute?
- If the average number of cells in a particular volume counted is 25, what is the chance of only counting 20 cells in that volume?

Using StatMate to calculate the binomial and Poisson distribution

Follow these steps:

1. Indicate whether you wish to use the Binomial or Poisson distribution.
2. For the binomial distribution, enter the proportion success in the overall population, and the number of subjects per experiment.

   For the Poisson distribution, enter the average number of counts per unit time or objects per unit volume in the entire population.
3. Press Calculate and view the table.

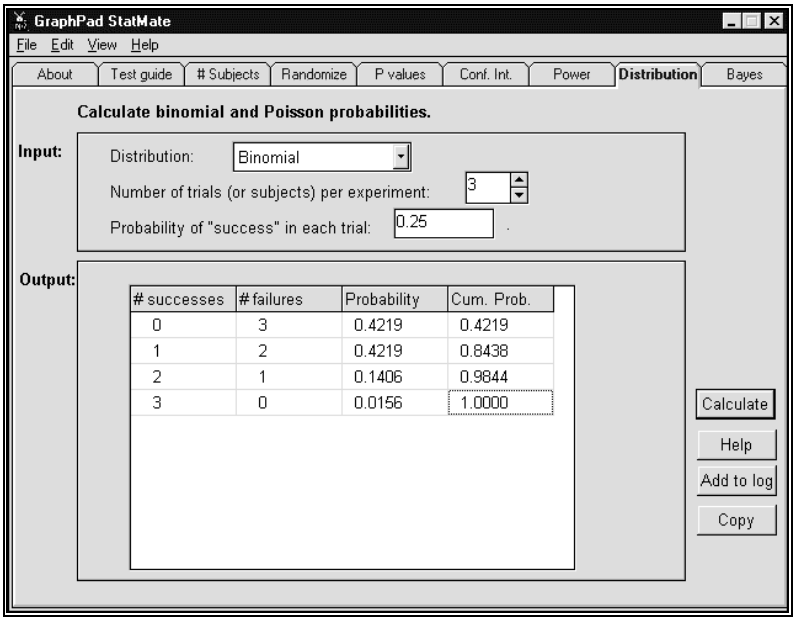
Review the uses of the binomial and Poisson distributions.

Example of binomial calculations.

Example of binomial calculations

When two heterozygotes (carriers) for a disease have a child, there is a 25% chance that the child will have the disease. If they have three children, what is the chance of having zero, one, two or three affected children?

Choose the binomial distribution, and enter 3 trials (subjects) per experiment with a 0.25 probability of success in each. Press Calculate. StatMate tells you that the chance of zero or one affected children is about 42%; the chance of having 2 affected children is about 14% and the chance of having all three children affected is 1.6%.


GraphPad StatMate
File Edit View Help
About
Test guide
Subjects
Randomize
P values
Conf. Int.
Power
Distribution
Bayes
Calculate binomial and Poisson probabilities.
Input:
Distribution:
Binomial
Number of trials (or subjects) per experiment:
3
Probability of "success" in each trial:
0.25
Output:
successes
failures
Probability
Cum. Prob.
0
3
0.4219
0.4219
1
2
0.4219
0.8438
2
1
0.1406
0.9844
3
0
0.0156
1.0000
Calculate
Help
Add to log
Copy

Bayesian calculations

Introduction to Bayesian logic

When interpreting the results of an experiment (or a laboratory test) you should keep in mind the context of the experiment. For example, one shouldn t interpret a P value from a screening assay the same way as a P value from a follow up study. You need to interpret results in the context of previous knowledge and theory. Bayesian calculations help you do this.

See also:

Bayesian analysis of statistically significant results

Bayesian analysis of clinical lab tests.

Bayesian analysis of linkage data.

The likelihood ratio.

Why are Bayesian calculations controversial?

Bayesian equations.

Bayesian analysis of statistically significant results

When you calculate a P value less than 0.05 (or whatever value you set for α) you conclude that the result is statistically significant. But that does not mean that there definitely is a real difference between groups. It is possible that the two groups are identical, and you just happened to get a small P value by chance. In other words, you may have made a Type I error.

What is the chance of making a Type I error? This question is ambiguous. To make a Type I error, two things must be true: the two populations must really have the same mean, and the P value must be less than α (and thus statistically significant). Therefore the question what is the chance of making a Type I error has (at least) can mean different things to different people:

1. Assume that the two populations really have the same mean. Now imagine all the results that could be obtained from experiments with

these populations. What fraction of these experiments will yield a statistically significant result?

2. Imagine all experiments that yield a statistically significant result. In what fraction of these experiments do the two populations really have the same mean?

By definition, the answer to the first question is α, usually set to 0.05.

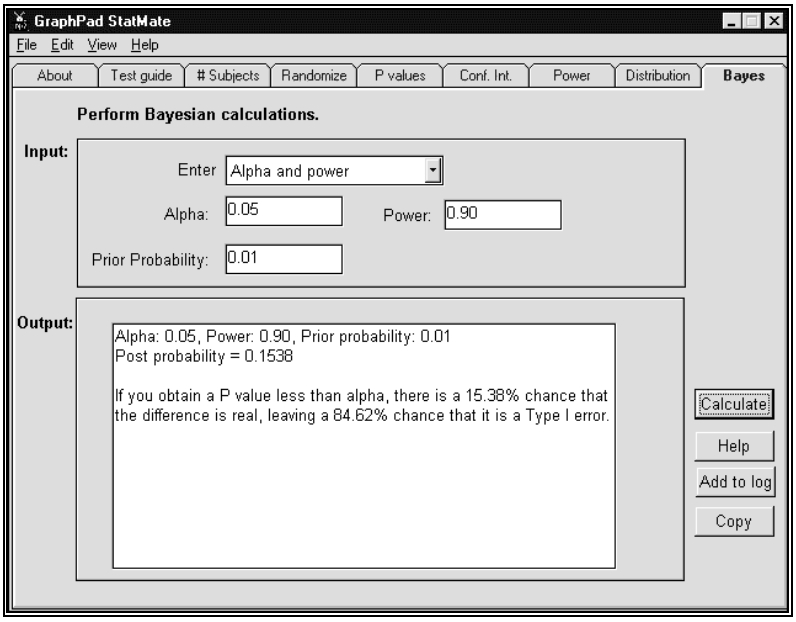
The second question is different, and the answer is not equal to α. You cannot answer that question without defining the prior probability that the experimental hypothesis is true and applying Bayesian logic. The prior probability must be assigned a value based on information you knew before collecting the data.

Theory and examples

Let's assume that you are performing a screening assay for new drugs. The drugs you screen are not selected for any particular reason. The chance that any particular drug will be effective in your assay is tiny, say 1%. In this context, you expect that most statistically significant results are Type I errors. The next paragraph works through an example to justify this conclusion.

What do you expect to find in 1000 experiments? You d expect that 10 of these (1%) would be with drugs that are really effective. Not all of these will yield a statistically significant result. It depends on the power of the study. Let s say that your sample size is large enough to give you 90% power. You d expect that 9 of those 10 experiments would give you a P value less than 0.05 and be called statistically significant. The other 990 experiments are with drugs that are not effective. But since you set $\alpha = 0.05$, 5% of these experiments will result in a P value less than 0.05 and will be called statistically significant. So you expect 5% of 990, or about 49, of the experiments with ineffective drugs to give you a statistically significant result (Type I errors). Out of the 1000 experiments, then, 58 (9 + 49) gave statistically significant results. In only 9 of these experiments, or about 15%, was the drug really effective. When you see a statistically significant result in this context, you can conclude that there is a 15% chance that the drug is truly effective and a 85% chance that you ve made a Type I error. The vast majority of the statistically significant results are Type I errors.

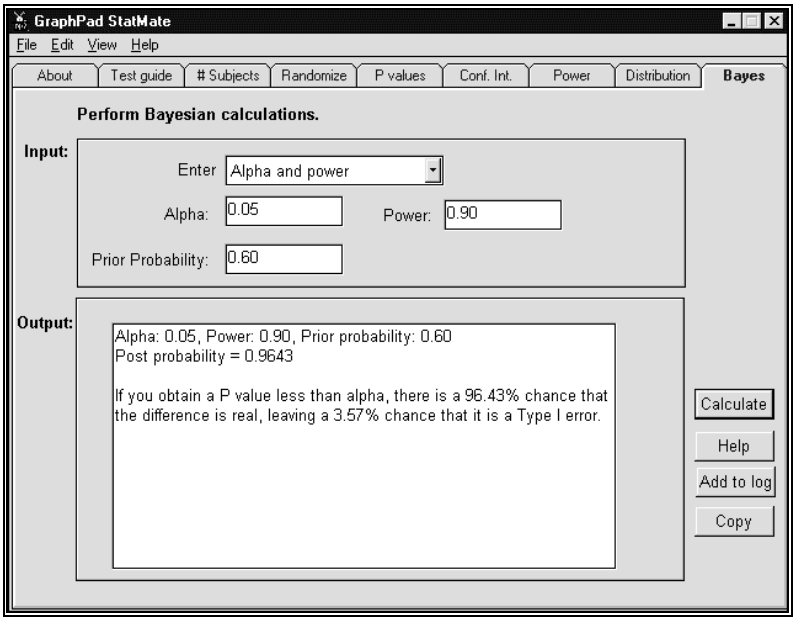
StatMate performs these calculations for you. Set $\alpha = 0.05$, power = 90%, and prior probability = 1%.



Now let's assume that you are testing drugs that have been carefully chosen. You expect that many of these drugs, say 60%, will be active in your assay. In this context, you expect that most of the statistically significant results will be with active drugs. The next paragraph shows you calculations to justify this conclusion.

What would you expect to see in 1000 experiments? In 600 of the experiments, the drug is really effective. Given 90% power (depends on sample size), you d expect 540 statistically significant results. The other 400 experiments are with inactive drugs. You expect 5% of these, or 20, to give statistically significant results by chance, even though the drug is really ineffective. Altogether you expect 560 statistically significant results and you expect that 540/560 or 96% of these cases will the drug really be effective. When you see a statistically significant result in this context, you can conclude that there is a 96% chance that the drug is truly effective and a 4% chance that you ve made a Type I error.

To make these calculations with StatMate, set $\alpha = 0.05$, power = 90%, and prior probability = 60%.



These examples demonstrate that the interpretation of a statistically significant P value depends on the context of the experiment. Few scientists perform Bayesian calculations explicitly, as shown in these examples. But most good scientists adjust for prior probability informally they just account for prior knowledge and theory when interpreting the data. A good scientist would not interpret a P value of 0.04 in a screening assay the same as a P value of 0.04 in a confirming experiment. Bayesian reasoning explains why one should interpret data in the context of what you already knew from theory and previous data. Since different people have different ideas about previous data and theory, different people will interpret the same data differently.

Step-by-step instructions

1. Choose that you want to enter alpha and power.
2. Enter a value for α. This is how you define statistical significance. If a P value is less than α, the results will be deemed statistically significant. If you plan to report a two-tailed P value, enter the two-tailed value for α. If you plan to report a one-tailed P value, then enter the one-tailed value for α. Most commonly, α is set to 0.05.
3. Enter a value for power. Power is a probability. It answers this question: If the two populations are really different, what is the chance of

obtaining a P value less than α in your study? To learn how to calculate the power of your study, see Power on page 37.

4. Enter a value for the prior probability. This is your best estimate for the probability that a difference really exists between the populations.
5. Press Calculate.

StatMate expects you to enter both power and the prior probability as fractions i.e. 0.80. If you enter a percentage i.e. 80 StatMate will automatically convert the value to a fraction.

Bayesian analysis of clinical lab tests

Many lab tests report a result as normal or abnormal. When the result is abnormal, that may be because the patient has the disease you are testing for. Or it may be a false positive.

What is the probability of obtaining a false positive result? This question is ambiguous. For a result to be a false positive, two things must be true: The patient must really be disease free, and the result must be positive. Therefore the question what is the probability of obtaining a false positive result? can be interpreted in two ways:

- What fraction of all disease free individuals will have a positive test result?
- What fraction of all positive test results are false positive? The answer is defined to be the rate of false positives. To calculate the answer to this question, you need to know the prior probability that the patient had the disease based on information you had before you saw the test results.

Theory and examples

Let s follow an example. Acute intermittent porphyria is a very rare disease difficult to diagnose clinically. It can be diagnosed by observing a low level of an enzyme, but the diagnosis isn t always clear because the enzyme levels vary in both the normal population and in people with the disease.

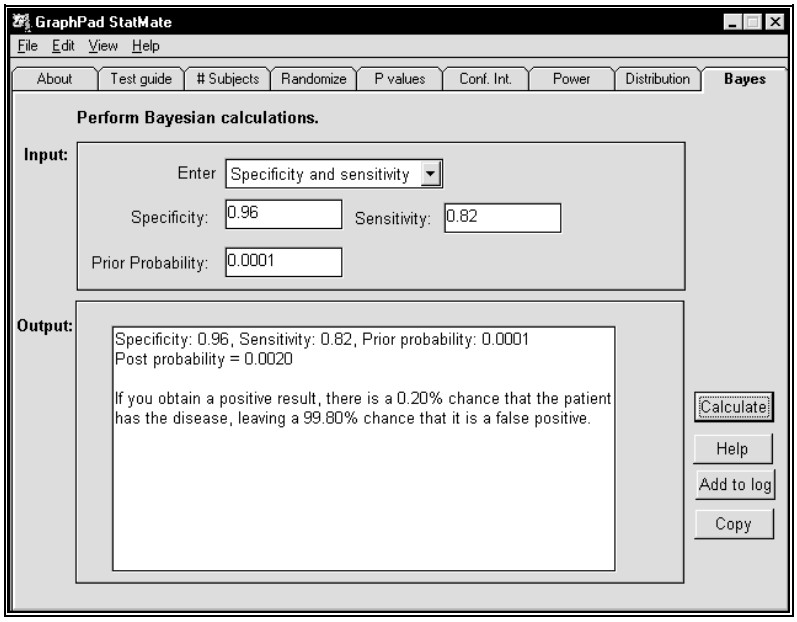
The accuracy of the test can be quantified by sensitivity and specificity. Sensitivity is the fraction of patients with the disease you are testing for who will have a positive test. For this example, the sensitivity is 82%. If you have the disease, there is an 82% chance that the lab test will be abnormal. Specificity is the fraction of patients without the disease you are testing for who will have a negative test. In this example, the specificity is 96%. That is the chance that someone without the disease will (correctly) get a negative lab

test. That leaves a 4% chance that someone without the disease will have a positive test result. This is the answer to the first question posed above.

A patient has an abnormal lab test. What is the chance that he or she has the disease? You need Bayesian analysis to answer the question, as it depends on the prior probability.

Patient A is healthy, and there is no reason to suspect this disease. The test was done as a screening test. The disease is rare (1 in 10,000 people) so the chance that Patient A has the defective gene is tiny. A positive test in such a person is very likely to be a false positive. From the specificity, you know that about 4% of normal people will get a positive test result. Since the disease is so rare, the 4% of the normals who have a false positive test will be far more numerous than true positives. A positive result is very likely to be due to a false positive. Bayesian calculations quantify this.

Enter specificity = 0.96, sensitivity = 0.82, and prior probability = 0.0001. Then press Calculate.

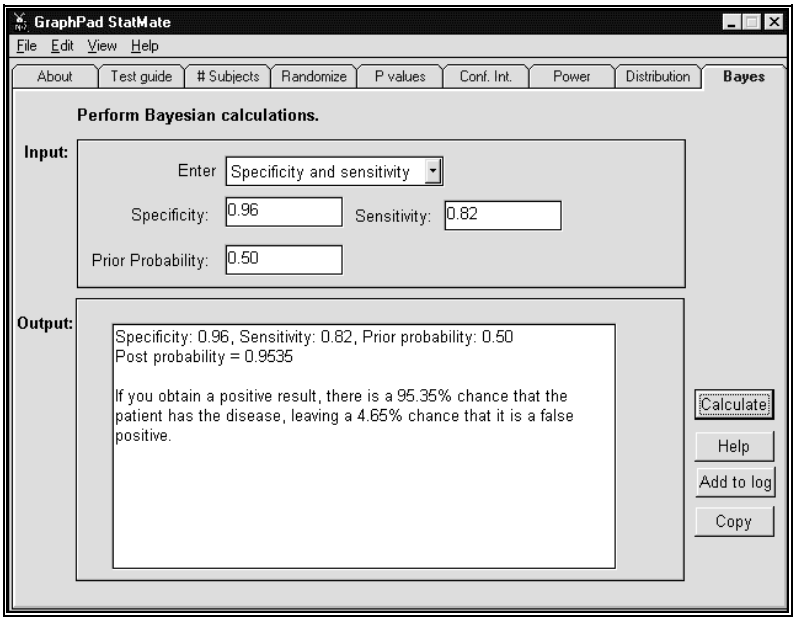


StatMate calculates that the post test probability is 0.0020. Despite the abnormal test result, there is only a 0.2% chance that Patient A has the defective gene. The chances are overwhelming (>99%) that the result is a false positive.

Patient B has a brother with the disease. Since the disease is autosomal dominant, you knew before doing the test that there is a 50% chance that Pa-

tient B has the defective gene. If he does have the defective gene, there is a 86% chance that the test will be positive (sensitivity). If he does not have the defective gene, there is a 4% chance that the test will be positive (false positive, calculated from the specificity). Given this information, we know that it is likely that the positive test is correct, and that there is a low chance that this is a false positive result. Bayes calculations quantify this intuition.

Enter specificity = 0.96, sensitivity = 0.82, and prior probability = 0.50. Then press Calculate.



StatMate calculates that the post test probability is 0.95. Given the family history and the abnormal test result, there is a 95% chance that Patient B has the defective gene. There is only a 5% chance that the finding is a false positive.

Step-by-step instructions

1. Choose that you want to enter sensitivity and specificity.
2. Enter a value for specificity. Specificity is the fraction of patients without the disease you are testing for who will have a negative test.
3. Enter a value for sensitivity. Sensitivity is the fraction of patients with the disease you are testing for who will have a positive test.
4. Enter a value for the prior probability. This is your best estimate for the fraction of the population you are testing that has the disease. This

may come from epidemiological data, from genetic theory, or from clinical intuition.

5. Press Calculate.

StatMate expects you to enter sensitivity, specificity and the prior probability as fractions i.e. 0.80. If you enter a percentage i.e. 80 StatMate will automatically convert the value to a fraction.

Bayesian analysis of linkage data

When two loci (genes or DNA sequences) are located near each other on the same chromosome, they are said to be linked. If the two loci are very close, crossing over or recombination between the two loci occurs rarely. Thus. alleles of linked loci tend to be inherited together.

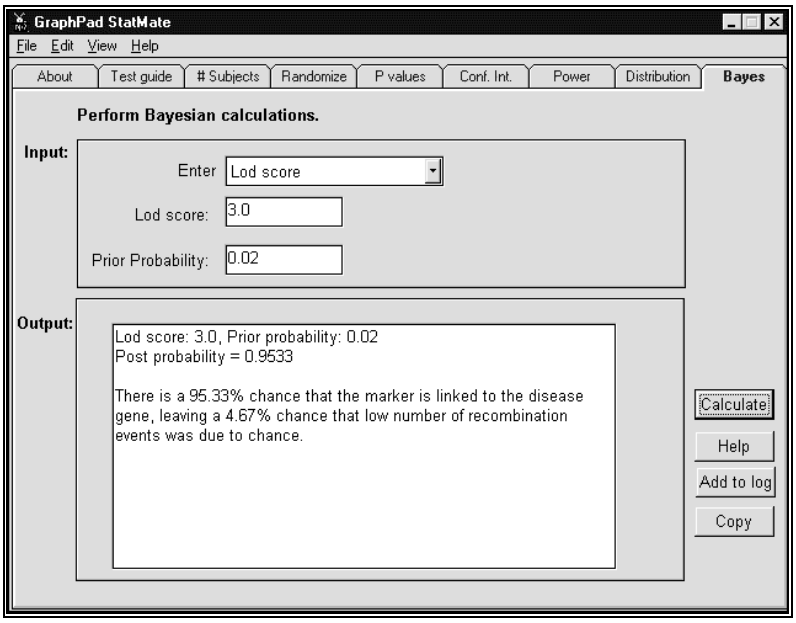
Linkage is useful in genetic diagnosis and mapping. Since it is not possible to detect all abnormal genes directly, geneticists try to identify a marker gene or variable DNA sequence that is linked to the disease gene. The presence of the marker (that you can identify) then can be used to predict the presence of the disease gene (that you can t directly measure).

How can you tell if a marker is linked to a disease gene? Geneticists study large families and observe how often the marker and disease are inherited together, and how often there is recombination. If there are few recombination events, there are two possible explanations. The marker and gene could be linked. Or the two could not be linked, but just happened to be inherited together by coincidence.

The results of linkage studies are reported as the lod score (for log of odds). If you take the antilogarithm of the lod score, you get the likelihood ratio. For example if the lod score is 3, the likelihood ratio is 1000 (10^3). This means that the observed small number of recombination events is 1000 times more likely to occur if the marker is linked to the disease than if the two are not linked. (Calculating a lod score is difficult, and this calculation is not done by StatMate.)

What is the chance that the marker and disease are linked? To answer this question, you must perform Bayesian calculations. To do this, you must know the prior probability the probability that a random marker would be linked to a particular gene. Since there are 23 pairs of chromosomes, the probability that a marker and gene would be on the same chromosome is 1/23 or 4.3%. But it is not enough to be on the same chromosome the marker and disease gene must be near each other in order to be linked. Genetic data tell us that 2% of randomly selected markers will be linked to any particular disease gene (this depends on exactly how you define linked).

If the lod score is 3, what is the probability that the marker and disease are actually linked? Enter lod = 3.0 and prior probability = 0.02. Then press Calculate.



StatMate calculates that the post test probability is 0.95. Combining the prior probability and the lod score, there is a 95% chance that the marker and disease are linked. There is only a 5% chance that the two are not linked and that the low number of recombination events was a coincidence. This answer is based on the assumption that the marker was randomly selected. If you had a reason to think that the marker was linked to the disease (before doing the experiment), the prior probability would be higher and the results would be different.

The likelihood ratio

When performing the Bayesian calculations, StatMate s first step is to calculate the likelihood ratio, which it reports in the results. Its meaning depends on the context:

- Statistical comparisons. The likelihood ratio is power/α. It is the chance of obtaining a statistically significant result if there really is a difference between populations divided by the chance of obtaining a statistically significant result if there really is no difference between populations.

- Lab tests. The likelihood ratio is sensitivity/(1-specificity). It is a chance of obtaining an abnormal lab test from someone with the disease divided by the chance of obtaining an abnormal lab test from someone without the disease.
- Linkage. The likelihood ratio is the antilogarithm of the lod score (10^{lod}). It is the chance of observing so few recombination events if the marker and disease gene are linked divided by the chance of observing so few recombination events if the marker and disease gene are not linked.

StatMate also lets you enter the likelihood ratio directly. This lets you perform Bayesian calculations in other situations.

Why are Bayesian calculations controversial?

Bayesian calculations are not widely used. There are two problems.

Is the prior probability really a probability?

To perform Bayesian calculations, you must state the prior probability.

In some situations, you can reliably state the value of the prior probability. For example, in genetic linkage the probability that a randomly selected marker is linked to a random disease is known quite accurately. Another example is clinical testing. You often know the prior probability from epidemiological data (the prevalence) or from genetic theory (each sibling of someone with an autosomal dominant disease has a 50% of having the abnormal gene). When the prior probabilities are known, there is nothing controversial about using Bayesian calculations.

In other situations, you cannot reliably state the value of the prior probability. Instead, the best you can do is estimate the value based on experience or hunch. Thus you might think that an experiment is based on flimsy evidence and that the experimental hypothesis is very unlikely to be true. To use Bayesian calculations, you need to convert that subjective feeling of uncertainty into a number I think there is only 5% chance that the experimental hypothesis is true. At best, prior probabilities determined this way are estimates. Many statisticians think it is OK to enter these estimates as prior probabilities into Bayesian calculations, so long as you realize that the results of the Bayesian calculations are no more accurate than are the values upon which the calculations are based. Others statisticians say that you should never equate a subjective feeling of uncertainty with a probability. These statisticians would say that a vague feeling that you are 30% sure is not the same as a prior probability of 30%, and that Bayesian calculations should not be done unless you have a solid basis for knowing the prior probability.

Are you counting the prior probability twice?

When interpreting experimental evidence, you need to combine the results of the current experiment with previous data and theory. Many people combine the two informally, without performing explicit Bayesian calculations. For example, when examining results of screening assays for new drugs, you don t get too excited by a P value of 0.04 because you know it is likely to be a Type I error. When examining the results of a confirming experiment, you might get quite excited by the same P value.

You can also combine prior evidence with experimental evidence using Bayesian calculations. When you do this, you must remember that the resulting post probability already includes the prior evidence. Once you ve performed the Bayesian calculations, you should interpret a post test probability of 0.01 the same, regardless of whether it came from a screening test or a confirming experiment. It would be a mistake to informally combine the Bayesian result with your prior knowledge. If you did so, you would be counting the prior probability twice.

The results log

Adding results to the log

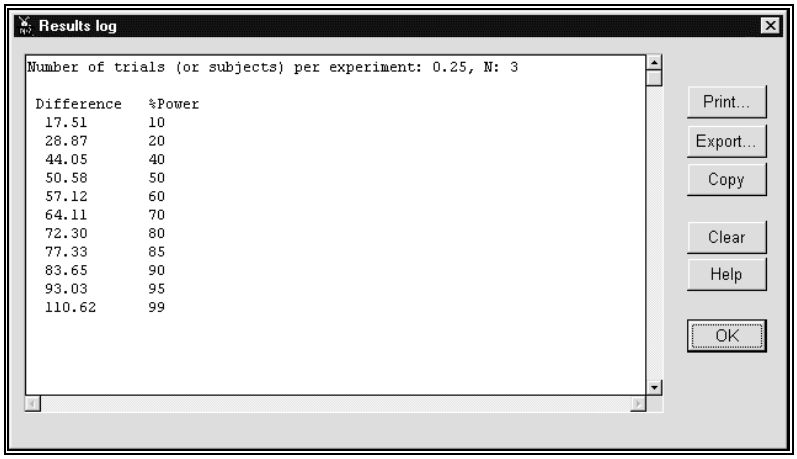
The results and tables produced by StatMate are compact. To save paper, StatMate lets you accumulate several sets of results before you print a page.

To bring up the results log:

- After using any of StatMate s calculators, press the button labeled Add to log. This appends the results to the log, and brings up the results log dialog.

or

- Pull down the View menu and choose Results Log. This brings up the results log dialog, without adding to it.



To	Do this
Edit the log.	Click the mouse and type. You may delete portions or add comments or labels.
Go back to the main StatMate Calculator Window.	Press OK. All the results remain in the log.
Erase the log.	Press Clear.
Copy the log to the clipboard.	Press Copy. If you select a portion of the log first, only the selection will be copied.
Print the log.	Press Print.
Save the log to a text file.	Press Export. See Exporting the results log on page 59.
Resize the log window.	Put your mouse over an edge and drag.

Printing

After accumulating several sets of results in the log, you can print.

To print the results log:

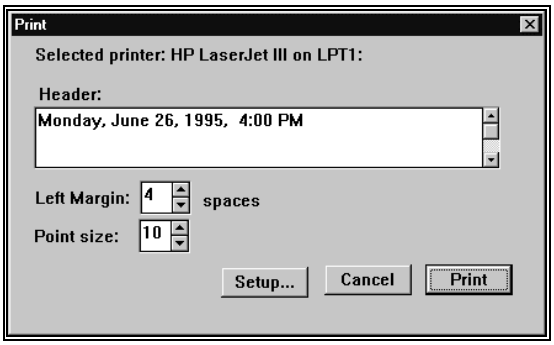
1. If the results log is not visible, do one of the following:

   Press the button labeled Add to log. This appends the current results to the log.

   or

   Pull down the View menu, and choose Results log. This brings up the results log without adding to it.

2. Press the Print button to bring up the Print dialog.



3. Change the header, margin and point size.
4. Press Print.

StatMate always prints using the Courier font. This is a nonproportional font (all characters have the same width) which ensures that the tables (i.e. sample size results) are aligned.

Initially the header contains the current date and time. These are formatted according to the settings in the Windows control panel. You can change the date format in the International (Windows 3.1) or Regional Settings (Windows 95) portion of the Windows control panel.

It is not possible to print directly from a StatMate calculator. You must first add results to the results log, and then print the entire log (which may have just one set of results).

Exporting the results log

To export the results log to a text file:

1. If the results log is not visible, do one of the following:

   Press the button labeled Add to log. This appends the current results to the log.

   or

   Pull down the View menu, and choose Results log. This brings up the results log without adding to it.
2. Press the Export button to bring up the Export dialog.
3. Enter a file name.

The results log will be a plain text file. You may import this file into any word processor.

After importing tabular results (i.e. sample size) into your word processor, the columns may not be aligned. There are two ways to deal with this problem:

- Use a nonproportional font such as Courier. The tables will be aligned when you use a nonproportional font.
- Adjust your tab stops in the program, and add tabs (and erase spaces as necessary) to get things to line up.

An alternative approach may work better for bringing tabular results into a word processor. Instead of exporting the results log to a text file, copy the results table to the clipboard. Copy directly from a StatMate calculator, not from the results log. Then paste into your word processor. Finally, instruct your word processor to convert those lines into a table. For example, in Mi-

crosoft Word, select the text, pull down the Table menu and choose Convert text to table.

Copying results to the clipboard

You may copy results to the clipboard either directly from one of StatMate s calculators or from the results log. The distinction is important for tabular results:

- When a table is copied from a StatMate calculator, adjacent columns are separated by tabs. This is the format that spreadsheet and graphics programs expect.
- When tabular results are copied from the results log, adjacent columns are separated by one or more spaces. This lines up the columns when printed or displayed using a nonproportional font.

Your choice between the two formats depends on where the results will be pasted.

- When transferring results tables to a spreadsheet or graphics program, you should copy the results from a StatMate calculator, and not from the results log.
- When transferring to a word processor, copy from the results log if you are satisfied with using a nonproportional font. If you want to use a proportional font, copy from a StatMate calculator, but plan to spend a bit of time reformatting the data in the word processor.

To copy results to the clipboard (either from a calculator or from the results log):

1. If you only want to copy a portion of the results, select that portion. If you want to copy all the results, you don t need to select anything.
2. Press the Copy button.

   or

   Pull down the Edit menu and choose Copy results.

Algorithms used

Algorithm for sample size and power. Comparing two means.

StatMate follows the equations in Jacob Cohen, Statistical Power Analysis for the Behavioral Sciences, Second Edition, 1988, Lawrence Erlbaum Associates, Hillsdale NJ.

Calculating sample size and power is not entirely straightforward. Different books give different equations, and there is no strong reason to pick one over another. The differences between different methods are minor, and are not important so long as you remember that the goal is to estimate sample size and power.

This equation computes the detectable difference between means as a function of alpha, beta, N and SD.

$$\Delta \approx \frac{\left(z_{1-b} + z_{1-a/2}\right) \cdot SD \cdot \left(2(N-1) + 1.21 \cdot (z_{1-a/2} - 1.06)\right)}{(N-1)\sqrt{2N}}$$

The variables are:

Variable	Meaning
Δ	Difference between means that can be detected with specified power.
z_{1-b}	The value on a standard Gaussian distribution such that a fraction equal to β of the distribution is greater. For example, if power is 95%, then $\beta = 0.05$, and $z_{1-\beta} = 1.65$. This means that 5% of a standard Gaussian distribution has a value greater than 1.65. In other words, 5% of the values in a Gaussian distribution are more than 1.65 standard deviations away from the mean.

$Z_{1-a/2}$	If you set $\alpha = 0.05$ (two tailed), then $z_{1-\alpha/2} = 1.96$. This means that 2.5% of a standard Gaussian distribution has a value greater than 1.96, and that 5% of a standard Gaussian distribution has a value either greater than 1.96 or less than -1.96.
N	The sample size in each group. When calculating power from groups with unequal sample size, N is the harmonic mean of the two sample sizes.
SD	The standard deviation. When calculating sample size, the SD is assumed equal in both groups. When calculating power, this is the pooled standard deviation.

When calculating the power of a test, N is the harmonic mean of the two sample sizes N_1 and N_2, and SD is the pooled standard deviation calculated from SD_1 and SD_2.

$$\mathrm{N} = \text{Harmonic Mean} = \frac{2 \cdot \mathrm{N}_1 \cdot \mathrm{N}_2}{\mathrm{N}_1 + \mathrm{N}_2}$$

$$\mathrm{SD} = \text{Pooled SD} = \sqrt{\frac{(\mathrm{N}_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}}$$

Algorithm for sample size and power. Comparing two proportions.

StatMate follows the equations in Jacob Cohen, Statistical Power Analysis for the Behavioral Sciences, Second edition, 1988, Lawrence Erlabaum Associates, Hillsdale NJ.

The power of a test to find a difference between two proportions, p_1 and p_2, is not a simple function of the difference between two proportions. Nor is it a simple function of the ratio of two proportions. The power to find a difference between 0.2 and 0.4 is not the same as the power to find a difference between 0.3 and 0.5 (same difference) or between 0.3 and 0.6 (same ratio).

It is possible to transform the two proportions into a measure that is simply related to power. This transform is:

$$h = 2\left|\text{Arcsine}\sqrt{p_1}) - \text{Arcsine}(\sqrt{p_2})\right| \qquad \text{(Equation 1)}$$

The arcsine is the inverse of the sine. Arcsine(x) is the angle in radians whose sine equals x. It is surprising to see trigonometric functions show up in statistical computations. But it turns out that power is a consistent function of the difference of the arcsines of the square roots of the proportions. The power of study will be the same for any two proportions that yield the same value of h.

If you know p_1 and p_2, you can calculate h. If you also choose values for α and β (1.0 minus power), you can calculate the sample size in each group, N:

$$N = 2\left(\frac{z_{1-b} + z_{1-a/2}}{h}\right) \qquad \text{(Equation 2)}$$

StatMate rearranges this equation to calculate h from known values of α, β and N. It then rearranges the previous equation to calculate p_2 from known values of p_1 and h. Finally, it displays the difference p_2-p_1 or the ratio p_2/p_1.

For the sample size calculations, you enter the anticipated proportion in the control group and this becomes p1. Knowing h and p_1, equation 1 can be solved to find two different values for p_2, one greater than p_1 and the other smaller than p_1. If you stated that you are looking for an increase, then StatMate chooses the value of p_2 that is larger than p_1. If you are looking for a decrease, then StatMate chooses the other value, where p_2 is less than p_1.

For the power calculations, you enter the observed proportions for each group. StatMate computes the average (not weighted for sample size) of these two numbers, p_{av}, and ignores their individual values. For each value of power, StatMate solves equation 2 to find the value of h. N is the harmonic mean of the two sample sizes.

$$\text{N} = \text{Harmonic Mean} = \frac{2 \cdot \text{N}_1 \cdot \text{N}_2}{\text{N}_1 + \text{N}_2}$$

To find values of p_1 and p_2 , StatMate solves Equation 1 as well as equation 3 below. Equation 3 ensures that the hypothetical values p_1 and p_2 are centered on the value of p_{av} that you observed in your experiment.

$$\text{Arcsine}(\sqrt{p_{av}}) - \text{Arcsine}(\sqrt{p_1}) = \text{Arcsine}\sqrt{p_2}) - \text{Arcsine}(\sqrt{p_{av}}) \qquad \text{(Eq. 3)}$$

Randomizing method

StatMate uses the random number generator built in to the Borland Delphi (version 1.0) compiler. The Randomize function is used first so that the random numbers are different every time you run the program.

P value computations

To compute P values from z, r, t, F or chi-square, StatMate uses algorithms from William H. Press, Saul A. Teukolsky, Willliam T. Vetterling, and Brian P. Flannery, Numerical Recipes in C. Second Edition, 1992.

Confidence intervals

The equations for calculating the confidence intervals of a count of proportion were adapted from Lothar Sachs, Applied Statistics, Second edition, 1982.

The 95% confidence interval of a count C is calculated from the chi-square distribution with equation 1.180 on page 182. The 95% confidence interval of a proportion (N/D) is calculated from the F distribution with Equation 4.20 on page 333.

Bayesian equations

The Bayesian calculations follow these steps:

1. Calculate the likelihood ratio:

Context	Equation
Statistical comparisons.	power/α.
Lab tests.	sensitivity/(1-specificity)
Linkage.	10^{lod}

2. Convert the prior probability to odds.

$$\text{Prior Odds} = \frac{\text{Prior Probbility}}{1 - \text{Prior Probability}}$$

3. Calculate the post odds.

$$\text{Post Odds} = \text{Prior Odds} \cdot \text{Likelihood Ratio}$$

4. Convert odds to probability.

$$\text{Post Probability} = \frac{\text{Post Odds}}{1 + \text{Post Odds}}$$

GraphPad Software

Software license agreement

If you purchased StatMate via the Internet, you agreed to this license before completing the purchase. If you purchased Statmate by mail, you agreed to this license when you opened the disk envelope.

GraphPad Software, Inc. ("GraphPad") grants you a limited license to use In-Stat (the "Licensed Software") on a single computer. You may only install StatMate on more than one computer if you are the sole user. If more than one person will use StatMate, it must be installed on only one shared computer. Any other installation is an infringement of GraphPad's copyright.

GraphPad provides the Licensed Software "as is", and does not warrant that it is error-free. You should check important results carefully before drawing conclusions. The Licensed Software is designed for research purposes only, and should not be used for the diagnosis or treatment of patients.

GraphPad shall not be liable for any consequential, incidental, special, punitive, exemplary or other damages, even if advised of the possibility of such damages. GraphPad expressly disclaims any and all warranties, expressed or implied. GraphPad's entire liability, and your exclusive remedy, shall be to replace the Licensed Software or refund the license fee.

Title to the Licensed Software remains with GraphPad at all times. You may not sublicense, rent, lease, convey, or modify the software. The laws of the State of California, U.S.A, govern this agreement.

Technical support

GraphPad Software provides free technical support to the registered user for one year after the date of purchase.

- Our free support does not include free statistical consulting. We can sometimes provide limited statistical consulting for an hourly fee.

- If several people in your lab share one copy of StatMate (on a single computer), please route all technical support calls through one person when possible.
- If possible, please be near your computer when you call.

To contact GraphPad Software:

EMAIL:	support@graphpad.com *or* sales@graphpad.com
PHONE:	(USA) 619-457-3909
FAX:	(USA) 619-457-8141
WEB:	www.graphpad.com
MAIL:	GraphPad Software, Inc.
	10855 Sorrento Valley Road #203
	San Diego, CA 92121 USA

Citing GraphPad StatMate

When citing analyses performed by the program, include the name of the analysis, and Statmate version number (including the second number after the decimal point). Use this example as a guide: Sample sizes were calculated using GraphPad StatMate version 1.01i, GraphPad Software, San Diego California USA, www.graphpad.com .

To find the full version number, pull down the Help menu and choose About GraphPad StatMate.

Uninstalling Statmate

Delete the StatMate folder (directory) to uninstall StatMate. Also delete StatMate s icons (shortcuts). Statmate consists of three files: statmate.exe, statmate.cnt and statmate.hlp. With Windows 95 or NT, the file statmate.gid is created.

Intuitive Biostatistics (the book)

If you like the style of this manual, you'll probably also like *Intuitive Biostatistics*, by Harvey Motulsky, President of GraphPad Software (and primary author of this manual). Here is the description from the back of the book:

"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.

Index

assigning groups randomly 25

Bayesian calculations, why they are controversial 54
Beta 15
binomial distribution 41

chi-square, calcuating P value from 27
choosing statistical tests 11
confidence interval defined 33
copying results 60
count, confidence interval of 33

exporting results 59

F, calculating a P value from 27

Gaussian distribution 11
genetic linkage 52
GraphPad Prism 8
GraphPad Software 68

hypothesis testing 14

Installing 8

License, software 67
likelihood ratio 53
linkage 52
lod score 52

McNemar's test 30
mean, confidence interval of 33

N, determining sample size 13

P value defined 31
Poisson distribution 41
power 15
printing 58
Prism 8
proprotion, confidence interval of 33
pseudo-random numbers 26

r, calculating a P value from 27
randomizing 25
results log 57
runs test 28

sample size, calculating 13
sensitivity 49, 51
sign test 30
Software license 67
specificity 49, 51

t, calculating a P value from 27
technical support 67
test guide 11
Type I vs. Type II error 14

unequal sample sizes 21

z, calculating a P value from 27