

# **STATA POWER AND SAMPLE-SIZE REFERENCE MANUAL RELEASE 14**



A Stata Press Publication  
StataCorp LP  
College Station, Texas



Copyright © 1985–2015 StataCorp LP  
All rights reserved  
Version 14

Published by Stata Press, 4905 Lakeway Drive, College Station, Texas 77845  
Typeset in  $\text{T}_{\text{E}}\text{X}$

ISBN-10: 1-59718-163-3  
ISBN-13: 978-1-59718-163-1

This manual is protected by copyright. All rights are reserved. No part of this manual may be reproduced, stored in a retrieval system, or transcribed, in any form or by any means—electronic, mechanical, photocopy, recording, or otherwise—without the prior written permission of StataCorp LP unless permitted subject to the terms and conditions of a license granted to you by StataCorp LP to use the software and documentation. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted by this document.

StataCorp provides this manual “as is” without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. StataCorp may make improvements and/or changes in the product(s) and the program(s) described in this manual at any time and without notice.

The software described in this manual is furnished under a license agreement or nondisclosure agreement. The software may be copied only in accordance with the terms of the agreement. It is against the law to copy the software onto DVD, CD, disk, diskette, tape, or any other medium for any purpose other than backup or archival purposes.

The automobile dataset appearing on the accompanying media is Copyright © 1979 by Consumers Union of U.S., Inc., Yonkers, NY 10703-1057 and is reproduced by permission from CONSUMER REPORTS, April 1979.

Stata, **STATA** Stata Press, Mata, **MATA** and NetCourse are registered trademarks of StataCorp LP.

Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations.

NetCourseNow is a trademark of StataCorp LP.

Other brand and product names are registered trademarks or trademarks of their respective companies.

For copyright information about the software, type `help copyright` within Stata.

The suggested citation for this software is

StataCorp. 2015. *Stata: Release 14*. Statistical Software. College Station, TX: StataCorp LP.

# Contents

intro	Introduction to power and sample-size analysis	1
GUI	Graphical user interface for power and sample-size analysis	14
power	Power and sample-size analysis for hypothesis tests	27
power, graph	Graph results from the power command	56
power, table	Produce table of results from the power command	79
power onemean	Power analysis for a one-sample mean test	88
power twomeans	Power analysis for a two-sample means test	102
power pairedmeans	Power analysis for a two-sample paired-means test	120
power oneproportion	Power analysis for a one-sample proportion test	136
power twoproportions	Power analysis for a two-sample proportions test	154
power pairedproportions	Power analysis for a two-sample paired-proportions test	177
power onevariance	Power analysis for a one-sample variance test	199
power twovariances	Power analysis for a two-sample variances test	213
power onecorrelation	Power analysis for a one-sample correlation test	228
power twocorrelations	Power analysis for a two-sample correlations test	241
power oneway	Power analysis for one-way analysis of variance	255
power twoway	Power analysis for two-way analysis of variance	276
power repeated	Power analysis for repeated-measures analysis of variance	297
power cmh	Power and sample size for the Cochran–Mantel–Haenszel test	325
power mcc	Power analysis for matched case–control studies	347
power trend	Power analysis for the Cochran–Armitage trend test	364
power cox	Power analysis for the Cox proportional hazards model	381
power exponential	Power analysis for the exponential test	401
power logrank	Power analysis for the log-rank test	436
unbalanced designs	Specifications for unbalanced designs	462
Glossary		475
Subject and author index		485



# Cross-referencing the documentation

When reading this manual, you will find references to other Stata manuals. For example,

[U] [26 Overview of Stata estimation commands](#)

[R] [regress](#)

[D] [reshape](#)

The first example is a reference to chapter 26, *Overview of Stata estimation commands*, in the *User's Guide*; the second is a reference to the `regress` entry in the *Base Reference Manual*; and the third is a reference to the `reshape` entry in the *Data Management Reference Manual*.

All the manuals in the Stata Documentation have a shorthand notation:

[GSM]	<i>Getting Started with Stata for Mac</i>
[GSU]	<i>Getting Started with Stata for Unix</i>
[GSW]	<i>Getting Started with Stata for Windows</i>
[U]	<i>Stata User's Guide</i>
[R]	<i>Stata Base Reference Manual</i>
[BAYES]	<i>Stata Bayesian Analysis Reference Manual</i>
[D]	<i>Stata Data Management Reference Manual</i>
[FN]	<i>Stata Functions Reference Manual</i>
[G]	<i>Stata Graphics Reference Manual</i>
[IRT]	<i>Stata Item Response Theory Reference Manual</i>
[XT]	<i>Stata Longitudinal-Data/Panel-Data Reference Manual</i>
[ME]	<i>Stata Multilevel Mixed-Effects Reference Manual</i>
[MI]	<i>Stata Multiple-Imputation Reference Manual</i>
[MV]	<i>Stata Multivariate Statistics Reference Manual</i>
[PSS]	<i>Stata Power and Sample-Size Reference Manual</i>
[P]	<i>Stata Programming Reference Manual</i>
[SEM]	<i>Stata Structural Equation Modeling Reference Manual</i>
[SVY]	<i>Stata Survey Data Reference Manual</i>
[ST]	<i>Stata Survival Analysis Reference Manual</i>
[TS]	<i>Stata Time-Series Reference Manual</i>
[TE]	<i>Stata Treatment-Effects Reference Manual: Potential Outcomes/Counterfactual Outcomes</i>
[I]	<i>Stata Glossary and Index</i>
[M]	<i>Mata Reference Manual</i>



# Title

**intro** — Introduction to power and sample-size analysis

[Description](#)

[Remarks and examples](#)

[References](#)

[Also see](#)

## Description

Power and sample-size (PSS) analysis is essential for designing a statistical study. It investigates the optimal allocation of study resources to increase the likelihood of the successful achievement of a study objective.

## Remarks and examples

Remarks are presented under the following headings:

*Power and sample-size analysis*

*Hypothesis testing*

*Components of PSS analysis*

*Study design*

*Statistical method*

*Significance level*

*Power*

*Clinically meaningful difference and effect size*

*Sample size*

*One-sided test versus two-sided test*

*Another consideration: Dropout*

*Survival data*

*Sensitivity analysis*

*An example of PSS analysis in Stata*

*Video example*

This entry describes statistical methodology for PSS analysis and terminology that will be used throughout the manual. For a list of supported PSS methods and the description of the software, see [\[PSS\] power](#). To see an example of PSS analysis in Stata, see *An example of PSS analysis in Stata*. For more information about PSS analysis, see [Lachin \(1981\)](#), [Cohen \(1988\)](#), [Cohen \(1992\)](#), [Wickramaratne \(1995\)](#), [Lenth \(2001\)](#), [Chow, Shao, and Wang \(2008\)](#), [Julious \(2010\)](#), and [Ryan \(2013\)](#), to name a few.

## Power and sample-size analysis

Power and sample-size (PSS) analysis is a key component in designing a statistical study. It investigates the optimal allocation of study resources to increase the likelihood of the successful achievement of a study objective.

How many subjects do we need in a study to achieve its research objectives? A study with too few subjects may have a low chance of detecting an important effect, and a study with too many subjects may offer very little gain and will thus waste time and resources. What are the chances of achieving the objectives of a study given available resources? Or what is the smallest effect that can be detected in a study given available resources? PSS analysis helps answer all of these questions. In what follows, when we refer to PSS analysis, we imply any of these goals.

We consider prospective PSS analysis (PSS analysis of a future study) as opposed to retrospective PSS analysis (analysis of a study that has already happened).

Statistical inference, such as hypothesis testing, is used to evaluate research objectives of a study. In this manual, we concentrate on the PSS analysis of studies that use hypothesis testing to investigate the objectives of interest. The supported methods include one-sample and two-sample tests of means, variances, proportions, correlations, and more. See [PSS] **power** for a full list of methods.

Before we discuss the components of PSS analysis, let us first revisit the basics of hypothesis testing.

### Hypothesis testing

Recall that the goal of hypothesis testing is to evaluate the validity of a hypothesis, a statement about a population parameter of interest  $\theta$ , a target parameter, based on a sample from the population. For simplicity, we consider a simple hypothesis test comparing a population parameter  $\theta$  with 0. The two complementary hypotheses are considered: the null hypothesis  $H_0: \theta = 0$ , which typically corresponds to the case of “no effect”, and the alternative hypothesis  $H_a: \theta \neq 0$ , which typically states that there is “an effect”. An effect can be a decrease in blood pressure after taking a new drug, an increase in SAT scores after taking a class, an increase in crop yield after using a new fertilizer, a decrease in the proportion of defective items after the installation of new equipment, and so on.

The data are collected to obtain evidence against the postulated null hypothesis in favor of the alternative hypothesis, and hypothesis testing is used to evaluate the obtained data sample. The value of a test statistic (a function of the sample that does not depend on any unknown parameters) obtained from the collected sample is used to determine whether the null hypothesis can be rejected. If that value belongs to a rejection or critical region (a set of sample values for which the null hypothesis will be rejected) or, equivalently, falls above (or below) the critical values (the boundaries of the rejection region), then the null is rejected. If that value belongs to an acceptance region (the complement of the rejection region), then the null is not rejected. A critical region is determined by a hypothesis test.

A hypothesis test can make one of two types of errors: a type I error of incorrectly rejecting the null hypothesis and a type II error of incorrectly accepting the null hypothesis. The probability of a type I error is  $\Pr(\text{reject } H_0 | H_0 \text{ is true})$ , and the probability of a type II error is commonly denoted as  $\beta = \Pr(\text{fail to reject } H_0 | H_0 \text{ is false})$ .

A power function is a function of  $\theta$  defined as the probability that the observed sample belongs to the rejection region of a test for a given parameter  $\theta$ . A power function unifies the two error probabilities. A good test has a power function close to 0 when the population parameter belongs to the parameter’s null space ( $\theta = 0$  in our example) and close to 1 when the population parameter belongs to the alternative space ( $\theta \neq 0$  in our example). In a search for a good test, it is impossible to minimize both error probabilities for a fixed sample size. Instead, the type-I-error probability is fixed at a small level, and the best test is chosen based on the smallest type-II-error probability.

An upper bound for a type-I-error probability is a significance level, commonly denoted as  $\alpha$ , a value between 0 and 1 inclusively. Many tests achieve their significance level—that is, their type-I-error probability equals  $\alpha$ ,  $\Pr(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$ —for any parameter in the null space. For other tests,  $\alpha$  is only an upper bound; see [example 6](#) in [PSS] **power oneproportion** for an example of a test for which the nominal significance level is not achieved. In what follows, we will use the terms “significance level” and “type-I-error probability” interchangeably, making the distinction between them only when necessary.

Typically, researchers control the type I error by setting the significance level to a small value such as 0.01 or 0.05. This is done to ensure that the chances of making a more serious error are very small. With this in mind, the null hypothesis is usually formulated in a way to guard against what a researcher considers to be the most costly or undesirable outcome. For example, if we were to use hypothesis testing to determine whether a person is guilty of a crime, we would choose the



null hypothesis to correspond to the person being not guilty to minimize the chances of sending an innocent person to prison.

The power of a test is the probability of correctly rejecting the null hypothesis when the null hypothesis is false. Power is inversely related to the probability of a type II error as  $\pi = 1 - \beta = \Pr(\text{reject } H_0 | H_0 \text{ is false})$ . Minimizing the type-II-error probability is equivalent to maximizing power. The notion of power is more commonly used in PSS analysis than is the notion of a type-II-error probability. Typical values for power in PSS analysis are 0.8, 0.9, or higher depending on the study objective.

Hypothesis tests are subdivided into one sided and two sided. A one-sided or directional test asserts that the target parameter is large (an upper one-sided test  $H: \theta > \theta_0$ ) or small ( $H: \theta \leq \theta_0$ ), whereas a two-sided or nondirectional test asserts that the target parameter is either large or small ( $H: \theta \neq \theta_0$ ). One-sided tests have higher power than two-sided tests. They should be used in place of a two-sided test only if the effect in the direction opposite to the tested direction is irrelevant; see *One-sided test versus two-sided test* below for details.

Another concept important for hypothesis testing is that of a  $p$ -value or observed level of significance.  $P$ -value is a probability of obtaining a test statistic as extreme or more extreme as the one observed in a sample assuming the null hypothesis is true. It can also be viewed as the smallest level of  $\alpha$  that leads to the rejection of the null hypothesis. For example, if the  $p$ -value is less than 0.05, a test is considered to reject the null hypothesis at the 5% significance level.

For more information about hypothesis testing, see, for example, [Casella and Berger \(2002\)](#).

Next we review concepts specific to PSS analysis.

## Components of PSS analysis

The general goal of PSS analysis is to help plan a study such that the chosen statistical method has high power to detect an effect of interest if the effect exists. For example, PSS analysis is commonly used to determine the size of the sample needed for the chosen statistical test to have adequate power to detect an effect of a specified magnitude at a prespecified significance level given fixed values of other study parameters. We will use the phrase “detect an effect” to generally mean that the collected data will support the alternative hypothesis. For example, detecting an effect may be detecting that the means of two groups differ, or that there is an association between the probability of a disease and an exposure factor, or that there is a nonzero correlation between two measurements.

The general goal of PSS analysis can be achieved in several ways. You can

- compute sample size directly given specified significance level, power, effect size, and other study parameters;
- evaluate the power of a study for a range of sample sizes or effect sizes for a given significance level and fixed values of other study parameters;
- evaluate the magnitudes of an effect that can be detected with reasonable power for specific sample sizes given a significance level and other study parameters;
- evaluate the sensitivity of the power or sample-size requirements to various study parameters.

The main components of PSS analysis are

- study design;
- statistical method;
- significance level,  $\alpha$ ;
- power,  $1 - \beta$ ;

- a magnitude of an effect of interest or clinically meaningful difference, often expressed as an **effect size**,  $\delta$ ;
- sample size,  $N$ .

Below we describe each of the main components of PSS analysis in more detail.

### Study design

A well-designed statistical study has a carefully chosen study design and a clearly specified research objective that can be formulated as a statistical hypothesis. A study can be observational, where subjects are followed in time, such as a cross-sectional study, or it can be experimental, where subjects are assigned a certain procedure or treatment, such as a randomized, controlled clinical trial. A study can involve one, two, or more samples. A study can be prospective, where the outcomes are observed given the exposures, such as a cohort study, or it can be retrospective, where the exposures are observed given the outcomes, such as a case–control study. A study can also use matching, where subjects are grouped based on selected characteristics such as age or race. A common example of matching is a paired study, consisting of pairs of observations that share selected characteristics.

### Statistical method

A well-designed statistical study also has well-defined methods of analysis to be used to evaluate the objective of interest. For example, a comparison of two independent populations may involve an independent two-sample  $t$  test of means or a two-sample  $\chi^2$  test of variances, and so on. PSS computations are specific to the chosen statistical method and design. For example, the power of a balanced- or equal-allocation design is typically higher than the power of the corresponding unbalanced design.

### Significance level

A significance level  $\alpha$  is an upper bound for the probability of a type I error. With a slight abuse of terminology and notation, we will use the terms “significance level” and “type-I-error probability” interchangeably, and we will also use  $\alpha$  to denote the probability of a type I error. When the two are different, such as for tests with discrete sampling distributions of test statistics, we will make a distinction between them. In other words, unless stated otherwise, we will assume a size- $\alpha$  test, for which  $\Pr(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$  for any  $\theta$  in the null space, as opposed to a level- $\alpha$  test, for which  $\Pr(\text{reject } H_0 | H_0 \text{ is true}) \leq \alpha$  for any  $\theta$  in the null space.

As we mentioned earlier, researchers typically set the significance level to a small value such as 0.01 or 0.05 to protect the null hypothesis, which usually represents a state for which an incorrect decision is more costly.

Power is an increasing function of the significance level.

### Power

The power of a test is the probability of correctly rejecting the null hypothesis when the null hypothesis is false. That is,  $\pi = 1 - \beta = \Pr(\text{reject } H_0 | H_0 \text{ is false})$ . Increasing the power of a test decreases the probability of a type II error, so a test with high power is preferred. Common choices for power are 90% and 80%, depending on the study objective.

We consider prospective power, which is the power of a future study.

## Clinically meaningful difference and effect size

Clinically meaningful difference and effect size represent the magnitude of an effect of interest. In the context of PSS analysis, they represent the magnitude of the effect of interest to be detected by a test with a specified power. They can be viewed as a measure of how far the alternative hypothesis is from the null hypothesis. Their values typically represent the smallest effect that is of clinical significance or the hypothesized population effect size.

The interpretation of “clinically meaningful” is determined by the researcher and will usually vary from study to study. For example, in clinical trials, if no prior knowledge is available about the performance of the considered clinical procedure, then a standardized effect size (adjusted for standard deviation) between 0.25 and 0.5 may be considered clinically meaningful.

The definition of effect size is specific to the study design, analysis endpoint, and employed statistical model and test. For example, for a comparison of two independent proportions, an effect size may be defined as the difference between two proportions, the ratio of the two proportions, or the odds ratio. Effect sizes also vary in magnitude across studies: a treatment effect of 1% corresponding to an increase in mortality may be clinically meaningful, whereas a treatment effect of 10% corresponding to a decrease in a circumference of an ankle affected by edema may be of little importance. Effect size is usually defined in such a way that power is an increasing function of it (or its absolute value).

More generally, in PSS analysis, effect size summarizes the disparity between the alternative and null sampling distributions (sampling distributions under the alternative hypothesis and the null hypothesis, respectively) of a test statistic. The larger the overlap between the two distributions, the smaller the effect size and the more difficult it is to reject the null hypothesis, and thus there is less power to detect an effect.

For example, consider a  $z$  test for a comparison of a mean  $\mu$  with 0 from a population with a known standard deviation  $\sigma$ . The null hypothesis is  $H_0: \mu = 0$ , and the alternative hypothesis is  $H_a: \mu \neq 0$ . The test statistic is a sample mean or sample average. It has a normal distribution with mean 0 and standard deviation  $\sigma$  as its null sampling distribution, and it has a normal distribution with mean  $\mu$  different from 0 and standard deviation  $\sigma$  as its alternative sampling distribution. The overlap between these distributions is determined by the mean difference  $\mu - 0 = \mu$  and standard deviation  $\sigma$ . The larger  $\mu$  or, more precisely, the larger its absolute value, the larger the difference between the two populations, and thus the smaller the overlap and the higher the power to detect the differences  $\mu$ . The larger the standard deviation  $\sigma$ , the more overlap between the two distributions and the lower the power to detect the difference. Instead of being viewed as a function of  $\mu$  and  $\sigma$ , power can be viewed as a function of their combination expressed as the standardized difference  $\delta = (\mu - 0)/\sigma$ . Then, the larger  $|\delta|$ , the larger the power; the smaller  $|\delta|$ , the smaller the power. The effect size is then the standardized difference  $\delta$ .

To read more about effect sizes in Stata, see [R] [esize](#), although PSS analysis may sometimes use different definitions of an effect size.

## Sample size

Sample size is usually the main component of interest in PSS analysis. The sample size required to successfully achieve the objective of a study is determined given a specified significance level, power, effect size, and other study parameters. The larger the significance level, the smaller the sample size, with everything else being equal. The higher the power, the larger the sample size. The larger the effect size, the smaller the sample size.

When you compute sample size, the actual power (power corresponding to the obtained sample size) will most likely be different from the power you requested, because sample size is an integer. In the computation, the resulting fractional sample size that corresponds to the requested power is

usually rounded to the nearest integer. To be conservative, to ensure that the actual power is at least as large as the requested power, the sample size is rounded up. For multiple-sample designs, fractional sample sizes may arise when you specify sample size to compute power or effect size. For example, to accommodate an odd total sample size of, say, 51 in a balanced two-sample design, each individual sample size must be 25.5. To be conservative, sample sizes are rounded down on input. The actual sample sizes in our example would be 25, 25, and 50. See [Fractional sample sizes in \[PSS\] unbalanced designs](#) for details about sample-size rounding.

For multiple samples, the allocation of subjects between groups also affects power. A balanced- or equal-allocation design, a design with equal numbers of subjects in each sample or group, generally has higher power than the corresponding unbalanced- or unequal-allocation design, a design with different numbers of subjects in each sample or group.

### One-sided test versus two-sided test

Among other things that affect power is whether the employed test is directional (upper or lower one sided) or nondirectional (two sided). One-sided or one-tailed tests are more powerful than the corresponding two-sided or two-tailed tests. It may be tempting to choose a one-sided test over a two-sided test based on this fact. Despite having higher power, one-sided tests are generally not as common as two-sided tests. The direction of the effect, whether the effect is larger or smaller than a hypothesized value, is unknown in many applications, which requires the use of a two-sided test. The use of a one-sided test in applications in which the direction of the effect may be known is still controversial. The use of a one-sided test precludes the possibility of detecting an effect in the opposite direction, which may be undesirable in some studies. You should exercise caution when you decide to use a one-sided test, because you will not be able to rule out the effect in the opposite direction, if one were to happen. The results from a two-sided test have stronger justification.

### Another consideration: Dropout

During the planning stage of a study, another important consideration is whether the data collection effort may result in missing observations. In clinical studies, the common term for this is dropout, when subjects fail to complete the study for reasons unrelated to study objectives.

If dropout is anticipated, its rate must be taken into consideration when determining the required sample size or computing other parameters. For example, if subjects are anticipated to drop out from a study with a rate of  $R_d$ , an ad hoc way to inflate the estimated sample size  $n$  is as follows:  $n_d = n/(1 - R_d)^2$ . Similarly, the input sample size must be adjusted as  $n = n_d(1 - R_d)^2$ , where  $n_d$  is the anticipated sample size.

### Survival data

The prominent feature of survival data is that the outcome is the time from an origin to the occurrence of a given event (failure), often referred to as the analysis time. Analyses of such data use the information from all subjects in a study, both those who experience an event by the end of the study and those who do not. However, inference about the survival experience of subjects is based on the event times and therefore depends on the number of events observed in a study. Indeed, if none of the subjects fails in a study, then the survival rate cannot be estimated and survivor functions of subjects from different groups cannot be compared. Therefore, power depends on the number of events observed in a study and not directly on the number of subjects recruited to the study. As a result, to obtain the estimate of the required number of subjects, the probability that a subject experiences an event during the course of the study needs to be estimated in addition to the required number of events. This distinguishes sample-size determination for survival studies from that for other studies in which the endpoint is not measured as a time to failure.

All the above leads us to consider the following two types of survival studies. The first type (a type I study) is a study in which all subjects experience an event by the end of the study (no censoring), and the second type (a type II study) is a study that terminates after a fixed period regardless of whether all subjects experienced an event by that time. For a type II study, subjects who did not experience an event at the end of the study are known to be right-censored. For a type I study, when all subjects fail by the end of the study, the estimate of the probability of a failure in a study is one and the required number of subjects is equal to the required number of failures. For a type II study, the probability of a failure needs to be estimated and therefore various aspects that affect this probability (and usually do not come into play at the analysis stage) must be taken into account for the computation of the sample size.

Under the assumption of random censoring (Lachin 2011, 431; Lawless 2003, 52; Chow and Liu 2014, 391), the type of censoring pattern is irrelevant to the analysis of survival data in which the goal is to make inferences about the survival distribution of subjects. It becomes important, however, for sample-size determination because the probability that a subject experiences an event in a study depends on the censoring distribution. We consider the following two types of random censoring: administrative censoring and loss to follow-up.

Under administrative censoring, a subject is known to have experienced either of the two outcomes at the end of a study: survival or failure. The probability of a subject failing in a study depends on the duration of the study. Often in practice, subjects may withdraw from a study, say, because of severe side effects from a treatment or may be lost to follow-up because of moving to a different location. Here the information about the outcome that subject would have experienced at the end of the study had he completed the course of the study is unavailable, and the probability of experiencing an event by the end of the study is affected by the process governing withdrawal of subjects from the study. In the literature, this type of censoring is often referred to as subject loss to follow-up, subject withdrawal, or sometimes subject dropout (Freedman 1982, Machin and Campbell 2005). Generally, great care must be taken when using this terminology because it may have slightly different meanings in different contexts. `power logrank` and `power cox` apply a conservative adjustment to the estimate of the sample size for withdrawal. `power exponential` assumes that losses to follow-up are exponentially distributed.

Another important component of sample-size and power determination that affects the estimate of the probability of a failure is the pattern of accrual of subjects into the study. The duration of a study is often divided into two phases: an accrual phase, during which subjects are recruited to the study, and a follow-up phase, during which subjects are followed up until the end of the study and no new subjects enter the study. For a fixed-duration study, fast accrual increases the average analysis time (average follow-up time) and increases the chance of a subject failing in a study, whereas slow accrual decreases the average analysis time and consequently decreases this probability. `power logrank` and `power exponential` provide facilities to take into account uniform accrual, and for `power exponential` only, truncated exponential accrual.

All sample-size formulas used by `power`'s survival methods rely on the proportional-hazards assumption, that is, the assumption that the hazard ratio does not depend on time. See the documentation entry of each subcommand for the additional assumptions imposed by the methods it uses. In the case when the proportional-hazards assumption is suspect, or in the presence of other complexities associated with the nature of the trial (for example, lagged effect of a treatment, more than two treatment groups, clustered data) and with the behavior of participants (for example, noncompliance of subjects with the assigned treatment, competing risks), one may consider obtaining required sample size or power by simulation. Feiveson (2002) demonstrates an example of such simulation for clustered survival data. Also see Royston (2012) and Crowther and Lambert (2012) for ways of simulating complicated survival data. Barthel et al. (2006); Barthel, Royston, and Babiker (2005);

Royston and Babiker (2002); Barthel, Royston, and Parmar (2009); and Royston and Barthel (2010) present sample-size and power computation for multiarm trials under more flexible design conditions.

## Sensitivity analysis

Due to limited resources, it may not always be feasible to conduct a study under the original ideal specification. In this case, you may vary study parameters to find an appropriate balance between the desired detectable effect, sample size, available resources, and an objective of the study. For example, a researcher may decide to increase the detectable effect size to decrease the required sample size, or, rarely, to lower the desired power of the test. In some situations, it may not be possible to reduce the required sample size, in which case more resources must be acquired before the study can be conducted.

Power is a complicated function of all the components we described in the previous section—none of the components can be viewed in isolation. For this reason, it is important to perform sensitivity analysis, which investigates power for various specifications of study parameters, and refine the sample-size requirements based on the findings prior to conducting a study. Tables of power values (see [PSS] [power, table](#)) and graphs of power curves (see [PSS] [power, graph](#)) may be useful for this purpose.

## An example of PSS analysis in Stata

Consider a study of math scores from the SAT exam. Investigators would like to test whether a new coaching program increases the average SAT math score by 20 points compared with the national average in a given year of 514. They do not anticipate the standard deviation of the scores to be larger than the national value of 117. Investigators are planning to test the differences between scores by using a one-sample  $t$  test. Prior to conducting the study, investigators would like to estimate the sample size required to detect the anticipated difference by using a 5%-level two-sided test with 90% power. We can use the `power onemean` command to estimate the sample size for this study; see [PSS] [power onemean](#) for more examples.

Below we demonstrate PSS analysis of this example interactively, by typing the commands; see [PSS] [GUI](#) for point-and-click analysis of this example.

We specify the reference or null mean value of 514 and the comparison or alternative value of 534 as command arguments following the command name. The values of standard deviation and power are specified in the corresponding `sd()` and `power()` options. `power onemean` assumes a 5%-level two-sided test, so we do not need to specify any additional options.

```
. power onemean 514 534, sd(117) power(0.9)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =    0.1709
      m0 =    514.0000
      ma =    534.0000
      sd =    117.0000
Estimated sample size:
      N =        362
```

The estimated required sample size is 362.

Investigators do not have enough resources to enroll that many subjects. They would like to estimate the power corresponding to a smaller sample of 300 subjects. To compute power, we replace the `power(0.9)` option with the `n(300)` option in the above command.

```
. power onemean 514 534, sd(117) n(300)
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha =    0.0500
      N =      300
    delta =    0.1709
     m0 =    514.0000
     ma =    534.0000
     sd =    117.0000
Estimated power:
    power =    0.8392
```

For a smaller sample of 300 subjects, the power decreases to 84%.

Investigators would also like to estimate the minimum detectable difference between the scores given a sample of 300 subjects and a power of 90%. To compute the standardized difference between the scores, or effect size, we specify both the power in the `power()` option and the sample size in the `n()` option.

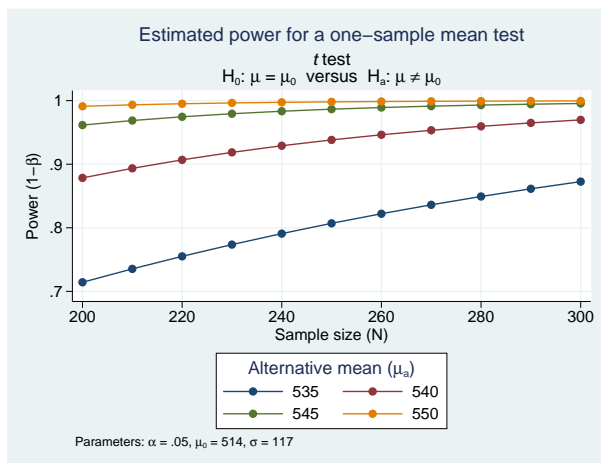
```
. power onemean 514, sd(117) power(0.9) n(300)
Performing iteration ...
Estimated target mean for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0; ma > m0
Study parameters:
    alpha =    0.0500
    power =    0.9000
      N =      300
     m0 =    514.0000
     sd =    117.0000
Estimated effect size and target mean:
    delta =    0.1878
     ma =    535.9671
```

The minimum detectable standardized difference given the requested power and sample size is 0.19, which corresponds to an average math score of roughly 536 and a difference between the scores of 22.

Continuing their analysis, investigators want to assess the impact of different sample sizes and score differences on power. They wish to estimate power for a range of alternative mean scores between 530 and 550 with an increment of 5 and a range of sample sizes between 200 and 300 with an increment of 10. They would like to see results on a graph.

We specify the range of alternative means as a *numlist* (see [U] 11.1.8 *numlist*) in parentheses as the second command argument. We specify the range of sample sizes as a *numlist* in the `n()` option. We request a graph by specifying the `graph` option.

```
. power onemean 514 (535(5)550), sd(117) n(200(10)300) graph
```



The default graph plots the estimated power on the  $y$  axis and the requested sample size on the  $x$  axis. A separate curve is plotted for each of the specified alternative means. Power increases as the sample size increases or as the alternative mean increases. For example, for a sample of 220 subjects and an alternative mean of 535, the power is approximately 75%; and for an alternative mean of 550, the power is nearly 1. For a sample of 300 and an alternative mean of 535, the power increases to 87%. Investigators may now determine a combination of an alternative mean and a sample size that would satisfy their study objective and available resources.

If desired, we can also display the estimated power values in a table by additionally specifying the `table` option:

```
. power onemean 514 (530(5)550), sd(117) n(200(10)300) graph table
(output omitted)
```

The `power` command performs PSS analysis for a number of hypothesis tests for continuous, binary, and survival outcomes; see [PSS] [power](#) and method-specific entries for more examples. Also, in the absence of readily available PSS methods, consider performing PSS analysis by simulation; see, for example, [Feiveson \(2002\)](#) and [Hooper \(2013\)](#) for examples of how you can do this in Stata. You can also add your own methods to the `power` command as described in `help power userwritten`; `power userwritten` is not part of official Stata, because it is under active development.

## Video example

[A conceptual introduction to power and sample-size calculations](#)

## References

- Agresti, A. 2013. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. 2002. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *Journal of the American Medical Association* 288: 2981–2997.
- Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley.



- Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.
- Barthel, F. M.-S., A. G. Babiker, P. Royston, and M. K. B. Parmar. 2006. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Statistics in Medicine* 25: 2521–2542.
- Barthel, F. M.-S., P. Royston, and A. G. Babiker. 2005. A menu-driven facility for complex sample size calculation in randomized controlled trials with a survival or a binary outcome: Update. *Stata Journal* 5: 123–129.
- Barthel, F. M.-S., P. Royston, and M. K. B. Parmar. 2009. A menu-driven facility for sample-size calculation in novel multiarm, multistage randomized controlled trials with a time-to-event outcome. *Stata Journal* 9: 505–523.
- Casagrande, J. T., M. C. Pike, and P. G. Smith. 1978a. The power function of the “exact” test for comparing two binomial distributions. *Journal of the Royal Statistical Society, Series C* 27: 176–180.
- . 1978b. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* 34: 483–486.
- Casella, G., and R. L. Berger. 2002. *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury.
- Chernick, M. R., and C. Y. Liu. 2002. The saw-toothed behavior of power versus sample size and software solutions: Single binomial proportion using exact methods. *American Statistician* 56: 149–155.
- Chow, S.-C., and J.-P. Liu. 2014. *Design and Analysis of Clinical Trials: Concepts and Methodologies*. 3rd ed. Hoboken, NJ: Wiley.
- Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- . 1992. A power primer. *Psychological Bulletin* 112: 155–159.
- Collett, D. 2003. *Modelling Survival Data in Medical Research*. 2nd ed. London: Chapman & Hall/CRC.
- Connor, R. J. 1987. Sample size for testing differences in proportions for the paired-sample design. *Biometrics* 43: 207–211.
- Cox, D. R. 1972. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34: 187–220.
- Cox, D. R., and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman & Hall/CRC.
- Crowther, M. J., and P. C. Lambert. 2012. Simulating complex survival data. *Stata Journal* 12: 674–687.
- Davis, B. R., J. A. Cutler, D. J. Gordon, C. D. Furberg, J. T. Wright, Jr, W. C.ushman, R. H. Grimm, J. LaRosa, P. K. Whelton, H. M. Perry, M. H. Alderman, C. E. Ford, S. Oparil, C. Francis, M. Proschan, S. Pressel, H. R. Black, and C. M. Hawkins, for the ALLHAT Research Group. 1996. Rationale and design for the antihypertensive and lipid lowering treatment to prevent heart attack trial (ALLHAT). *American Journal of Hypertension* 9: 342–360.
- Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw-Hill.
- Feiveson, A. H. 2002. Power by simulation. *Stata Journal* 2: 107–124.
- Fisher, R. A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10: 507–521.
- Flleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley.
- Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.
- Graybill, F. A. 1961. *An Introduction to Linear Statistical Models, Vol. 1*. New York: McGraw-Hill.
- Guenther, W. C. 1977. Desk calculation of probabilities for the distribution of the sample correlation coefficient. *American Statistician* 31: 45–48.
- Harrison, D. A., and A. R. Brady. 2004. Sample size and power calculations using the noncentral t-distribution. *Stata Journal* 4: 142–153.
- Hemming, K., and J. Marsh. 2013. A menu-driven facility for sample-size calculations in cluster randomized controlled trials. *Stata Journal* 13: 114–135.
- Hooper, R. 2013. Versatile sample-size calculation using simulation. *Stata Journal* 13: 21–38.

- Hosmer, D. W., Jr., S. A. Lemeshow, and S. May. 2008. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. 2nd ed. New York: Wiley.
- Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.
- Howell, D. C. 2002. *Statistical Methods for Psychology*. 5th ed. Belmont, CA: Wadsworth.
- Hsieh, F. Y. 1992. Comparing sample size formulae for trials with unbalanced allocation using the logrank test. *Statistics in Medicine* 11: 1091–1098.
- Irwin, J. O. 1935. Tests of significance for differences between percentages based on small numbers. *Metron* 12: 83–94.
- Julious, S. A. 2010. *Sample Sizes for Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC.
- Kahn, H. A., and C. T. Sempos. 1989. *Statistical Methods in Epidemiology*. New York: Oxford University Press.
- Klein, J. P., and M. L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer.
- Krishnamoorthy, K., and J. Peng. 2007. Some properties of the exact and score methods for binomial proportion and sample size calculation. *Communications in Statistics—Simulation and Computation* 36: 1171–1186.
- Kunz, C. U., and M. Kieser. 2011. Simon’s minimax and optimal and Jung’s admissible two-stage designs with or without curtailment. *Stata Journal* 11: 240–254.
- Lachin, J. M. 1981. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 2: 93–113.
- . 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.
- Lawless, J. F. 2003. *Statistical Models and Methods for Lifetime Data*. 2nd ed. New York: Wiley.
- Lenth, R. V. 2001. Some practical guidelines for effective sample size determination. *American Statistician* 55: 187–193.
- Machin, D. 2004. On the evolution of statistical methods as applied to clinical trials. *Journal of Internal Medicine* 255: 521–528.
- Machin, D., and M. J. Campbell. 2005. *Design of Studies for Medical Research*. Chichester, UK: Wiley.
- Machin, D., M. J. Campbell, S. B. Tan, and S. H. Tan. 2009. *Sample Size Tables for Clinical Studies*. 3rd ed. Chichester, UK: Wiley–Blackwell.
- Marubini, E., and M. G. Valsecchi. 1997. *Analysing Survival Data from Clinical Trials and Observational Studies*. Chichester, UK: Wiley.
- McNemar, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12: 153–157.
- Newson, R. B. 2004. Generalized power calculations for generalized linear models and more. *Stata Journal* 4: 379–401.
- Pagano, M., and K. Gauvreau. 2000. *Principles of Biostatistics*. 2nd ed. Belmont, CA: Duxbury.
- Royston, P. 2012. Tools to simulate realistic censored survival-time distributions. *Stata Journal* 12: 639–654.
- Royston, P., and A. G. Babiker. 2002. A menu-driven facility for complex sample size calculation in randomized controlled trials with a survival or a binary outcome. *Stata Journal* 2: 151–163.
- Royston, P., and F. M.-S. Barthel. 2010. Projection of power and events in clinical trials with a time-to-event outcome. *Stata Journal* 10: 386–394.
- Ryan, T. P. 2013. *Sample Size Determination and Power*. Hoboken, NJ: Wiley.
- Saunders, C. L., D. T. Bishop, and J. H. Barrett. 2003. Sample size calculations for main effects and interactions in case-control studies using Stata’s `nchi2` and `npnchi2` functions. *Stata Journal* 3: 47–56.
- Schoenfeld, D. A., and J. R. Richter. 1982. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* 38: 163–170.
- Schork, M. A., and G. W. Williams. 1980. Number of observations required for the comparison of two correlated proportions. *Communications in Statistics—Simulation and Computation* 9: 349–357.
- Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods*. 8th ed. Ames, IA: Iowa State University Press.

- Tamhane, A. C., and D. D. Dunlop. 2000. *Statistics and Data Analysis: From Elementary to Intermediate*. Upper Saddle River, NJ: Prentice Hall.
- Væth, M., and E. Skovlund. 2004. A simple approach to power and sample size calculations in logistic regression and Cox regression models. *Statistics in Medicine* 23: 1781–1792.
- van Belle, G., L. D. Fisher, P. J. Heagerty, and T. S. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences*. 2nd ed. New York: Wiley.
- Wickramaratne, P. J. 1995. Sample size determination in epidemiologic studies. *Statistical Methods in Medical Research* 4: 311–337.
- Winer, B. J., D. R. Brown, and K. M. Michels. 1991. *Statistical Principles in Experimental Design*. 3rd ed. New York: McGraw–Hill.
- Wittes, J. 2002. Sample size calculations for randomized control trials. *Epidemiologic Reviews* 24: 39–53.

## Also see

- [PSS] **GUI** — Graphical user interface for power and sample-size analysis
- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **unbalanced designs** — Specifications for unbalanced designs
- [PSS] **Glossary**

# Title

## GUI — Graphical user interface for power and sample-size analysis

[Description](#)   [Menu](#)   [Remarks and examples](#)   [Also see](#)

## Description

This entry describes the graphical user interface (GUI) for the `power` command. See [\[PSS\] power](#) for a general introduction to the `power` command.

## Menu

Statistics > Power and sample size

## Remarks and examples

Remarks are presented under the following headings:

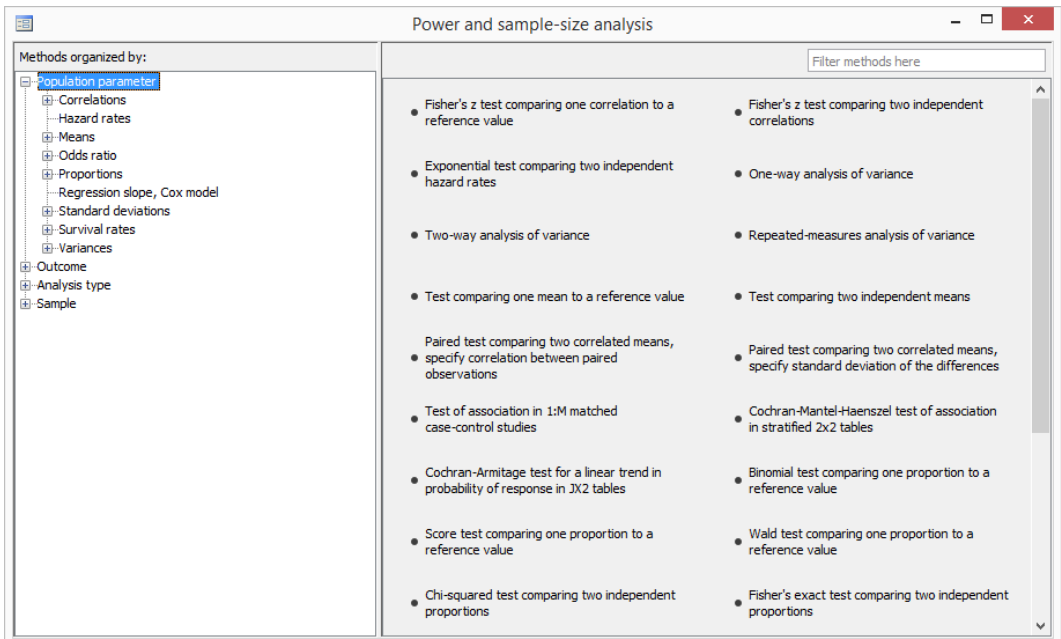
*[PSS Control Panel](#)*

*[Example with PSS Control Panel](#)*

## PSS Control Panel

You can perform PSS analysis interactively by typing the `power` command or by using a point-and-click GUI available via the PSS Control Panel.

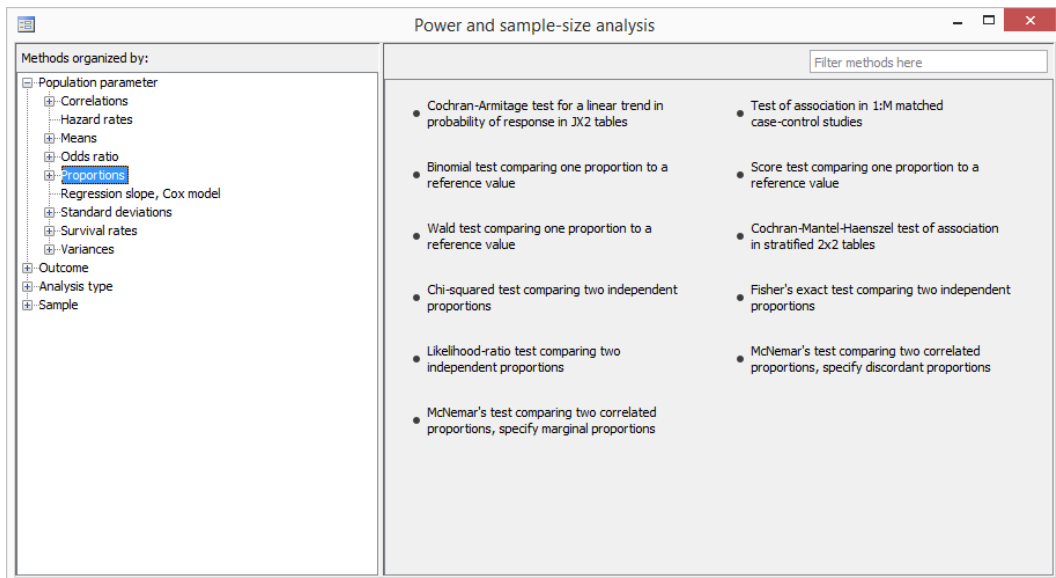
The PSS Control Panel can be accessed by selecting **Statistics > Power and sample size** from the Stata menu. It includes a tree-view organization of the PSS methods.



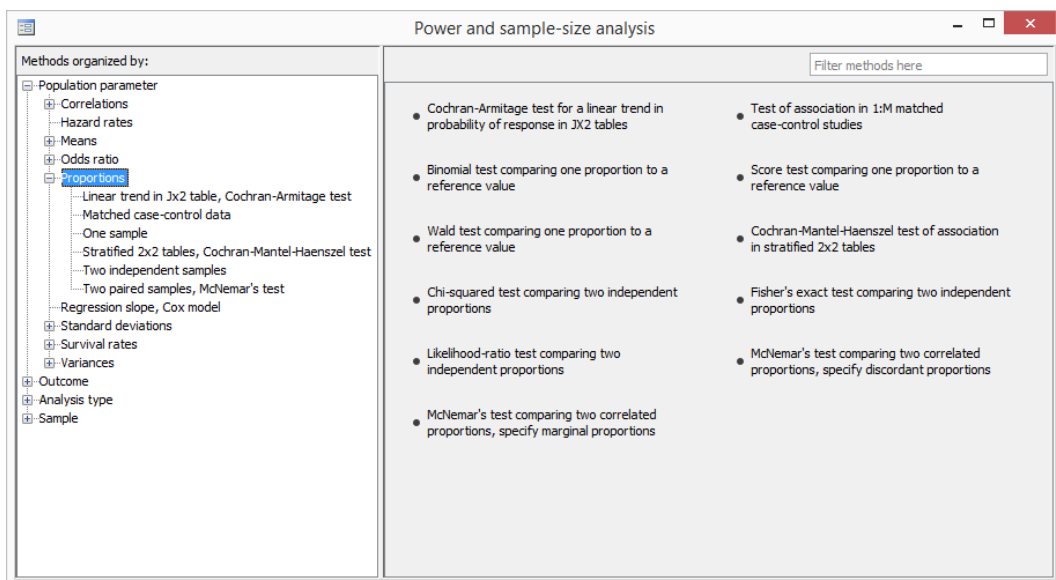
The left pane organizes the methods, and the right pane displays the methods corresponding to the selection in the left pane. On the left, the methods are organized by the type of population parameter, such as mean or proportion; the type of outcome, such as continuous or binary; the type of analysis, such as  $t$  test or  $\chi^2$  test; and the type of sample, such as one sample or two samples. You click on one of the methods shown in the right pane to launch the dialog box for that method.

By default, methods are organized by **Population parameter**. We can find the method we want to use by looking for it in the right pane, or we can narrow down the type of method we are looking for by selecting one of the expanded categories in the left pane.

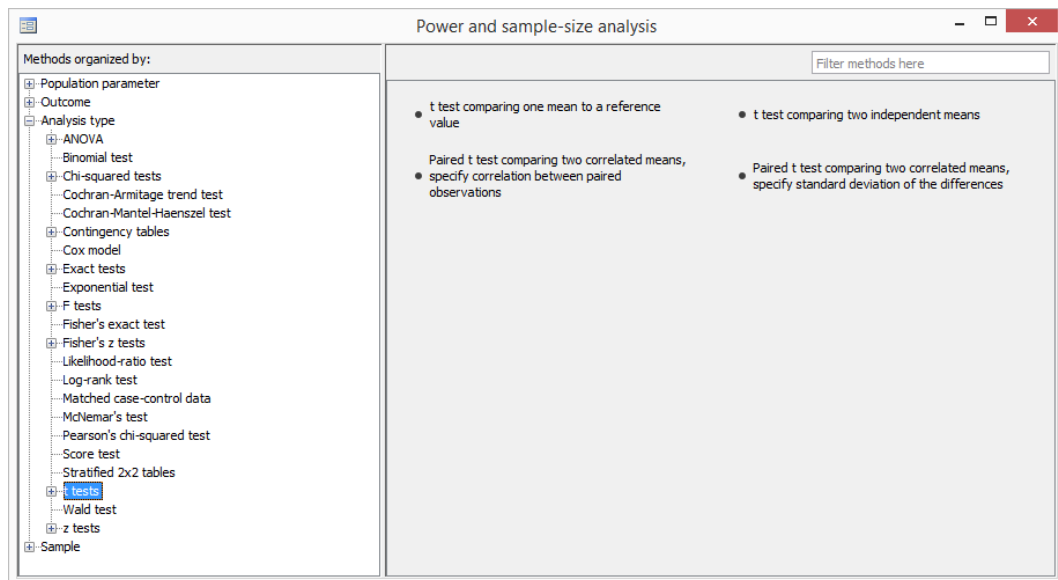
For example, if we are interested in proportions, we can click on **Proportions** within **Population parameter** to see all methods comparing proportions in the right pane.



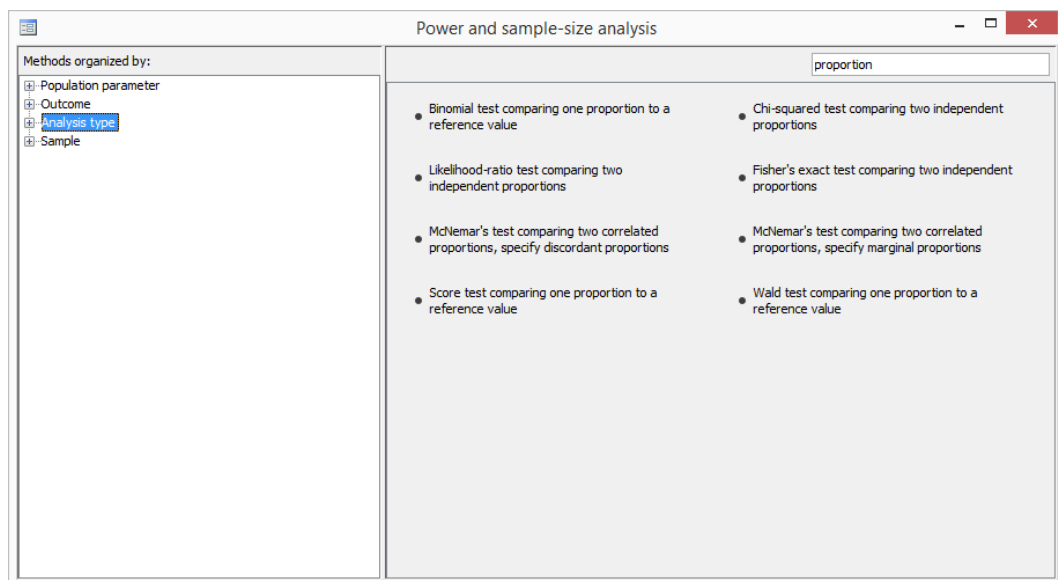
We can expand **Proportions** to further narrow down the choices by clicking on the symbol to the left of **Proportions**.



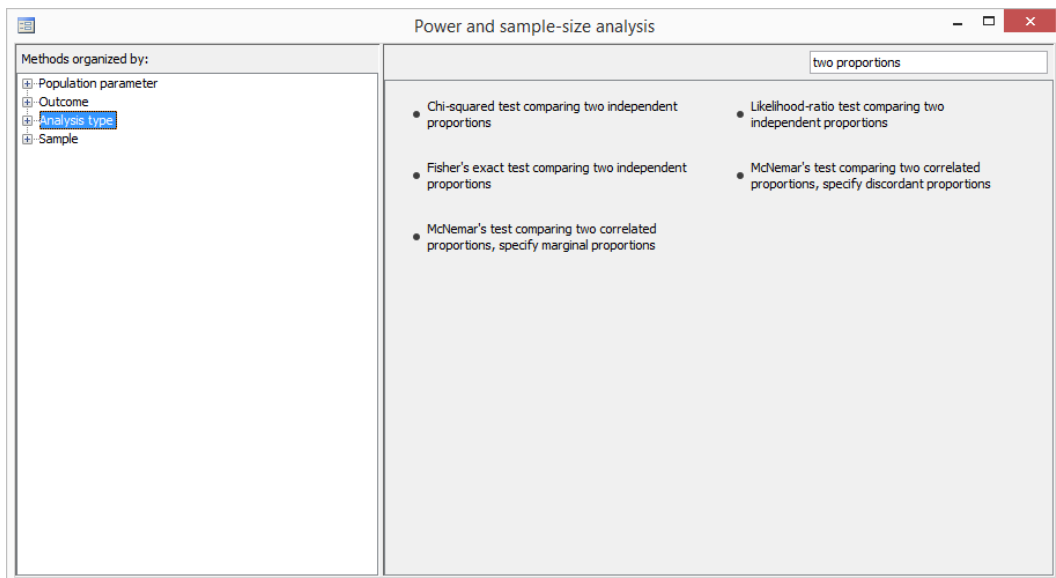
Or we can choose a method by the type of analysis by expanding **Analysis type** and selecting, for example, **t tests**:



We can also locate methods by searching the titles of methods. You specify the search string of interest in the *Filter* box at the top right of the PSS Control Panel. For example, if we type “proportion” in the *Filter* box while keeping the focus on **Analysis type**, only methods with a title containing “proportion” will be listed in the right pane.



We can specify multiple words in the *Filter* box, and only methods with all the specified words in their titles will appear. For example, if we type “two proportions”, only methods with the words “two” and “proportions” in their titles will be shown:



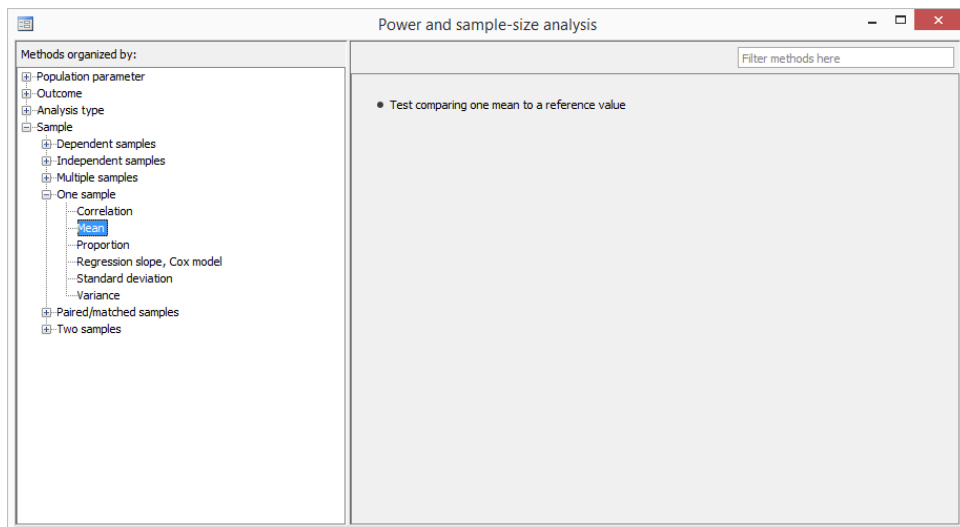
The search is performed within the group of methods selected by the choice in the left pane. In the above example, the search was done within **Analysis type**. When you select one of the top categories in the left pane, the same set of methods appears in the right pane but in the order determined by the specific organization. To search all methods, you can first select any of the four top categories, and you will get the same results but possibly in a slightly different order determined by the selected top-level category.



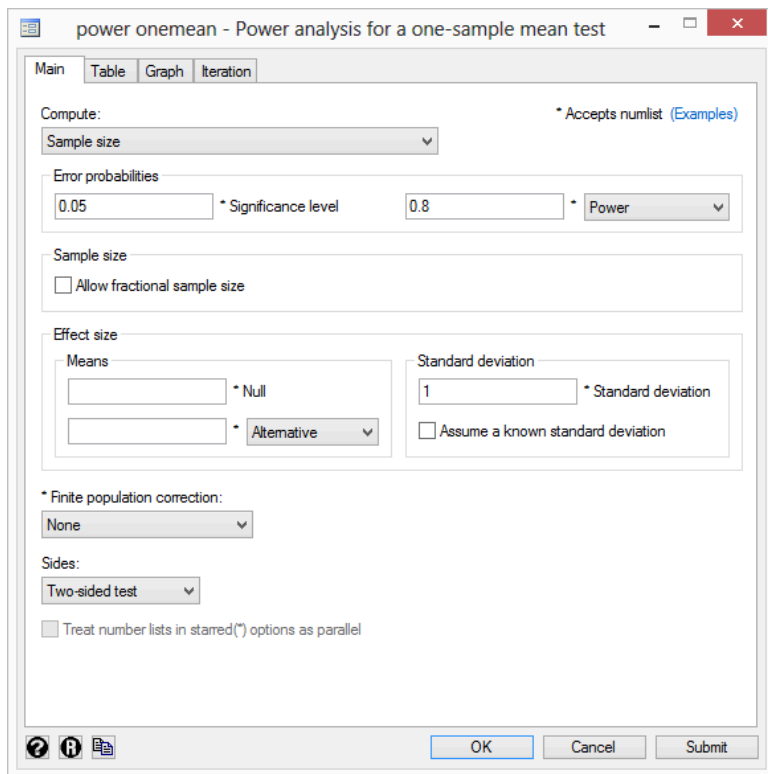
## Example with PSS Control Panel

In *An example of PSS analysis in Stata* in [PSS] [intro](#), we performed PSS analysis interactively by typing commands. We replicate the analysis by using the PSS Control Panel and dialog boxes.

We first launch the PSS Control Panel from the **Statistics > Power and sample size** menu. We then narrow down to the desired dialog box by first choosing **Sample** in the left pane, then choosing **One sample** within that, and then choosing **Mean**. In the right pane, we see **Test comparing one mean to a reference value**.



We invoke the dialog box by clicking on the method title in the right pane. The following appears:



Following the example from *An example of PSS analysis in Stata* in [PSS] [intro](#), we now compute sample size. The first step is to choose which parameter to compute. The *Compute* drop-down box specifies *Sample size*, so we leave it unchanged. The next step is to specify error probabilities. The default significance level is already set to our desired value of 0.05, so we leave it unchanged. We change power from the default value of 0.8 to 0.9. We then specify a null mean of 514, an alternative mean of 534, and a standard deviation of 117 in the *Effect size* group of options. We leave everything else unchanged and click on the **Submit** button to obtain results.

The following command is displayed in the Results window and executed:

```
. power onemean 514 534, sd(117) power(0.9)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =    0.1709
      m0 =    514.0000
      ma =    534.0000
      sd =    117.0000
Estimated sample size:
      N =        362
```

We can verify that the command and results are exactly the same as what we specified in [An example of PSS analysis in Stata](#) in [\[PSS\] intro](#).

Continuing our PSS analysis, we now want to compute power for a sample of 300 subjects. We return to the dialog box and select **Power** under *Compute*. The only thing we need to specify is the sample size of 300:

The following command is issued after we click on the **Submit** button:

```
. power onemean 514 534, sd(117) n(300)
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha =    0.0500
    N      =    300
    delta  =    0.1709
    m0     = 514.0000
    ma     = 534.0000
    sd     = 117.0000
Estimated power:
    power  =    0.8392
```

To compute effect size, we select **Effect size and target mean** under *Compute*. All the previously used values for power and sample size are preserved, so we do not need to specify anything additional.

We click on the **Submit** button and get the following:

```
. power onemean 514, sd(117) power(0.9) n(300)
Performing iteration ...
Estimated target mean for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0; ma > m0
Study parameters:
alpha = 0.0500
power = 0.9000
N = 300
m0 = 514.0000
sd = 117.0000
Estimated effect size and target mean:
delta = 0.1878
ma = 535.9671
```

To produce the graph from *An example of PSS analysis in Stata*, we first select **Power** under *Compute*. Then we specify the *numlists* for sample size and alternative mean in the respective edit boxes:

power onemean - Power analysis for a one-sample mean test

Main Table Graph Iteration

Compute: \* Accepts numlist (Examples)  
Power

Error probabilities  
0.05 \* Significance level

Sample size  
200(10)300 \* Sample size

Effect size

Means  
514 \* Null  
535(5)550 \* Alternative

Standard deviation  
117 \* Standard deviation  
 Assume a known standard deviation

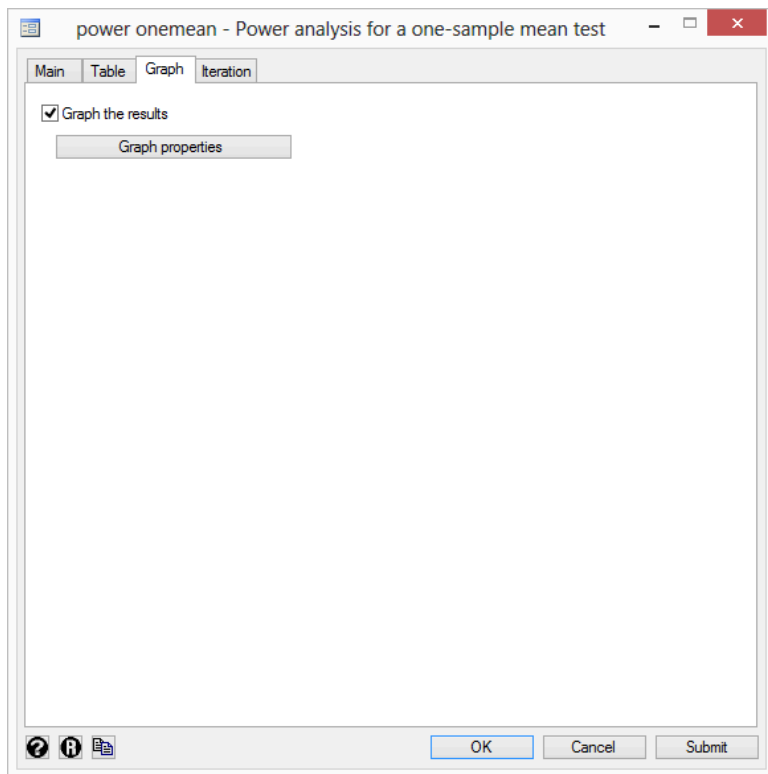
\* Finite population correction:  
None

Sides:  
Two-sided test

Treat number lists in starred(\*) options as parallel

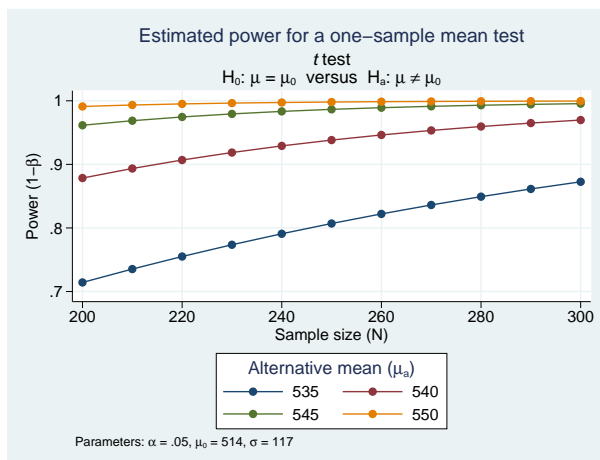
OK Cancel Submit

We also check the *Graph the results* box on the **Graph** tab:



We click on the **Submit** button and obtain the following command and graph:

```
. power onemean 514 (535(5)550), sd(117) n(200(10)300) graph
```



## Also see

[PSS] [power](#) — Power and sample-size analysis for hypothesis tests

[PSS] [intro](#) — Introduction to power and sample-size analysis

[PSS] [Glossary](#)



# Title

**power** — Power and sample-size analysis for hypothesis tests

Description

Remarks and examples

Also see

Menu

Stored results

Syntax

Methods and formulas

Options

References

## Description

The `power` command is useful for planning studies. It performs power and sample-size analysis for studies that use hypothesis testing to form inferences about population parameters. You can compute sample size given power and effect size, power given sample size and effect size, or the minimum detectable effect size and the corresponding target parameter given power and sample size. You can display results in a table ([PSS] **power, table**) and on a graph ([PSS] **power, graph**).

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

```
power method ... [ , power(numlist) power_options ... ]
```

*Compute power*

```
power method ... , n(numlist) [power_options ... ]
```

*Compute effect size and target parameter*

```
power method ... , n(numlist) power(numlist) [power_options ... ]
```

<i>method</i>	Description
One sample	
<a href="#">onemean</a>	One-sample mean test (one-sample $t$ test)
<a href="#">oneproportion</a>	One-sample proportion test
<a href="#">onecorrelation</a>	One-sample correlation test
<a href="#">onevariance</a>	One-sample variance test
Two independent samples	
<a href="#">twomeans</a>	Two-sample means test (two-sample $t$ test)
<a href="#">two proportions</a>	Two-sample proportions test
<a href="#">twocorrelations</a>	Two-sample correlations test
<a href="#">two variances</a>	Two-sample variances test
Two paired samples	
<a href="#">pairedmeans</a>	Paired-means test (paired $t$ test)
<a href="#">pairedproportions</a>	Paired-proportions test (McNemar's test)
Analysis of variance	
<a href="#">oneway</a>	One-way ANOVA
<a href="#">two way</a>	Two-way ANOVA
<a href="#">repeated</a>	Repeated-measures ANOVA
Contingency tables	
<a href="#">cmh</a>	Cochran–Mantel–Haenszel test (stratified $2 \times 2$ tables)
<a href="#">mcc</a>	Matched case–control studies
<a href="#">trend</a>	Cochran–Armitage trend test (linear trend in $J \times 2$ table)
Survival analysis	
<a href="#">cox</a>	Cox proportional hazards model
<a href="#">exponential</a>	Two-sample exponential test
<a href="#">logrank</a>	Log-rank test

---

<i>power_options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
* <u>n1</u> ( <i>numlist</i> )	sample size of the control group
* <u>n2</u> ( <i>numlist</i> )	sample size of the experimental group
* <u>nratio</u> ( <i>numlist</i> )	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<u>compute</u> ( <i>n1</i>   <i>n2</i> )	solve for $N1$ given $N2$ or for $N2$ given $N1$
<u>nfractional</u>	allow fractional sample sizes
<u>direction</u> ( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<u>saving</u> ( <i>filename</i> [ , <u>replace</u> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<u>init</u> (#)	initial value of the estimated parameter; default is method specific
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
[ <u>no</u> ] <u>log</u>	suppress or display iteration log
[ <u>no</u> ] <u>dots</u>	suppress or display iterations as dots
<u>notitle</u>	suppress the title

\* Starred options may be specified either as one number or as a list of values; see [U] [11.1.8 numlist](#).

Options `n1()`, `n2()`, `nratio()`, and `compute()` are available only for two-independent-samples methods.

Iteration options are available only with computations requiring iteration.

`notitle` does not appear in the dialog box.

## Options

### Main

- `alpha(numlist)` sets the significance level of the test. The default is `alpha(0.05)`.
- `power(numlist)` sets the power of the test. The default is `power(0.8)`. If `beta()` is specified, this value is set to be  $1 - \text{beta}()$ . Only one of `power()` or `beta()` may be specified.
- `beta(numlist)` sets the probability of a type II error of the test. The default is `beta(0.2)`. If `power()` is specified, this value is set to be  $1 - \text{power}()$ . Only one of `beta()` or `power()` may be specified.
- `n(numlist)` specifies the total number of subjects in the study to be used for power or effect-size determination. If `n()` is specified, the power is computed. If `n()` and `power()` or `beta()` are specified, the minimum effect size that is likely to be detected in a study is computed.
- `n1(numlist)` specifies the number of subjects in the control group to be used for power or effect-size determination.
- `n2(numlist)` specifies the number of subjects in the experimental group to be used for power or effect-size determination.
- `nratio(numlist)` specifies the sample-size ratio of the experimental group relative to the control group,  $N2/N1$ , for power or effect-size determination for two-sample tests. The default is `nratio(1)`, meaning equal allocation between the two groups.
- `compute(n1 | n2)` requests that the `power` command compute one of the group sample sizes given the other one instead of the total sample size for two-sample tests. To compute the control-group sample size, you must specify `compute(n1)` and the experimental-group sample size in `n2()`. Alternatively, to compute the experimental-group sample size, you must specify `compute(n2)` and the control-group sample size in `n1()`.
- `nfractional` specifies that fractional sample sizes be allowed. When this option is specified, fractional sample sizes are used in the intermediate computations and are also displayed in the output.
- Also see the description and the use of options `n()`, `n1()`, `n2()`, `nratio()`, and `compute()` for two-sample tests in [\[PSS\] unbalanced designs](#).
- `direction(upper | lower)` specifies the direction of the effect for effect-size determination. For most methods, the default is `direction(upper)`, which means that the postulated value of the parameter is larger than the hypothesized value. For survival methods, the default is `direction(lower)`, which means that the postulated value is smaller than the hypothesized value.
- `onesided` indicates a one-sided test. The default is two sided.
- `parallel` reports results in parallel over the list of numbers supplied to command arguments and options allowing *numlist*. By default, results are computed over all combinations of the number lists. If the specified number lists are of different sizes, the last value in each of the shorter lists will be used in the remaining computations.

### Table

- `notable`, `table`, and `table()` control whether or not results are displayed in a tabular format. `table` is implied if any number list contains more than one element. `notable` is implied with graphical output—when either the `graph` or the `graph()` option is specified. `table()` is used to produce custom tables. See [\[PSS\] power, table](#) for details.
- `saving(filename [, replace])` creates a Stata data file (`.dta` file) containing the table values with variable names corresponding to the displayed *columns*. `replace` specifies that *filename* be overwritten if it exists. `saving()` is only appropriate with tabular output.

## Graph

`graph` and `graph()` produce graphical output; see [PSS] [power](#), [graph](#) for details.

The following options control an iteration procedure used by the `power` command for solving nonlinear equations.

## Iteration

`init(#)` specifies an initial value for the estimated parameter. Each `power` method sets its own default value. See the documentation entry of the method for details.

`iterate(#)` specifies the maximum number of iterations for the Newton method. The default is `iterate(500)`.

`tolerance(#)` specifies the tolerance used to determine whether successive parameter estimates have converged. The default is `tolerance(1e-12)`. See *Convergence criteria* in [M-5] [solvenl\(\)](#) for details.

`ftolerance(#)` specifies the tolerance used to determine whether the proposed solution of a nonlinear equation is sufficiently close to 0 based on the squared Euclidean distance. The default is `ftolerance(1e-12)`. See *Convergence criteria* in [M-5] [solvenl\(\)](#) for details.

`log` and `nolog` specify whether an iteration log is to be displayed. The iteration log is suppressed by default. Only one of `log`, `nolog`, `dots`, or `nodots` may be specified.

`dots` and `nodots` specify whether a dot is to be displayed for each iteration. The iteration dots are suppressed by default. Only one of `dots`, `nodots`, `log`, or `nolog` may be specified.

The following option is available with `power` but is not shown in the dialog box:

`notitle` prevents the command title from displaying.

## Remarks and examples

Remarks are presented under the following headings:

*Using the power command*

*Specifying multiple values of study parameters*

*One-sample tests*

*Two-sample tests*

*Paired-sample tests*

*Analysis of variance models*

*Contingency tables*

*Survival analysis*

*Tables of results*

*Power curves*

This section describes how to perform power and sample-size analysis using the `power` command. For a software-free introduction to power and sample-size analysis, see [PSS] [intro](#).

## Using the power command

The `power` command computes sample size, power, or minimum detectable effect size and the corresponding target parameter for various hypothesis tests. You can also add your own methods to the `power` command as described in `help power userwritten`; `power userwritten` is not part of official Stata, because it is under active development.

All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

By default, the `power` command computes sample size for the default power of 0.8. You may change the value of power by specifying the `power()` option. Instead of power, you can specify the probability of a type II error in the `beta()` option.

To compute power, you must specify the sample size in the `n()` option.

To compute power or sample size, you must also specify a magnitude of the effect desired to be detected by a hypothesis test. `power`'s methods provide several ways in which an effect can be specified. For example, for a one-sample mean test, you can specify either the target mean or the difference between the target mean and a reference mean; see [PSS] [power onemean](#).

You can also compute the smallest magnitude of the effect or the minimum detectable effect size (MDES) and the corresponding target parameter that can be detected by a hypothesis test given power and sample size. To compute MDES, you must specify both the desired power in the `power()` option or the probability of a type II error in the `beta()` option and the sample size in the `n()` option. In addition to the effect size, `power` also reports the estimated value of the parameter of interest, such as the mean under the alternative hypothesis for a one-sample test or the experimental-group proportion for a two-sample test of independent proportions. By default, when the postulated value is larger than the hypothesized value, the `power` command assumes an effect in the upper direction, the `direction(upper)` option. You may request an estimate of the effect in the opposite, lower, direction by specifying the `direction(lower)` option.

For hypothesis tests comparing two independent samples, you can compute one of the group sizes given the other one instead of the total sample size. In this case, you must specify the label of the group size you want to compute in the `compute()` option and the value of the other group size in the respective `n#()` option. For example, if we wanted to find the size of the second group given the size of the first group, we would specify the combination of options `compute(n2)` and `n1(#)`.

A balanced design is assumed by default for two independent-sample tests, but you can request an unbalanced design. For example, you can specify the allocation ratio  $n_2/n_1$  between the two groups in the `nratio()` option or the individual group sizes in the `n1()` and `n2()` options. See [PSS] [unbalanced designs](#) for more details about various ways of specifying an unbalanced design.

For sample-size determination, the reported integer sample sizes may not correspond exactly to the specified power because of rounding. To obtain conservative results, the `power` command rounds up the sample size to the nearest integer so that the corresponding power is at least as large as the requested power. You can specify the `nfractional` option to obtain the corresponding fractional sample size.

Some of `power`'s computations require iteration. The defaults chosen for the iteration procedure should be sufficient for most situations. In a rare situation when you may want to modify the defaults, the `power` command provides options to control the iteration procedure. The most commonly used is the `init()` option for supplying an initial value of the estimated parameter. This option can be useful in situations where the computations are sensitive to the initial values. If you are performing computations for a large number of combinations of various study parameters, you may consider reducing the default maximum number of iterations of 500 in the `iterate()` option so that the command is not spending time on calculations in difficult-to-compute regions of the parameter space. By default, `power` suppresses the iteration log. If desired, you can specify the `log` option to display the iteration log or the `dots` option to display iterations as dots to monitor the progress of the iteration procedure.

The `power` command can produce results for one study scenario or for multiple study scenarios when multiple values of the parameters are specified; see [Specifying multiple values of study parameters](#) below for details.

For a single result, `power` displays results as text. For multiple results or if the `table` option is specified, `power` displays results in a table. You can also display multiple results on a graph by specifying the `graph` option. Graphical output suppresses the table of the results; use the `table` option to also see the tabular output. You can customize the default tables and graphs by specifying suboptions within the respective options `table()` and `graph()`; see [\[PSS\] power, table](#) and [\[PSS\] power, graph](#) for details.

You can also save the tabular output to a Stata dataset by using the `saving()` option.

## Specifying multiple values of study parameters

The `power` command can produce results for one study scenario or for multiple study scenarios when multiple values of the parameters are supplied to the supported options. The options that support multiple values specified as a *numlist* are marked with a star in the syntax diagram.

For example, the `power()` option supports multiple values. You can specify multiple powers as individual values, `power(0.8 0.85 0.9)`, or as a range of values, `power(0.8(0.05)0.9)`; see [\[U\] 11.1.8 numlist](#) for other specifications.

In addition to options, you may specify multiple values of command arguments, values specified after the command name. For example, let  $\#_1$  and  $\#_2$  be the first and the second command arguments in

```
. power method #1 #2, ...
```

If we want to specify multiple values for the command arguments, we must enclose these values in parentheses. For example,

```
. power method (1 2) (1 2 3), ...
```

or, more generally,

```
. power method (numlist) (numlist), ...
```

When multiple values are specified in multiple options or for multiple command arguments, the `power` command computes results for all possible combinations formed by the values from every option and command argument. In some cases, you may want to compute results in parallel for specific sets of values of the specified parameters. To request this, you can specify the `parallel` option. If the specified number lists are of varying sizes, *numlist* with the maximum size determines the number of final results produced by `power`. The last value from *numlist* of smaller sizes will be used in the subsequent computations.

For example,

```
. power method (1 2), power(0.8 0.9)
```

is equivalent to

```
. power method 1, power(0.8)
. power method 2, power(0.8)
. power method 1, power(0.9)
. power method 2, power(0.9)
```

When the `parallel` option is specified,

```
. power method (1 2), power(0.8 0.9) parallel
```

is equivalent to

```
. power method 1, power(0.8)
```

```
. power method 2, power(0.9)
```

When the `parallel` option is specified and `numlist` is of different sizes, the last value of the shorter `numlist` is used in the subsequent computations. For example,

```
. power method (1 2 3), power(0.8 0.9) parallel
```

is equivalent to

```
. power method 1, power(0.8)
```

```
. power method 2, power(0.9)
```

```
. power method 3, power(0.9)
```

## One-sample tests

The `power` command provides PSS computations for four one-sample tests. `power onemean` performs PSS analysis for a one-sample mean test; `power oneproportion` performs PSS analysis for a one-sample proportion test; `power onecorrelation` performs PSS analysis for a one-sample correlation test; and `power onevariance` performs PSS analysis for a one-sample variance test.

`power onemean` provides PSS computations for a one-sample  $t$  test assuming known or unknown population standard deviation. It also provides a way to adjust computations for a finite population sample. See [PSS] [power onemean](#).

`power oneproportion` provides PSS computations for a test that compares one proportion with a reference value. By default, the computations are based on a large-sample  $z$  test that uses the normal approximation of the distribution of the test statistic. You may choose between two large-sample tests: the score test or Wald test. You may also compute power for the small-sample binomial test by specifying the `test(binomial)` option. See [PSS] [power oneproportion](#).

`power onecorrelation` provides PSS computations for a test that compares one correlation with a reference value. The computations are based on a Fisher's  $z$  transformation of a correlation coefficient. See [PSS] [power onecorrelation](#).

`power onevariance` provides PSS computations for a test that compares one variance with a reference value. The computations are based on a  $\chi^2$  test of the ratio of the variance to its reference value. You can perform computations in the variance or standard deviation metric. See [PSS] [power onevariance](#).

All one-sample methods compute sample size given power and target parameter, power given sample size and target parameter, or MDES and the corresponding target parameter given power and sample size.

For PSS determination, an effect may be supplied by specifying the null and alternative values of the target parameter as command arguments `#0` and `#a`:

```
. power onesample #0 #a, ...
```

Instead of the alternative value `#a`, you can specify the ratio of the alternative value to the null value in the `ratio()` option and the null value as `#0` for `power onevariance`,

```
. power onevariance #0, ratio(#) ...
```



or you can specify the difference between the alternative value and the null value in the `diff()` option and the null value as `#0` for other methods,

```
. power onesample #0, diff(#) ...
```

For sample-size determination, the reported sample size is rounded up to the nearest integer. This ensures that the corresponding actual power is at least as large as the specified power. You can specify the `nfractional` option to obtain the corresponding fractional sample size, or you can recompute the actual power using the reported rounded value; see [Fractional sample sizes](#) in [PSS] [unbalanced designs](#) for details.

Below we show a quick example of PSS analysis for a one-sample mean test. See entries of the one-sample methods for more examples.

### ► Example 1: PSS analysis for a one-sample mean test

A group of pediatricians would like to study the exposure of infants to television. The group wants to investigate whether the average number of hours watched per day by infants between 3 and 12 months of age is greater than 2 hours. Before conducting a study, pediatricians would like to determine how many infants they need to enroll in the study. The analysis will use the one-sample  $t$  test to compare the mean of the obtained sample with the reference value. An earlier pilot study reported an average of 2.5 hours watched per day with a standard deviation of 0.8. Pediatricians would like to compute the sample size required to detect a mean of 2.5 using a two-sided test with 5% significance level and 80% power. Although pediatricians suspect that the effect is in the upper direction—more than two hours watched on average—they prefer to obtain the required sample size for a two-sided test instead of a one-sided test.

We use `power onemean` to compute the required sample size. We specify the reference or null value of 2 and the comparison or alternative value of 2.5 as command arguments. We also specify the standard deviation of 0.8 in the `sd()` option. We omit the `alpha(0.05)` and `power(0.8)` options because the desired values are the defaults for these options. The default test is two sided, so we do not need to supply any additional information to the command.

```
. power onemean 2 2.5, sd(0.8)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.6250
    m0 =     2.0000
    ma =     2.5000
    sd =     0.8000

Estimated sample size:
    N =         23
```

All `power` commands have a similar output format. Information about the test and tested hypothesis is displayed first. The input and implied values of the study parameters are displayed next under `Study parameters`. The estimated parameters, such as the sample size in this example, are displayed last.

Pediatricians need to enroll 23 infants in the study to detect a standardized difference of 0.625 between the alternative mean of 2.5 and the null mean of 2 given a standard deviation of 0.8 using a 5%-level two-sided one-sample  $t$  test with 80% power.

The pediatricians believe that they have resources to enroll more infants. They wish to compute the power that corresponds to the sample size of 50. To compute the corresponding power, we specify a sample size of 50 in the `n()` option:

```
. power onemean 2 2.5, sd(0.8) n(50)
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha = 0.0500
      N = 50
    delta = 0.6250
     m0 = 2.0000
     ma = 2.5000
     sd = 0.8000
Estimated power:
    power = 0.9911
```

The power increases to 99% for a larger sample of 50 infants.

The pediatricians also want to find out what is the smallest mean difference they can detect with the larger sample of 50 infants while keeping the power at 80%. They assume the effect to be in the upper direction for this computation. To compute the minimum detectable difference, we specify both the sample size in the `n()` option and the power in the `power()` option.

```
. power onemean 2, sd(0.8) n(50) power(0.8)
Performing iteration ...
Estimated target mean for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0; ma > m0
Study parameters:
    alpha = 0.0500
    power = 0.8000
      N = 50
     m0 = 2.0000
     sd = 0.8000
Estimated effect size and target mean:
    delta = 0.4042
     ma = 2.3233
```

The smallest standardized difference that can be detected given the study parameters is about 0.4, with a corresponding mean of 2.32.

◀

## Two-sample tests

The `power` command provides PSS computations for four two-sample tests. `power twomeans` performs PSS analysis for a two-sample means test; `power twoproportions` performs PSS analysis for a two-sample proportions test; `power twocorrelations` performs PSS analysis for a two-sample correlations test; and `power twovariances` performs PSS analysis for a two-sample variances test.

`power twomeans` provides PSS computations for a two-sample means test that compares the means of two independent populations. The computations provided assume known or unknown and equal or unequal population standard deviations of the two groups. See [PSS] [power twomeans](#).

`power twoproportions` provides PSS computations for a two-sample proportions test that compares the proportions in two independent populations with binary outcomes. Three tests are supported: the large-sample Pearson's  $\chi^2$  test, the large-sample likelihood-ratio test, and the small-sample Fisher's exact test. Several effect specifications are available. For example, you can specify the effect of interest as a risk difference, or a relative risk, or an odds ratio. See [PSS] [power twoproportions](#).

`power twocorrelations` provides PSS computations for a two-sample correlations test that compares the correlation coefficients of two independent populations. The computations are based on a Fisher's  $z$  transformation of a correlation coefficient. See [PSS] [power twocorrelations](#).

`power twovariances` provides PSS computations for a two-sample variances test that compares the variances of two independent populations. The computations are based on an  $F$  test of the ratio of variances. You can perform computations in the variance or standard deviation metric. See [PSS] [power twovariances](#).

Also see [Survival analysis](#) for power and sample-size analysis for a two-sample comparison of survivor functions using the `power logrank` and `power exponential` commands.

All two-sample methods compute sample size given power and the control-group and experimental-group values of the target parameter, power given sample size and the control-group and experimental-group values of the target parameter, or MDES and the corresponding target value of the parameter in the experimental group given power, sample size, and the control-group parameter value.

To compute sample size or power, you can specify the magnitude of the effect of interest in two ways: by directly specifying the alternative values of the target parameter in two groups or by specifying the control-group alternative value and the corresponding relation of the experimental-group value to the control-group alternative value.

The two alternative values are specified as command arguments: the alternative value of the target parameter in the control or reference group, `#a1`, and the alternative value of the target parameter in the experimental or comparison group, `#a2`:

```
. power twosample #a1 #a2, ...
```

The experimental-group alternative value, `#a2`, may be omitted if an option containing the relationship between the two alternative values is specified. For example, for `power twomeans` and `power twocorrelations`, such an option is `diff()`, and it specifies the difference between the experimental-group and control-group alternative values:

```
. power twomeans #a1, diff(#) ...
```

For `power twovariances`, such an option is `ratio()`, and it contains the ratio of the experimental-group alternative value to the control-group value:

```
. power twovariances #a1, ratio(#) ...
```

`power twoproportions` provides several alternative specifications in which a difference between the two populations may be expressed. For example, you can express the “difference” as an odds ratio of the experimental group to the control group,

```
. power twoproportions #a1, orratio(#) ...
```

or as a relative risk,

```
. power twoproportions #a1, rrrisk() ...
```

In addition to the total sample size, two-sample methods provide a way to solve for one of the group sizes when the other group size is fixed. This can be achieved by specifying the `compute()` option. To compute the size of the first group, you must specify the `compute(n1)` option and the

size of the second group in the `n2()` option. To compute the size of the second group, you must specify the `compute(n2)` option and the size of the first group in the `n1()` option.

To compute power, you can specify a total sample size in the `n()` option, group sample sizes in the `n1()` and `n2()` options, or one of the group sample sizes and its ratio,  $n_2/n_1$ , in the `nratio()` option; see [PSS] **unbalanced designs** for more specifications.

Below we show a quick example of PSS analysis for a two-sample means test. See entries of the two-sample methods for more examples.

### ► Example 2: PSS analysis for a two-sample mean test

A pharmaceutical company would like to conduct a study to compare a new weight-loss drug with an older drug. Investigators are planning to use a two-sample  $t$  test to compare the average weight loss for the two drugs. The average weight loss of people taking the old drug for 3 months is 12 pounds, with a standard deviation of 5.5 pounds. The new drug is expected to produce an average weight loss of 16 pounds, with a standard deviation of 5 pounds for the same period of time. Investigators want to find out how many subjects they need to recruit into the study to detect the specified difference using a 5% level two-sided test with 90% power.

We use `power twomeans` to perform PSS analyses. We specify the control-group mean 12 and the experimental-group mean 16 as command arguments after the command name. We specify the respective standard deviations in the `sd1()` and `sd2()` options. The default power is set to 0.8, so we specify `power(0.9)` to request 90% power.

```
. power twomeans 12 16, sd1(5.5) sd2(5) power(0.9)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =    4.0000
      m1 =    12.0000
      m2 =    16.0000
      sd1 =    5.5000
      sd2 =    5.0000
Estimated sample sizes:
      N =          76
      N per group =    38
```

We need a sample of 76 subjects, 38 per group, to detect a difference of 4 between the control-group mean of 12 and the experimental-group mean of 16 given the respective standard deviations of 5.5 and 5 with 90% power using a 5%-level two-sided two-sample means  $t$  test.

The default test is two sided. You may specify the `onesided` option to request a one-sided test. The default design is also balanced; see [PSS] **unbalanced designs** for examples of unbalanced designs.

The investigators hope to keep the sample size under 60 and would like to compute the power corresponding to this sample size. To compute the corresponding power, we specify the `n(60)` option instead of the `power()` option:

```
. power twomeans 12 16, sd1(5.5) sd2(5) n(60)
Estimated power for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N =      60
N per group =    30
    delta =    4.0000
      m1 =   12.0000
      m2 =   16.0000
     sd1 =    5.5000
     sd2 =    5.0000
Estimated power:
    power =    0.8259
```

The power decreases to 83% for the smaller sample of 60 subjects.

To keep the power at 90%, the investigators want to compute the smallest difference between the experimental-group mean and the control-group mean (in the upper direction) given the sample of 60 subjects. For this computation, we specify both options `n(60)` and `power(0.9)`:

```
. power twomeans 12, sd1(5.5) sd2(5) n(60) power(0.9)
Performing iteration ...
Estimated experimental-group mean for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1; m2 > m1
Study parameters:
    alpha =    0.0500
    power =    0.9000
      N =      60
N per group =    30
      m1 =   12.0000
      sd1 =    5.5000
      sd2 =    5.0000
Estimated effect size and experimental-group mean:
    delta =    4.4744
      m2 =   16.4744
```

The smallest detectable mean difference is 4.47, with a corresponding value of the experimental-group mean of 16.47.

◀

## Paired-sample tests

The `power` command provides PSS computations for two tests of paired samples. `power pairedmeans` performs PSS analysis for a two-sample paired-means test, and `power pairedproportions` performs PSS analysis for a two-sample paired-proportions test.

`power pairedmeans` provides PSS computations for a two-sample paired  $t$  test assuming known or unknown population standard deviation of the differences between paired observations. You can specify standard deviations of each group and a correlation between paired observations, or you can directly specify the standard deviation of the differences between observations. You can obtain results for a nonzero null hypothesis of a difference between the two paired means. The command also provides a way to adjust computations for a finite population sample. See [PSS] [power pairedmeans](#).

`power pairedproportions` provides PSS computations for a two-sample paired-proportions test that compares proportions in two paired (correlated) samples. The computations are based on McNemar's test of marginal homogeneity. You can specify either the discordant proportions or the marginal proportions. A number of effect specifications are available. For example, you can specify the effect of interest as a relative risk or an odds ratio. See [PSS] **power pairedproportions**.

Both paired methods compute sample size given power and target parameter, power given sample size and target parameter, or MDES and the corresponding target parameter given power and sample size.

For power and sample-size determination of `power pairedmeans`, an effect may be supplied by specifying the alternative values of the two means, pretreatment and posttreatment, as command arguments  $m_{a1}$  and  $m_{a2}$ :

```
power pairedmeans  $m_{a1}$   $m_{a2}$ , ...
```

Instead of the alternative value  $m_{a2}$ , you can specify the difference between the two alternative values in the `altdiff()` option and the alternative pretreatment mean value  $m_{a1}$ :

```
power pairedmeans  $m_{a1}$ , altdiff() ...
```

You may omit both alternative values and specify only the difference between them in the `altdiff()` option:

```
power pairedmeans, altdiff() ...
```

By default, the null value of the difference between the pretreatment and posttreatment means is zero, but you may change it by specifying the `nulldiff()` option.

For PSS determination of `power pairedproportions`, there are a number of ways of specifying an effect of interest; see *Alternative ways of specifying effect* in [PSS] **power pairedproportions**. Two main specifications include the specification of discordant proportions and the specification of marginal probabilities. Specifically, you can supply the information about the effect of interest as discordant proportions  $p_{12}$  and  $p_{21}$ ,

```
power pairedproportions  $p_{12}$   $p_{21}$ , ...
```

or as marginal proportions  $p_{1+}$  and  $p_{+1}$ :

```
power pairedproportions  $p_{1+}$   $p_{+1}$ , corr(numlist) ...
```

When you specify marginal proportions, you must also specify the correlation between paired observations in the `corr()` option.

For sample-size determination, the reported sample size is rounded up to the nearest integer. This ensures that the corresponding actual power is at least as large as the specified power. You can specify the `nfractional` option to obtain the corresponding fractional sample size or you can recompute the actual power using the reported rounded value; see *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

Below we show a quick example of PSS analyses for a two-sample paired-means test. See [PSS] **power pairedmeans** and [PSS] **power pairedproportions** for more examples.

### ▷ Example 3: PSS analysis for a two-sample paired-means test

A forester would like to study the effects of a fertilizer treatment on heights of Virginia pine trees. The trees are planted in pairs with only one of them receiving the fertilizer treatment. The average height of untreated trees is 27.5 feet, with a standard deviation of 4.5 feet. The fertilizer treatment is expected to increase the average height to 30 feet, with the same standard deviation of 4.5 feet. The

correlation between the paired tree heights is expected to be 0.4. The forester would like to know how many pairs of pine trees need to be planted so that a 5%-level two-sided paired-means  $t$  test detects the anticipated difference with 80% power.

We use `power pairedmeans` for power and sample-size analysis. We supply the alternative pretreatment and posttreatment means of 27.5 and 30, respectively, as command arguments after the command name. The standard deviations of the two groups are the same, so we specify their common value in the `sd()` option. We specify the correlation of 0.4 in the `corr()` option. The default value for power is 0.8 and for significance level is 0.05, so we omit the corresponding options `power(0.8)` and `alpha(0.05)`.

```
. power pairedmeans 27.5 30, sd(4.5) corr(0.4)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test assuming sd1 = sd2 = sd
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha =    0.0500          ma1 =   27.5000
      power =    0.8000          ma2 =   30.0000
      delta =    0.5072          sd  =    4.5000
      d0 =     0.0000          corr =    0.4000
      da =     2.5000
      sd_d =    4.9295
Estimated sample size:
      N =          33
```

The forester needs 33 pairs of pine trees to run the experiment.

The forester has resources to plant more trees and would like to compute the power corresponding to the larger sample. To compute power given sample size, we specify sample size in the `n()` option:

```
. power pairedmeans 27.5 30, sd(4.5) corr(0.4) n(50)
Estimated power for a two-sample paired-means test
Paired t test assuming sd1 = sd2 = sd
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha =    0.0500          ma1 =   27.5000
      N =          50          ma2 =   30.0000
      delta =    0.5072          sd  =    4.5000
      d0 =     0.0000          corr =    0.4000
      da =     2.5000
      sd_d =    4.9295
Estimated power:
      power =    0.9400
```

The power increases to 0.94.

The forester may also wish to know the smallest detectable difference between average tree heights of the fertilized group and of the control group that can be detected with 80% power and sample size of 50. To compute this value, we specify both options `n(50)` and `power(0.8)`:

```

. power pairedmeans 27.5, sd(4.5) corr(0.4) n(50) power(0.8)
Performing iteration ...
Estimated target parameters for a two-sample paired-means test
Paired t test assuming sd1 = sd2 = sd
Ho: d = d0 versus Ha: d != d0; da > d0
Study parameters:
      alpha =    0.0500          ma1 =   27.5000
      power =    0.8000          sd   =    4.5000
      N      =     50           corr =    0.4000
      d0     =    0.0000
      sd_d   =    4.9295
Estimated effect size and target parameters:
      delta =    0.4042
      da    =    1.9924
      ma2   =   29.4924

```

The smallest detectable difference is 1.99, with a corresponding value of the average tree height for the fertilized trees of 29.5.

◀

## Analysis of variance models

The `power` command provides PSS computations for three types of analyses of variance (ANOVA) designs: one way, two way, and repeated measures. `power oneway` performs PSS analysis for a one-way ANOVA. `power twoway` performs PSS analysis for a two-way ANOVA. `power repeated` performs PSS analysis for a repeated-measures ANOVA.

`power oneway` provides PSS computations for a one-way ANOVA model. You can choose between the overall  $F$  test of the equality of group means and a test of a mean contrast. You can either specify group means or specify their variability in the computations. See [PSS] [power oneway](#).

`power twoway` provides PSS computations for a two-way fixed-effects ANOVA model. You can choose the overall  $F$  test of the main effect of a row factor, a column factor, or a row-by-column interaction. You can either specify cell means or specify the variance explained by the tested effect. See [PSS] [power twoway](#).

`power repeated` provides PSS computations for one-way and two-way fixed-effects repeated-measures ANOVA models. You can choose the overall  $F$  test of the main effect of a between-subjects factor, a within-subject factor, or a between-within factor interaction. You can either specify cell means or specify the variance explained by the tested effect. See [PSS] [power repeated](#).

All methods compute sample size given power and effect size, power given sample size and effect size, or effect size given power and sample size.

For power and sample-size determination of `power oneway`, an effect may be supplied by specifying the alternative values of group means as command arguments  $m_{a1}$ ,  $m_{a2}$ ,  $m_{a3}$ , and so on:

```
power oneway ma1 ma2 [ma3 ...], ...
```

Instead of the alternative group means, you can specify the variance of the group means in the `varmeans()` option and the number of groups in the `ngroups()` option:

```
power oneway, ngroups() varmeans() ...
```



For power and sample-size determination of `power twoway` and `power repeated`, an effect may be supplied by specifying the alternative values of cell means as command arguments  $m_{a1,1}$ ,  $m_{a1,2}$ , and so on, in a matrix form:

```
power twoway ma1,1 ma1,2 [...] \ ma2,1 ma2,2 [...], ...
```

```
power repeated ma1,1 ma1,2 [...] [\ ma2,1 ma2,2 [...]], ...
```

Instead of the alternative cell means, you can specify the variance of the tested effect in the `vareffect()` option and the dimensions of the cell-means matrix: number of rows and columns for `power twoway` and number of groups and repeated measures for `power repeated`:

```
power twoway, nrows() ncols() factor() vareffect() ...
```

```
power repeated, ngroups() nrepeated() factor() vareffect() ...
```

The means can also be supplied as a matrix at the command line. For example, suppose that we have three groups.

```
power oneway ma1 ma2 ma3, ...
```

The above command would be equivalent to

```
matrix means = (ma1, ma2, ma3)
```

```
power oneway means, ...
```

There are also other alternative specifications of an effect with these commands. See the specific entry of each command.

For sample-size determination, the reported sample size is rounded up to the nearest integer. This ensures that the corresponding actual power is at least as large as the specified power. You can specify the `nfractional` option to obtain the corresponding fractional sample size, or you can recompute the actual power using the reported rounded value; see [Fractional sample sizes](#) in [PSS] **unbalanced designs** for details.

Below we show a quick example of PSS analysis for a one-way ANOVA model. See [PSS] **power oneway**, [PSS] **power twoway**, and [PSS] **power repeated** for more examples.

#### ► Example 4: PSS analysis for a one-way ANOVA model

Researchers would like to compare the effects of four drugs on systolic blood pressure. They would like to use a one-way ANOVA model to test the equality of mean blood-pressure measurements across four drugs. To conduct a study, the researchers need an estimate for the number of subjects to be enrolled in a study. From a previous pilot study, the variance between group means was estimated to be 57, and the error variance was estimated to be 115. The researchers would like to compute the required sample size to detect the effect size of  $0.7040 = \sqrt{57/115}$  with 80% power using a 5%-level  $F$  test of the equality of means assuming a balanced design.

We use `power oneway` to compute the sample size. We specify the number of groups and the estimates of variances in the corresponding options. The default value for power is 0.8 and for significance level is 0.05, so we omit the corresponding options `power(0.8)` and `alpha(0.05)`.

```
. power oneway, ngroups(4) varmeans(57) varerror(115)
Estimated sample size for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.7040
      N_g =      4
      Var_m =   57.0000
      Var_e =  115.0000
Estimated sample sizes:
      N =        28
      N per group =    7
```

The researchers need to recruit 28 subjects, 7 subjects per group, for this study.

Unfortunately, the researchers can afford to recruit only 20 subjects. They wish to compute the power corresponding to this smaller sample size. To compute power, we additionally specify sample size in the `n()` option:

```
. power oneway, n(20) ngroups(4) varmeans(57) varerror(115)
Estimated power for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      N =        20
      N per group =    5
      delta =    0.7040
      N_g =      4
      Var_m =   57.0000
      Var_e =  115.0000
Estimated power:
      power =    0.6400
```

The power decreases to 0.64.

The researchers are not satisfied with such a low power. They now would like to compute the smallest effect size and the corresponding variance of means that can be detected with the power of 80% and the sample size of 20. To compute effect size, we specify both power and sample size in respective options:

```
. power oneway, n(20) power(0.8) ngroups(4) varerror(115)
Performing iteration ...
Estimated between-group variance for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
        N =         20
N per group =     5
      N_g =         4
      Var_e = 115.0000
Estimated effect size and between-group variance:
      delta =    0.8353
      Var_m =   80.2329
```

The smallest detectable effect size is 0.8353, with a corresponding value of the between-group variance of 80.2329.

◀

## Contingency tables

The `power` command provides PSS computations for three types of analyses of contingency tables.

`power cmh` performs PSS analysis for a Cochran–Mantel–Haenszel (CMH) test of association in stratified  $2 \times 2$  tables. The command accommodates unbalanced stratum sizes and unbalanced group sizes within each stratum. See [PSS] [power cmh](#).

`power mcc` performs PSS analysis for a test of association between a risk factor and a disease in  $1:M$  matched case–control studies. See [PSS] [power mcc](#).

`power trend` performs PSS analysis for a test of a linear trend in a probability of response in  $J \times 2$  tables, also known as a Cochran–Armitage test. It accommodates unbalanced designs and unequally spaced exposure levels (doses). With equally spaced exposure levels, a continuity correction is available. See [PSS] [power trend](#).

All methods compute sample size given power and effect size and power given sample size and effect size. `power cmh` and `power mcc` also compute effect size given power and sample size.

For sample-size determination, the reported sample sizes are rounded up to the nearest integer. This ensures that the corresponding actual power is at least as large as the specified power. You can specify the `nfractional` option to obtain the corresponding fractional sample sizes, or you can recompute the actual power using the reported rounded values; see [Fractional sample sizes](#) in [PSS] [unbalanced designs](#) for details.

Below we show a quick example of PSS analysis for a Cochran–Armitage test by using `power trend`; see [PSS] [power trend](#) for more examples.

### ► Example 5: Sample size for a Cochran–Armitage trend test

Consider a study investigating the effectiveness of a new topical antibiotic for the treatment of skin infections.

Suppose that in previous studies of the treatment, we observed the following proportions of successfully treated cases at different doses. We may hypothesize that these represent the probability of a successful treatment for each dose.

Doses/day	Proportion successes
1	0.80
2	0.85
3	0.90

We wish to determine the minimum sample size required for a clinical trial designed to detect a dose–response trend with 80% power using a two-sided 5%-level Cochran–Armitage test.

To compute the required sample size, we specify the values 0.80, 0.85, and 0.90 as the alternative success probabilities for each of the three doses after the command name. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults.

```
. power trend .80 .85 .90
note: exposure levels are assumed to be equally spaced
Performing iteration ...
Estimated sample size for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b != 0; logit(p) = a + b*x
Study parameters:
    alpha =    0.0500
    power =    0.8000
    N_g =      3
    p1 =     0.8000
    p2 =     0.8500
    p3 =     0.9000
Estimated sample sizes:
    N =      597
    N per group =    199
```

A total sample of 597 individuals, 199 individuals per group, must be obtained to detect a linear trend in probability of a successful treatment with 80% power using a two-sided 5%-level Cochran–Armitage test.

Suppose that we can recruit only 300 subjects. We can check how such a reduction in sample size affects the power. To compute power, we specify the alternative group probabilities, as before, and the total sample size in option `n()`.

```
. power trend .80 .85 .90, n(300)
note: exposure levels are assumed to be equally spaced
Estimated power for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b != 0; logit(p) = a + b*x
Study parameters:
    alpha =    0.0500
    N =      300
    N per group =    100
    N_g =      3
    p1 =     0.8000
    p2 =     0.8500
    p3 =     0.9000
Estimated power:
    power =    0.5082
```

With a sample of 300 subjects in this study, the power to detect a linear trend in probabilities decreases dramatically from 0.8 to 0.5, which is unacceptably low for practical purposes.

## Survival analysis

The `power` command provides PSS computations for survival analysis comparing two survivor functions using the log-rank test or the exponential test, as well as for more general survival analysis investigating the effect of a single covariate in a Cox proportional hazards regression model, possibly in the presence of other covariates. It provides the estimate of the number of events required to be observed (or the expected number of events) in a study. The minimal effect size (minimal detectable difference, expressed as the hazard ratio or the log hazard-ratio) may also be obtained for the log-rank test and for the Wald test on a single coefficient from the Cox model.

`power cox` provides estimates of sample size, power, or the minimal detectable value of the coefficient when an effect of a single covariate on subject survival is to be explored using Cox proportional hazards regression. It is assumed that the effect is to be tested using the partial likelihood from the Cox model (for example, score or Wald test) on the coefficient of the covariate of interest. See [PSS] [power cox](#).

`power exponential` reports estimates of sample size or power when the disparity in the two exponential survivor functions is to be tested using the exponential test, the parametric test comparing the two exponential hazard rates. In particular, we refer to the (exponential) hazard-difference test as the exponential test for the difference between hazards and the (exponential) log hazard-ratio test as the exponential test for the log of the hazard ratio or, equivalently, for the difference between log hazards. See [PSS] [power exponential](#).

`power logrank` reports estimates of sample size, power, or minimal detectable value of the hazard ratio (or log hazard-ratio) in the case when the two survivor functions are to be compared using the log-rank test. The only requirement about the distribution of the survivor functions is that the two survivor functions must satisfy the proportional-hazards assumption. See [PSS] [power logrank](#).

For sample-size and power computations, the default effect size corresponds to a value of the hazard ratio of 0.5 and may be changed by specifying the `hratio()` option. The hazard ratio is defined as a ratio of hazards of the experimental group to the control group (or the less favorable of the two groups). Other ways of specifying the effect size are available, and these are particular to each subcommand.

By default, all subcommands assume a type I study, that is, perform computations for uncensored survival data. The censoring information may be taken into account by specifying the appropriate arguments or options. See [PSS] [power cox](#), [PSS] [power logrank](#), and [PSS] [power exponential](#) for details.

### ► Example 6: Sample size for the test of the effect of a covariate in the Cox model

Consider a hypothetical study for which the goal is to investigate the effect of the expression of one gene on subject survival with the Cox proportional hazards regression model. Suppose that the Wald test is to be used to test the coefficient on the gene after fitting the Cox model. Gene expression values measure the level of activity of the gene. Consider the scenario described in [Simon, Radmacher, and Dobbins \(2002\)](#) in which the hazard ratio of 3 associated with a one-unit change in the  $\log_2$  intensity of a gene (or, respectively, with a twofold change in gene expression level) is desired to be detected with 95% power using a two-sided, 0.001-level test. The estimate of the standard deviation of the  $\log_2$ -intensity level of the gene over the entire set of samples is assumed to be 0.75.

```

. power cox, hratio(3) sd(0.75) power(0.95) alpha(0.001)
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 != 0
Study parameters:
      alpha =    0.0010
      power =    0.9500
      delta =    1.0986 (coefficient)
      hratio =    3.0000
      sd     =    0.7500

Censoring:
      Pr_E =    1.0000

Estimated number of events and sample size:
      E =     36
      N =     36

```

Provided that all subjects experience an event in this study, a total of 36 events is required to be observed in the study to ensure the specified power.

See [PSS] **power cox** for more details.

◀

### ▶ Example 7: Sample size for two-sample test of exponential survivor functions

Consider an example from [Lachin \(2011, 490\)](#) of a study comparing two therapies, the combination of a new therapy with the standard one versus the standard alone, in the treatment of lupus nephritis patients. From previous studies, the survivor function of the control group treated with the standard therapy was log linear with a constant yearly hazard rate of 0.3. The number of events (failures) required to ensure 90% power to detect a 50% risk reduction,  $\Delta = 0.5$ , (or, respectively, the log hazard-ratio of  $\ln(0.5) = -0.6931$ ) with a one-sided test at a 0.05 significance level was obtained to be 72 under equal-group allocation. In the absence of censoring, [Lachin \(2011\)](#) determined that a total of 72 subjects (36 per group) would have to be recruited to the study. To obtain this same estimate with **power exponential**, we supply the control hazard rate 0.3 as an argument and specify the **power(0.9)**, **onesided**, and **loghazard** options to request 90% power, a one-sided test, and sample-size determination for the exponential log hazard-ratio test (or test for the log-hazard difference), respectively.

```

. power exponential 0.3, power(0.9) onesided loghazard
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log hazard-ratio, conditional
Ho: ln(HR) = 0 versus Ha: ln(HR) < 0
Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =   -0.6931 (log hazard-ratio)

Survival information:
      h1 =    0.3000
      h2 =    0.1500
      hratio =    0.5000

Estimated sample sizes:
      N =     72
      N per group =    36

```

Further, the study was planned to continue for 6 years with a recruitment period of 4 years. Subjects who did not experience an event by the end of 6 years were censored. For this fixed-duration study with uniform entry (recruitment), the estimate of the sample size increases from 72 to 128. We specify the length of the accrual and the follow-up periods in the `aperiod()` and `fperiod()` options, respectively. We also request to display the expected number of events by using the `show` option.

```
. power exponential 0.3, power(0.9) onesided loghazard aperiod(4) fperiod(2) show
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log hazard-ratio, conditional
Ho: ln(HR) = 0 versus Ha: ln(HR) < 0

Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =   -0.6931 (log hazard-ratio)

Accrual and follow-up information:
      duration =    6.0000
      follow-up =    2.0000
      accrual   =    4.0000 (uniform)

Survival information:
      h1 =    0.3000
      h2 =    0.1500
      hratio =    0.5000

Estimated expected number of events:
      E|Ha =    72      E|Ho =    74
      E1|Ha =    44      E1|Ho =    37
      E2|Ha =    28      E2|Ho =    37

Estimated sample sizes:
      N =    128
      N per group =    64
```

Under the alternative hypothesis of  $H_a: \ln(\Delta) = -0.6931$ , where  $\ln(\Delta)$  denotes the log hazard-ratio of the experimental group to the control group, we expect to observe 44 events in the control group and 28 events in the experimental group. A total of 128 subjects (64 per group) is required to be enrolled into the study to observe an expected total of 72 events under the alternative.

See [PSS] [power exponential](#) for more examples.

◀

## ► Example 8: Sample size for the log-rank test

Consider an example from [Machin and Campbell \(2005\)](#) of a study comparing two forms of surgical resection for patients with gastric cancer. From a prestudy survey, the baseline 5-year survival rate was expected to be 20% and an anticipated increase in survival in the experimental group expressed as a hazard ratio of 0.6667 (corresponding to a 5-year survival rate of approximately 34%) was desired to be detected with 90% power using a two-sided, 0.05 level, log-rank test under 1:1 randomization. To obtain the estimate of the sample size for this study, we use `power logrank` with survival proportion in the control group 0.2 supplied as an argument, the `hratio(0.6667)` option to request a hazard ratio of 0.6667, and the `power(0.9)` option to request 90% power.

```

. power logrank 0.2, hratio(0.6667) power(0.9)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1
Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =    0.6667 (hazard ratio)
      hratio =    0.6667
Censoring:
      s1 =    0.2000
      s2 =    0.3420
      Pr_E =    0.7290
Estimated number of events and sample sizes:
      E =      263
      N =      362
      N per group =    181

```

From the output, 263 events (failures) are required to be observed in this study to ensure 90% power to detect a hazard ratio of 0.6667 by using the log-rank test. The respective estimate of the total number of subjects required to observe 263 events in a 5-year study is 362 with 181 subjects per surgical group. Our estimate, 181, of each group's sample size is close to the manually computed estimate of 180 from [Machin and Campbell \(2005\)](#). This is a fixed-duration study in which 20% of subjects were expected to survive (be censored) by the end of the study.

See [\[PSS\] power logrank](#) for more detailed examples and other available methods of sample-size computation for this type of analysis.

◀

## Tables of results

When `power` is used to perform computations for a single set of study parameters, the results can be displayed either as text or in a table. The default is to display results as text:

```

. power onemean 0 0.2
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.2000
      m0 =    0.0000
      ma =    0.2000
      sd =    1.0000
Estimated sample size:
      N =      199

```



You can specify the `table` option to display results in a table:

```
. power onemean 0 0.2, table
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd
.05	.8	199	.2	0	.2	1

For multiple sets of study parameters, when command arguments or options contain number lists, the results are automatically displayed in a table:

```
. power onemean 0 (0.2 0.5)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd
.05	.8	199	.2	0	.2	1
.05	.8	34	.5	0	.5	1

In this example we specified two values for the second argument.

Default tables can be modified by specifying the `table()` option. For example, we can change the order in which the columns are displayed:

```
. power onemean 0 (0.2 0.5), table(alpha power N m0 ma sd delta)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	m0	ma	sd	delta
.05	.8	199	0	.2	1	.2
.05	.8	34	0	.5	1	.5

Or we can change column labels:

```
. power onemean 0 (0.2 0.5), table(, labels(N "Sample size"))
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	Sample size	delta	m0	ma	sd
.05	.8	199	.2	0	.2	1
.05	.8	34	.5	0	.5	1

Or we can select which columns we want to display:

```
. power onemean 0 (0.2 0.5), table(alpha beta N m0 ma sd)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	beta	N	m0	ma	sd
.05	.2	199	0	.2	1
.05	.2	34	0	.5	1

For more examples, see [\[PSS\] power, table](#).

## Power curves

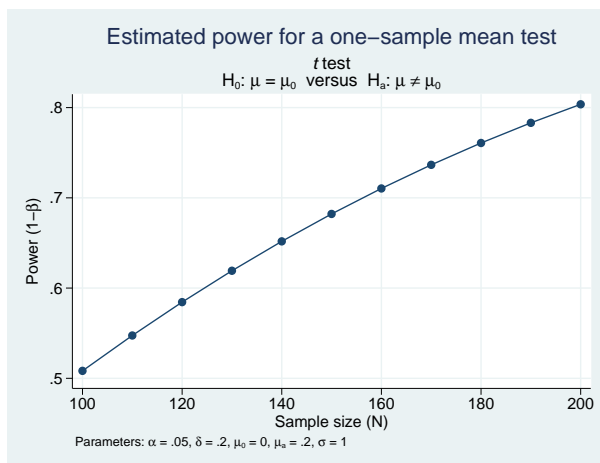
During the planning stage of a study, it is difficult to decide on a number of subjects to be enrolled in a study on the basis of only one set of study parameters. It is common to investigate the effect of various study scenarios on power. Power curves, or plots of estimated power versus one of the study parameters, are commonly used for this purpose.

The `power` command provides the `graph` and `graph()` options to produce power curves.

More precisely, when `graph` is specified, the estimated parameter such as power or sample size is plotted on the  $y$  axis, and the varying parameter is plotted on the  $x$  axis.

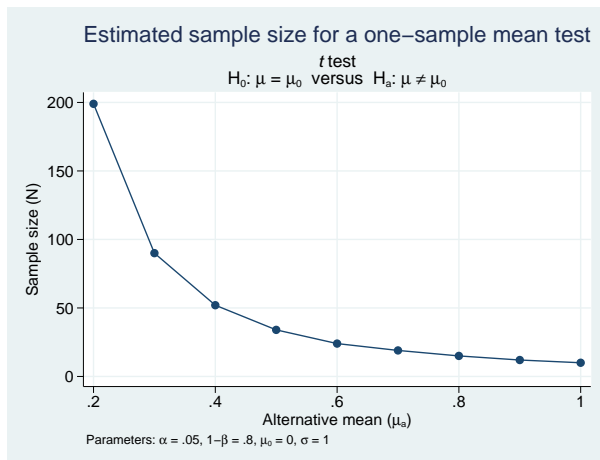
For example, we compute power and plot it as a function of sample size for a range of sample-size values between 100 and 200 with a step size of 10:

```
. power onemean 0 0.2, n(100(10)200) graph
```



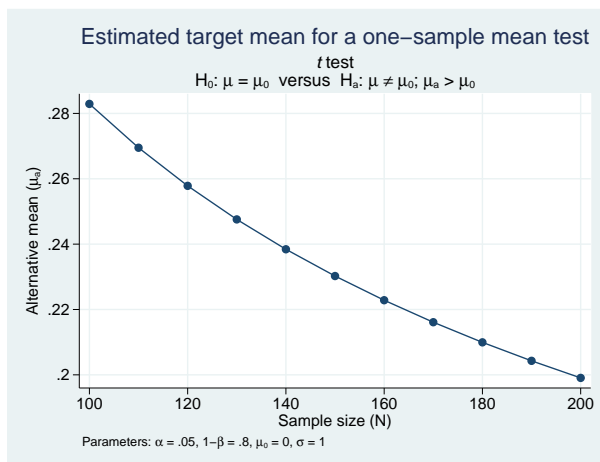
Or we can compute sample size and plot it as a function of the alternative mean when the mean ranges between 0.2 and 1 with a step size of 0.1:

```
. power onemean 0 (0.2(0.1)1), graph
```



Or we can compute the alternative mean for a given power of 80% and a range of sample-size values between 100 and 200 with a step size of 10, and plot it against the sample size:

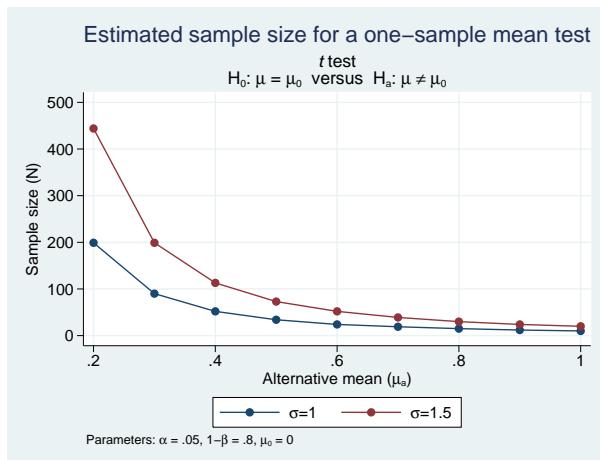
```
. power onemean 0, n(100(10)200) power(0.8) graph
```



The above graphs are the default graphs produced by `power, graph`. Similarly to tabular output, you can customize graphical output by specifying the `graph()` option.

For example, we modify the look of the default graph by using the `graph(nosimplelabels legend(title("")))` option. `nosimplelabels` requests that the graph legend include the column symbol and an equal sign; `legend(title(" "))` requests that the legend not have a title.

```
. power onemean 0 (0.2(0.1)1), sd(1 1.5) graph(nosimplelabels legend(title("")))
```



By default, when a graph is produced, the tabular output is suppressed. You can specify the `table` option if you also want to see results in a table.

For more examples, see [\[PSS\] power, graph](#).

## Stored results

`power` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value of the estimated parameter
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

Macros	
<code>r(type)</code>	test
<code>r(method)</code>	the name of the specified method
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats
Matrix	
<code>r(pss_table)</code>	table of results

Also see *Stored results* in the method-specific manual entries for additional stored results.

## Methods and formulas

By default, the `power` command rounds sample sizes to integers and uses integer values in the computations. To ensure conservative results, the commands round down the input sample sizes and rounds up the output sample sizes. See *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

Some of `power`'s methods require iteration. For example, the sample size for a two-sided test is usually solved iteratively from the two-sided power equation. Most methods use Mata function `solven1()` and its Newton's method described in *Newton-type methods* in [M-5] **solvenl()** to solve a nonlinear power equation. Other methods use a bisection method to find a root of a nonlinear power equation.

See *Methods and formulas* in the method-specific manual entries for details.

## References

- Batistatou, E., C. Roberts, and S. Roberts. 2014. *Sample size and power calculations for trials and quasi-experimental studies with clustering*. *Stata Journal* 14: 159–175.
- Lachin, J. M. 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.
- Machin, D., and M. J. Campbell. 2005. *Design of Studies for Medical Research*. Chichester, UK: Wiley.
- Simon, R., R. D. Radmacher, and K. Dobbin. 2002. Design of studies using DNA microarrays. *Genetic Epidemiology* 23: 21–36.

## Also see

- [PSS] **intro** — Introduction to power and sample-size analysis
- [PSS] **Glossary**

# Title

**power, graph** — Graph results from the power command

[Description](#)

[Remarks and examples](#)

[Menu](#)

[Also see](#)

[Syntax](#)

[Suboptions](#)

## Description

The `graph()` option of `power` specifies that results of the `power` command be graphed.

While there are many options for controlling the look of the graph, you will often merely need to specify the `graph` option on your `power` command.

## Menu

Statistics > Power and sample size

## Syntax

*Produce default graph*

```
power ..., graph ...
```

*Graph power against sample size*

```
power ..., graph(y(power) x(N)) ...
```

*Graph sample size against target parameter*

```
power ..., graph(y(N) x(target)) ...
```

*Graph effect size against sample size*

```
power ..., graph(y(delta) x(N)) ...
```

*Produce other custom graphs*

```
[power] ..., graph(graphopts) ...
```

<i>graphopts</i>	Description
Main	
<u>y</u> dimension( <i>dimlist</i> [, <i>dimopts</i> ])	use <i>dimlist</i> to define <i>y</i> axis
<u>x</u> dimension( <i>dimlist</i> [, <i>dimopts</i> ])	use <i>dimlist</i> to define <i>x</i> axis
<u>plot</u> dimension( <i>dimlist</i> [, <i>dimopts</i> ])	create plots for groups in <i>dimlist</i>
<u>by</u> dimension( <i>dimlist</i> [, <i>dimopts</i> ])	create subgraphs for groups in <i>dimlist</i>
<u>graph</u> dimension( <i>dimlist</i> [, <i>dimopts</i> ])	create graphs for groups in <i>dimlist</i>
<u>horizontal</u>	swap <i>x</i> and <i>y</i> axes
<u>schemegrid</u>	do not apply default <i>x</i> and <i>y</i> grid lines
<u>name</u> ( <i>name</i>   <i>stub</i> [, <i>replace</i> ])	name of graph, or <i>stub</i> if multiple graphs
Labels	
<u>y</u> regular	place regularly spaced ticks and labels on the <i>y</i> axis
<u>x</u> regular	place regularly spaced ticks and labels on the <i>x</i> axis
<u>y</u> values	place ticks and labels on the <i>y</i> axis for each distinct value
<u>x</u> values	place ticks and labels on the <i>x</i> axis for each distinct value
<u>collabels</u> ( <i>colspec</i> )	change default labels for columns
<u>no</u> labels	label groups with their values, not their labels
<u>allsimple</u> labels	forgo column label and equal signs in all labels
<u>nosimple</u> labels	include column label and equal signs in all labels
<u>eq</u> separator( <i>string</i> )	replace equal sign separator with <i>string</i>
<u>separator</u> ( <i>string</i> )	separator for labels when multiple columns are specified in a dimension
<u>no</u> separator	do not use a separator
<u>format</u> ( <i>%fmt</i> )	format for converting numeric values to labels
Plot	
<u>plot</u> opts( <i>plot_options</i> )	affect rendition of all plots
<u>plot</u> #opts( <i>plot_options</i> )	affect rendition of #th plot
<u>recast</u> ( <i>plotype</i> )	plot all plots using <i>plotype</i>
Add plots	
<u>add</u> plot( <i>plot</i> )	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall, By	
<u>twoway</u> _options	any options documented in [G-3] <i>twoway_options</i>
<u>by</u> opts( <i>byopts</i> )	how subgraphs are combined, labeled, etc.

*dimlist* may contain any of the following columns:

<i>column</i>	Description
<code>alpha</code>	significance level
<code>power</code>	power
<code>beta</code>	type II error probability
<code>N</code>	total number of subjects
<code>N1</code>	number of subjects in the control group
<code>N2</code>	number of subjects in the experimental group
<code>nratio</code>	ratio of sample sizes, experimental to control
<code>delta</code>	effect size
<code>target</code>	target parameter
<i>method_columns</i>	columns specific to the <code>method</code> specified with <code>power</code>

*colspec* is

```
column "label" [column "label" [...]]
```

<i>dimopts</i>	Description
<code>labels(<i>lablist</i>)</code>	list of quoted strings to label each level of the dimension
<code>elabels(<i>elablist</i>)</code>	list of enumerated labels
<code>nolabels</code>	label groups with their values, not their labels
<code>allsimplelabels</code>	forgo column name and equal signs in all labels
<code>nosimplelabels</code>	include column name and equal signs in all labels
<code>eqseparator(<i>string</i>)</code>	replace equal sign separator with <i>string</i> in the dimension
<code>separator(<i>string</i>)</code>	separator for labels when multiple columns are specified in the dimension
<code>noseparator</code>	do not use a separator
<code>format(<i>%fmt</i>)</code>	format for converting numeric values to labels

where *lablist* is defined as

```
"label" ["label" [...]]
```

*elablist* is defined as

```
# "label" [# "label" [...]]
```

and the #s are the levels of the dimension.

<i>plot_options</i>	Description
<i>marker_options</i>	change look of markers (color, size, etc.)
<i>marker_label_options</i>	add marker labels; change look or position
<i>cline_options</i>	change look of the line



## Suboptions

The following are suboptions within the `graph()` option of the `power` command.

Main

`ydimension()`, `xdimension()`, `plotdimension()`, `bydimension()`, and `graphdimension()` specify the dimension to be used for the graph's  $y$  axis,  $x$  axis, plots, `by()` subgraphs, and graphs.

The default dimensions are based on your analysis. The  $y$  dimension is power for power determination, sample size for sample-size determination, and target parameter for effect-size determination. If there is only one column containing multiple values, this column is plotted on the  $x$  dimension. Otherwise, the  $x$  dimension is sample size for power determination, target parameter for sample-size determination, and sample size for effect-size determination. Other columns that contain multiple values are used as plot dimensions. See *Default graphs* below for details. You may override the defaults and explicitly control which columns are used on each dimension of the graph using these dimension suboptions.

Each of these suboptions supports *suboptions* that control the labeling of the dimension—axis labels for `ydimension()` and `xdimension()`, plot labels for `plotdimension()`, subgraph titles for `bydimension()`, and graph titles for `graphdimension()`.

For examples using the dimension suboptions, see *Changing default graph dimensions* below.

`ydimension(dimlist [ , dimopts ])` specifies the columns for the  $y$  axis in *dimlist* and controls the content of those labels with *dimopts*.

`xdimension(dimlist [ , dimopts ])` specifies the columns for the  $x$  axis in *dimlist* and controls the content of those labels with *dimopts*.

`plotdimension(dimlist [ , dimopts ])` specifies in *dimlist* the columns whose levels determine the plots and optionally specifies in *dimopts* the content of the plots' labels.

`bydimension(dimlist [ , dimopts ])` specifies in *dimlist* the columns whose levels determine the `by()` subgraphs and optionally specifies in *dimopts* the content of the subgraphs' titles.

`graphdimension(dimlist [ , dimopts ])` specifies in *dimlist* the columns whose levels determine the graphs and optionally specifies in *dimopts* the content of the graphs' titles.

See the definition of *columns in graph* in [PSS] *Glossary*.

`horizontal` reverses the default  $x$  and  $y$  axes. By default, the values computed by `power` are plotted on the  $y$  axis, and the  $x$  axis represents one of the other columns. Specifying `horizontal` swaps the axes.

One common use is to put sample size on the  $x$  axis even when it is the value computed by `power`. This suboption can also be useful with the long labels produced when the `parallel` option is specified with `power`.

See *Parallel plots* below for an example of the `horizontal` suboption.

`schemegrid` specifies that  $x$  and  $y$  grid lines not always be drawn on the power graph. Instead, whether grid lines are drawn will be determined by the current scheme.

`name(name | stub [ , replace ])` specifies the name of the graph or graphs. If the `graphdimension()` suboption is specified, then the argument of `name()` is taken to be *stub*, and graphs named *stub1*, *stub2*, ... are created.

`replace` specifies that existing graphs of the same name may be replaced.

If `name()` is not specified, default names are used, and the graphs may be replaced by subsequent `power` graphs or other graphing commands.

## Labels

All the suboptions listed under the **Labels** tab may be specified directly within the `graph()` option. All of them except `yregular`, `xregular`, `yvalues`, and `xvalues` may be specified as *dimopts* within `ydimension()`, `xdimension()`, `plotdimension()`, `bydimension()`, and `graphdimension()`. When suboptions are specified in one of the dimension options, only the labels for that dimension are affected. When suboptions are specified outside the dimension options, all labels on all dimensions are affected. Specifications within the dimension options take precedence.

`yregular` and `yvalues` specify how tick marks and labels are to be placed on the *y* axis.

`yregular` specifies that regularly spaced ticks and labels be placed on the *y* axis.

`yvalues` specifies that a tick and label be placed for each distinct value.

If neither is specified, an attempt is made to choose the most reasonable option based on your results. Labeling may also be specified using the standard `graph twoway` [axis labeling rules and options](#).

`xregular` and `xvalues` do the same for tick marks and labels to be placed on the *x* axis.

`collabels(colspec)` specifies labels to be used on the graph for the specified columns. For example, `collabels(N "N")` specifies that wherever the column *N* is used on a graph—axis label, plot label, graph title, legend title, etc.—“*N*” be shown rather than the default label “Sample size”.

Multiple columns may be relabeled by typing, for example,

```
collabels(N "N" ma "Alternative mean")
```

and `SMCL` tags for Greek characters and other typesetting can be used by typing, for example,

```
collabels(alpha "&alpha;" ma "&mu;{sub:a}")
```

See the definition of *columns in graph* in [\[PSS\] Glossary](#).

`nolabels` specifies that value labels not be used to construct graph labels and titles for the levels in the dimension. By default, if a column in a dimension has value labels, those labels are used to construct labels and titles for axis ticks, plots, subgraphs, and graphs.

`allsimplelabels` and `nosimplelabels` control whether a graph’s labels and titles include just the values of the columns or also include column labels and equal signs. The default depends on whether the dimension is an axis dimension or one of the plot, by, and graph dimensions. It also depends on whether the values for the level of the dimension are labeled. An example of a simple label is “power” or “.8” and of a nonsimple label is “power=.8”.

In `power`, `graph` simple labels are almost universally best for *x* and *y* axes and also best for most plot labels. Labels with an equal sign are typically preferred for subgraph and graph titles. These are the defaults used by `power`, `graph`. The `allsimplelabels` and `nosimplelabels` suboptions let you override the default labeling.

`allsimplelabels` specifies that all titles and labels use just the value or value label of the column.

`nosimplelabels` specifies that all titles and labels include *dimname*= before the value or value label.

`eqseparator(string)` specifies a custom separator between column labels and values in labels. Use *string* in place of the default equal sign.

`separator(string)` and `noseparator` control the separator between label sections when more than one column is used to specify a dimension. The default separator is a comma followed by a space, but no separator may be requested with `noseparator`, or the default may be changed to any string with `separator()`.

For example, if `bydimension(a b)` is specified, the subgraph labels in our graph legend might be “a=1, b=1”, “a=1, b=2”, . . . . Specifying `separator(:)` would create labels “a=1:b=1”, “a=1:b=2”, . . . .

`format(%fmt)` specifies how numeric values are to be formatted for display as axis labels, labels on plots, and titles on subgraphs and graphs.

#### Plot

`plotopts(plot_options)` affects the rendition of all plots. The *plot\_options* can affect the size and color of markers, whether and how the markers are labeled, and whether and how the points are connected; see [G-3] [marker\\_options](#), [G-3] [marker\\_Label\\_options](#), and [G-3] [cline\\_options](#).

These settings may be overridden for specific plots by using the `plot#opts()` suboption.

`plot#opts(plot_options)` affects the rendition of the #th plot. The *plot\_options* can affect the size and color of markers, whether and how the markers are labeled, and whether and how the points are connected; see [G-3] [marker\\_options](#), [G-3] [marker\\_Label\\_options](#), and [G-3] [cline\\_options](#).

`recast(plottype)` specifies that results be plotted using *plottype*. *plottype* may be `scatter`, `line`, `connected`, `area`, `bar`, `spike`, `dropline`, or `dot`; see [G-2] [graph twoway](#). When `recast()` is specified, the plot-rendition options appropriate to the specified *plottype* may be used in lieu of *plot\_options*. For details on those suboptions, follow the appropriate link from [G-2] [graph twoway](#).

You may specify `recast()` within a `plotopts()` or `plot#opts()` suboption. It is better, however, to specify it as documented here, outside those suboptions. When it is specified outside those suboptions, you have greater access to the plot-specific rendition suboptions of your specified *plottype*.

#### Add plots

`addplot(plot)` provides a way to add other plots to the generated graph; see [G-3] [addplot\\_option](#).

If multiple graphs are drawn by a single `power` command or if *plot* specifies plots with multiple *y* variables, for example, `scatter y1 y2 x`, then the graph’s legend will not clearly identify all the plots and will require customization using the `legend()` suboption; see [G-3] [legend\\_options](#).

#### Y axis, X axis, Titles, Legend, Overall, By

*twoway\_options* are any of the options documented in [G-3] [twoway\\_options](#). These include options for titling the graph (see [G-3] [title\\_options](#)); for saving the graph to disk (see [G-3] [saving\\_option](#)); for controlling the labeling and look of the axes (see [G-3] [axis\\_options](#)); for controlling the look, contents, position, and organization of the legend (see [G-3] [legend\\_options](#)); for adding lines (see [G-3] [added\\_line\\_options](#)) and text (see [G-3] [added\\_text\\_options](#)); and for controlling other aspects of the graph’s appearance (see [G-3] [twoway\\_options](#)).

The `label1()` suboption of the `legend()` option has no effect on `power`, `graph`. Use the `order()` suboption instead.

`byopts(byopts)` affects the appearance of the combined graph when `bydimension()` is specified or when the default graph has subgraphs, including the overall graph title, the position of the legend, and the organization of subgraphs. See [G-3] [by\\_option](#).

## Remarks and examples

Remarks are presented under the following headings:

- Using power, graph*
- Graph symbols*
- Default graphs*
- Changing default graph dimensions*
- Changing the look of graphs*
- Parallel plots*

`power, graph` produces power curves and other graphical output from the `power` command. Power graphs are useful for visualizing the results of sensitivity analysis, which investigates the effect of varying study parameters on power, sample size, or other components of the study. The true values of study parameters are usually unknown. Power and sample-size analysis uses best guesses for these values. It is important to evaluate the sensitivity of the computed power or sample size to the chosen values of study parameters. For example, to evaluate variability of power values, you can compute powers for various ranges of values for the parameters of interest and display the resulting powers in a table (see [PSS] [power, table](#)) or plot them on a graph.

### Using power, graph

In most cases, you will probably be satisfied with the graphs that `power` produces by default when you specify the `graph` option. For other cases, `power, graph()` offers many options for you to produce the graph you desire.

Think of `power, graph()` as graphing the columns of `power, table`. One of the columns will be placed on the  $x$  axis, another will be placed on the  $y$  axis, and, if you have more columns with varying values, separate plots will be created for each. Similarly, we use the terms “column symbol”, “column name”, and “column label” to refer to symbols, names, and labels that appear in tables when tabular output is requested.

By default, `power, graph` plots the column corresponding to the estimated parameter on the  $y$  axis: `power`, when power is computed; `N`, when sample size is computed; and `target`, when the target parameter is computed. When there is only one varying column, the  $x$  axis uses this column by default. When there are multiple varying columns, the default  $x$  axis depends on what is being computed.

If power is computed (power determination), the default  $x$  axis is either the sample size, if sample size varies, or the target parameter, if the target parameter varies and sample size does not vary. If neither the sample size nor the target parameter varies, the power is plotted against one of the other varying parameters.

If sample size is computed (sample-size determination), the default  $x$  axis is either the target parameter, if the target parameter varies, or the power, if power varies and the target parameter does not vary. If neither the target parameter nor the power varies, the sample size is plotted against one of the other varying parameters.

If target parameter is computed (effect-size determination), the default  $x$  axis is either sample size, if sample size varies, or power, if power varies and sample size does not vary. If neither sample size nor power varies, the target parameter is plotted against one of the other varying parameters. You can also plot the effect size `delta` instead of the target parameter `target` by specifying the `ydimension(delta)` suboption within `graph()`; see [example 4](#). The graphs of `delta` may not reflect all the unique combinations of other study parameters, because effect size is not necessarily a one-to-one function of the constituent study parameters.

`power`, `graph()` provides great flexibility for customizing graphical output. You can make minor changes such as modifying the graph or axes titles or modifying the line color and style, or you can completely redesign the graph by changing the axes and style of the graph. The Graph Editor can also be used for additional customization; see [G-1] [graph editor](#).

When you produce a graph, the table of results is suppressed. You can request that the table be displayed in addition to the graph by specifying the `table` option with `graph()`.

## Graph symbols

Whenever space allows, such as on  $y$  and  $x$  axes, graphical output displays *extended column labels*, which include column labels and column symbols in parentheses. In other cases, such as in legend labels or by graph titles, graphical output includes only column (parameter) symbols for readability.

The following common symbols are used. See the documentation entry of the specified power method for additional symbols specific to that method.

Symbol	Description
$\alpha$	significance level
$\beta$	probability of a type II error
$1 - \beta$	power
$N$	total sample size
$N_1$	sample size of the control group
$N_2$	sample size of the experimental group
$\delta$	effect size
<i>method_symbols</i>	symbols specific to the <i>method</i> specified with <code>power</code>

## Default graphs

We start with a demonstration of several default power graphs and then show how you can produce custom power graphs in the subsequent sections.

In what follows, we graph the results of power and sample-size analysis for a two-sided 5%-level one-sample  $t$  test comparing the population mean with a hypothesized value; see [PSS] [power onemean](#).

### Example 1: Power curves

When we compute power given a range of sample sizes, `power`, `graph` plots power on the  $y$  axis and sample size on the  $x$  axis.

```
. power onemean 0 1, n(10(2)40) graph
```

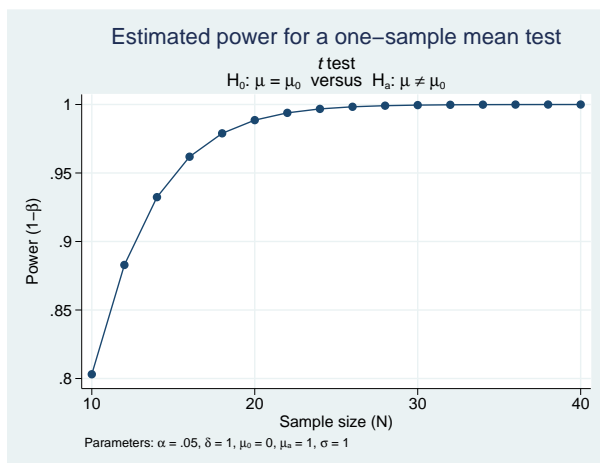


Figure 1.

As expected, power increases as sample size increases.

The default axis labels include column labels and column symbols in parentheses. The labels can be changed as we show in [example 6](#). The values of constant parameters are displayed in the note titled “Parameters”: significance level  $\alpha$  is 0.05, effect size  $\delta$  (the standardized mean difference  $(\mu_a - \mu_0)/\sigma$  for a one-sample  $t$  test) is 1, null mean  $\mu_0$  is 0, alternative mean  $\mu_a$  is 1, and standard deviation  $\sigma$  is 1.

In addition to varying sample size, we may compute powers for different alternative values of the mean.

```
. power onemean 0 (0.8 1), n(10(2)40) graph
```

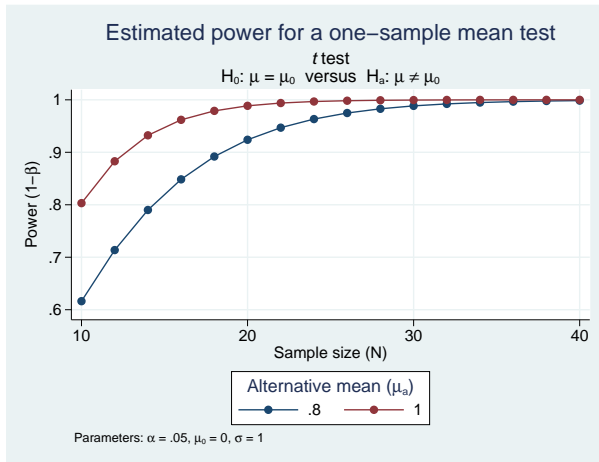


Figure 2.

For a given sample size, the larger the alternative value, the higher the power.

`power, graph` displays two power curves corresponding to the specified alternative values on one plot. The first curve is displayed in navy, and the second curve is displayed in maroon. The default colors of the lines and, in general, the overall look of the graph are determined by the current graph scheme. The scheme used here is `s2gco1or`; see [G-2] [set scheme](#) for details. We also show how to change the default look of the curves in [example 7](#).

We can obtain power curves for varying values of several parameters. For example, below we compute powers for varying values of alternative mean, sample size, and standard deviation.

```
. power onemean 0 (0.8 1), n(10(2)40) sd(1 1.5) graph
```

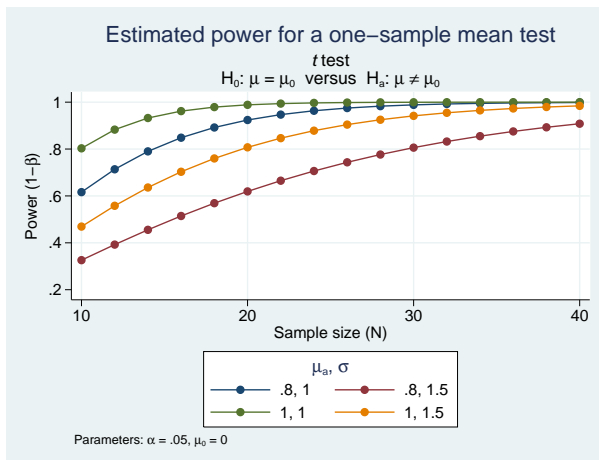


Figure 3.

The larger the standard deviation, the lower the power. This can be seen more easily in [figure 10](#) and [figure 19](#).

`power, graph` plots a separate curve for each unique combination of the values of the alternative mean and standard deviation on one plot. Alternatively, you can display curves on separate plots (by graphs) or even on separate graphs; see [example 5](#). Instead of the extended legend label, as displayed in [figure 2](#), the title of the legend now displays only column symbols because the legend contains more than one column. For an alternative look of the legend, see [figure 16](#).

If we specify only one sample size in the previous figure, the values of the alternative mean will be plotted on the  $x$  axis. You can try this yourself if you would like.



## Example 2: Sample-size curves

Instead of power curves, we can plot estimated sample sizes for a range of power values to get an idea of how the requirement on sample size changes for powers of interest.

```
. power onemean 0 1, power(0.8(0.05)0.95) graph
```

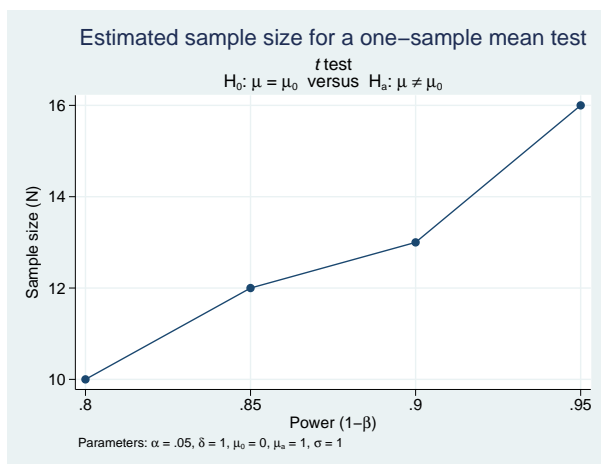


Figure 4.

The sample size increases as the power increases.

The look of this graph is the same as that of the graph in [figure 1](#), except sample size is plotted on the  $y$  axis and power is plotted on the  $x$  axis.



We may want to investigate how other study parameters such as alternative mean and standard deviation affect the sample-size requirement for given powers.

```
. power onemean 0 (0.3(0.1)1), power(0.8 0.9) sd(1 1.5) graph
```

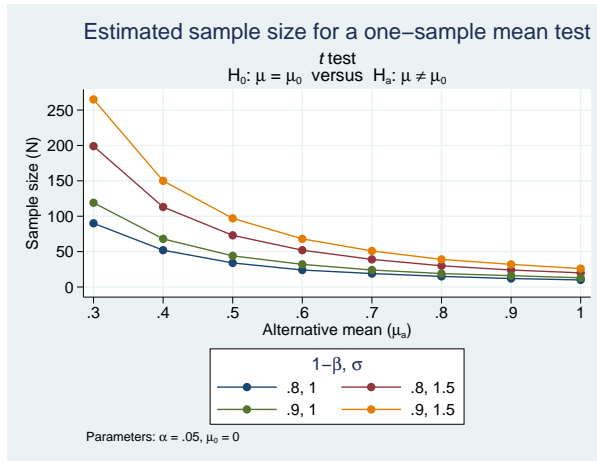


Figure 5.

The larger the alternative mean, or more precisely, the larger the standardized difference between the alternative and null means, the larger the required sample size.

When multiple study parameters each contain multiple values, as in the [above figure](#), the default  $x$  axis for sample-size curves is the target parameter (the alternative mean in our example) provided that the target parameter varies. You can plot a different parameter on the  $x$  axis such as standard deviation; see [example 4](#) about how to change the default axes.

◀

### ▷ Example 3: “Effect-size” curves

What we mean by “effect-size” curves are curves with an effect of interest, which may or may not be the actual effect size, plotted on the  $y$  axis. In fact, `power, graph` by default plots the target parameter on the  $y$  axis.

We can plot the alternative mean (the target parameter), or more precisely, the smallest value of the alternative mean that can be detected using the considered  $t$  test, against sample size for specified values of power and default values of other study parameters.

```
. power onemean 0, power(0.8 0.9) n(10(2)40) graph
```

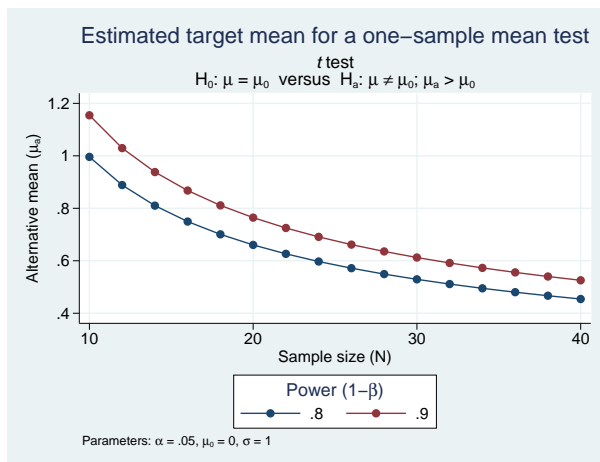


Figure 6.

4

The larger the sample size and the smaller the power, the smaller the values of the alternative mean that can be detected by the test. The default  $x$  axis is the sample size whenever the sample size varies.

If desired, you can plot the actual effect size instead of the target parameter on the  $y$  axis; see [example 4](#) for details.

## Changing default graph dimensions

So far, we have demonstrated the graphs that `power, graph` produces by default. In this section, we demonstrate how you can modify the default graphs.

We can use `power, graph()` to modify graphs by specifying suboptions to control the look of the graph.

### ► Example 4: Changing default graph axes

The default  $y$  axis corresponds to the computed study parameter—power for power determination, sample size for sample-size determination, and target parameter for effect-size determination. You would rarely need to change this dimension. One example when this may be useful is when you want to plot the estimated probability of a type II error,  $\beta$ , instead of the estimated power.

Following [figure 1](#), let's plot the estimated probability of a type II error instead of power on the  $y$  axis. We specify the name of the column to be displayed on the  $y$  axis, `beta`, in the `ydimension()` suboption of `power's graph()` option. We use the minimum abbreviation `y()` of the suboption.

```
. power onemean 0 1, n(10(2)40) graph(y(beta))
```

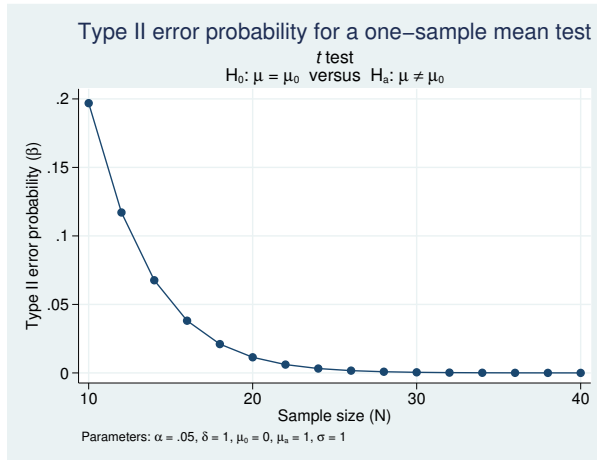


Figure 7.

The probability of a type II error decreases as the sample size increases. This is expected considering the relationship between power,  $1 - \beta$ , and  $\beta$ .

In [example 3](#), we plotted the minimum detectable values of the target parameter, alternative mean, against sample size. Instead of the alternative mean, we can plot the corresponding effect size, delta.

```
. power onemean 0, power(0.8 0.9) n(10(2)40) graph(y(delta))
```

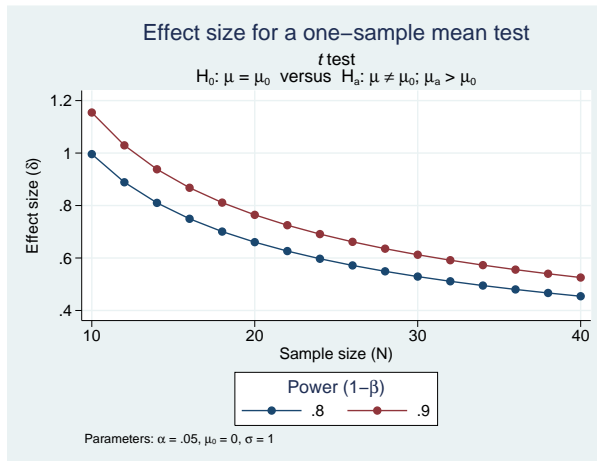


Figure 8.

The effect size is now plotted on the  $y$  axis. The  $y$  values are the same as in [figure 6](#) because the effect size corresponds to the alternative mean when the null mean is 0 and the standard deviation is 1. Note also that for a one-sample  $t$  test, the computed effect size is not affected by the specified values of the null mean or standard deviation, so the effect-size curves will stay the same if you vary these parameters.

When the `ydimension()` suboption is specified, the  $y$  axis is replaced with the specified column, and the column corresponding to the default  $y$  axis is used as a plot dimension. When the `ydimension()`

suboption contains delta, the target column is omitted from the graph. When the ydimension() suboption contains beta, the power column is omitted from the graph.

In figure 2, by default, the power is plotted against varying values of the sample size, and a separate curve is plotted for each of the varying values of the alternative mean. We can change the default  $x$  axis by specifying the xdimension() suboption (abbreviated to x()) within power's graph() option.

```
. power onemean 0 (0.3(0.1)1), n(10 20 40) graph(x(ma))
```

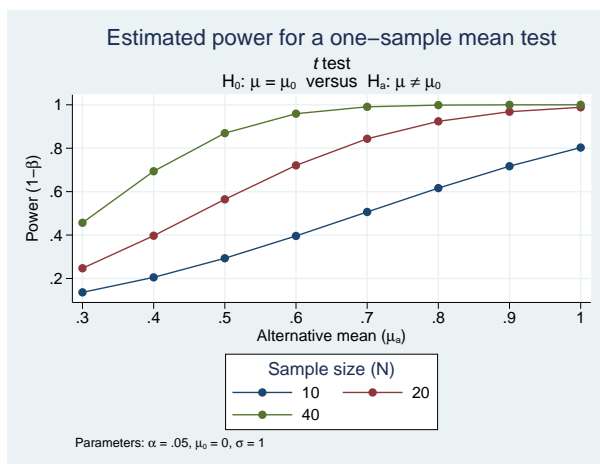


Figure 9.

The  $x$  axis now contains the values of the alternative mean, and a separate curve is now plotted for each sample size.

When the xdimension() suboption is specified, the  $x$  axis is replaced with the specified column, and the column corresponding to the default  $x$  axis is used as a plot dimension. When the xdimension() suboption contains delta, the target column is omitted from the graph. When the xdimension() suboption contains beta, the power column is omitted from the graph.

◀

## ► Example 5: By graphs and multiple graphs

Let's return to figure 3 demonstrating multiple power curves on one graph. Suppose we want to more easily see the impact of standard deviation on power given an alternative mean value. We can produce a separate plot for each of the mean values by specifying the column ma in the bydimension() suboption (abbreviated to by()) within power's graph() option.

```
. power onemean 0 (0.8 1), n(10(2)40) sd(1 1.5) graph(by(ma))
```

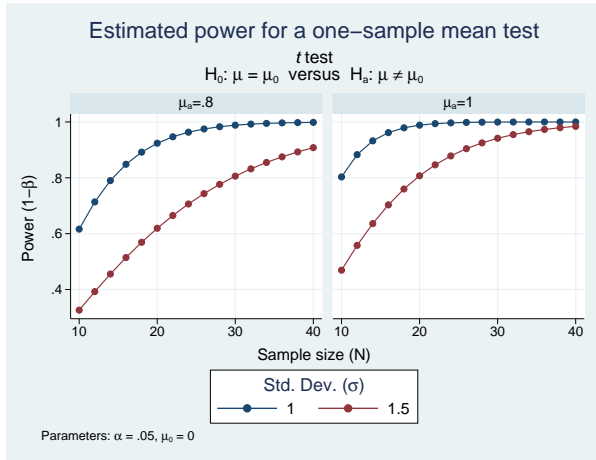


Figure 10.

For examples of how to modify the look of a by graph, see [example 8](#).

In the presence of many varying parameters, even by graphs may look crowded. In this case, you may consider producing multiple by graphs. In the example above, suppose that we also want to vary the significance level  $\alpha$ . We add the `alpha(0.05 0.1)` option to the previous command.

```
. power onemean 0 (0.8 1), n(10(2)40) sd(1 1.5) alpha(0.05 0.1) graph(by(ma))
(output omitted)
```

The above command produces a graph containing two by graphs. Each by graph contains four curves in which each corresponds to a unique combination of values of the standard deviation and significance level. We leave this for you to verify.

This is a lot of information for a single graph to convey, so instead, we request that a separate graph be produced for each of the significance levels by specifying the `graphdimension(alpha)` suboption (abbreviated to `graph()`) within `power`'s `graph()` option.

```
. power onemean 0 (0.8 1), n(10(2)40) sd(1 1.5) alpha(0.05 0.1)
> graph(by(ma) graph(alpha))
```

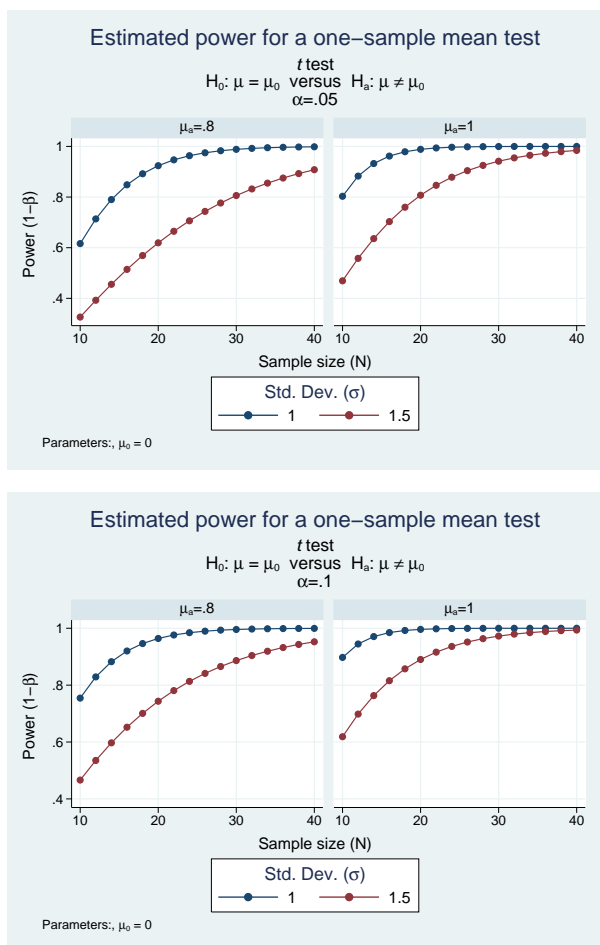


Figure 11.

◀

## Changing the look of graphs

### ► Example 6: Modifying axis labels

Reasonable defaults for axis labels are chosen based on your results. You can modify the defaults by using any of `power`, `graph()`'s labeling suboptions or `graph twoway`'s `axis_label_options`; see [G-3] [axis\\_label\\_options](#).

For example, we can request that ticks and labels of the  $y$  and  $x$  axes be placed for each distinct value instead of using equally spaced values as in figure 1.

```
. power onemean 0 1, n(10(2)20) graph(yvalues xvalues ylabel(, format(%4.3f)))
```

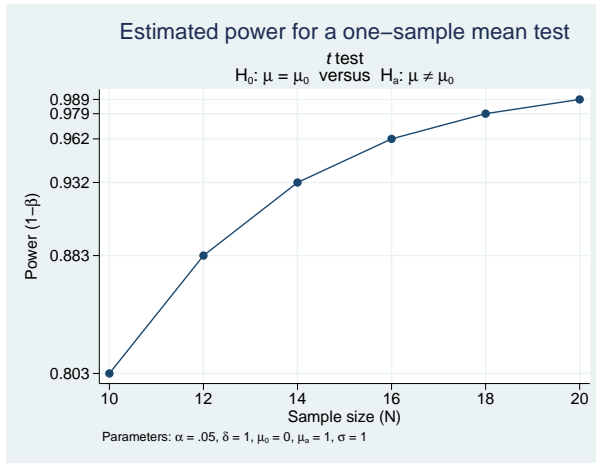


Figure 12.

In this example, we specified fewer sample sizes to obtain a more readable graph. To further improve readability, we also changed the default format of the values on the  $y$  axis to show only three decimal points by using `ylabel(, format(%4.3f))`.

We can use `ylabel()` and `xlabel()` to add text labels for some of the axis values. For example, suppose that our budget is 30 subjects. We can use `xlabel()` to label the sample-size value of 30 as “Budgeted”.

```
. power onemean 0 1, n(10(2)40) graph(xlabel(30 "Budgeted", add))
```

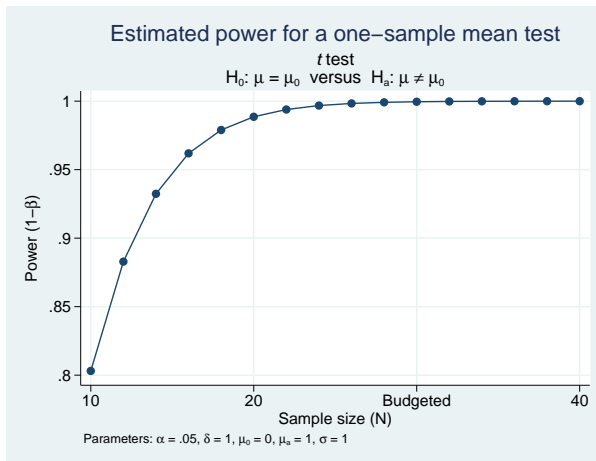


Figure 13.

We can use `ytittle()` and `xtittle()` to change the axis titles.

```
. power onemean 0 1, n(10(2)20) graph(ytitle("Power") xtitle("Sample size"))
> title("Estimated power") subtitle("") note("")
```

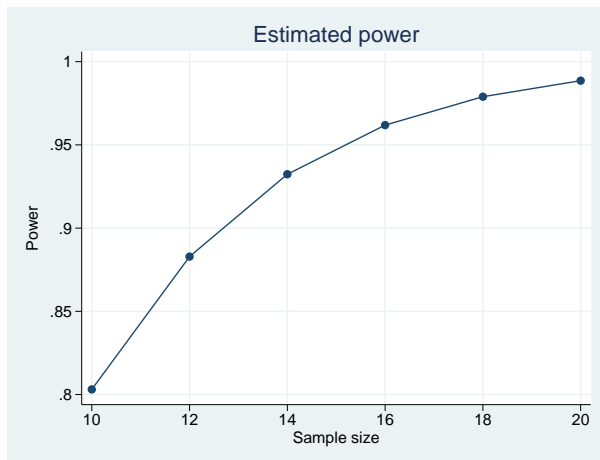


Figure 14.

In addition to modifying the axis titles, we also shortened the default title and suppressed the default subtitle and note.

You may find the `collabels()` suboption useful to override the default column labels. The specified column labels will be used wherever the corresponding column is used on the graph.

For example, change the default labels of the power, sample-size, and alternative-mean columns to, respectively, “Power”, “N”, and “Alternative mean” in [figure 2](#) as follows:

```
. power onemean 0 (0.8 1), n(10(2)40)
> graph(collabels(N "N" power "Power" ma "Alternative mean"))
```

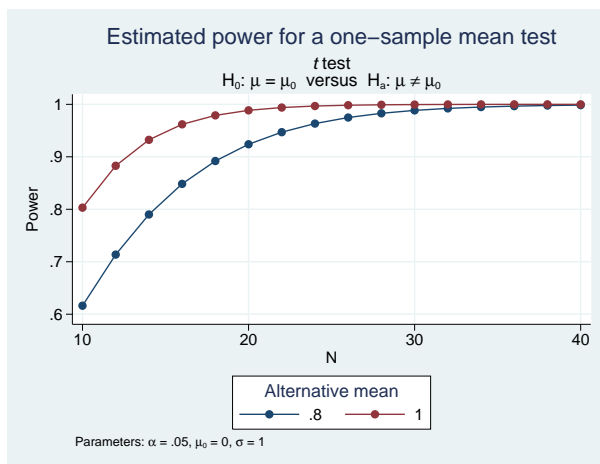


Figure 15.

For overlaid plots, we may consider alternative labeling of the plotted curves in the legend by using the `nosimplelabels` suboption (abbreviated to `nosimple`). We also suppress the legend title and request that an equality sign with a space on each side be used as a separator.



```
. power onemean 0 (0.8 1), n(10(2)40)
> graph(nosimple legend(title("")) eqsep(" = "))
```

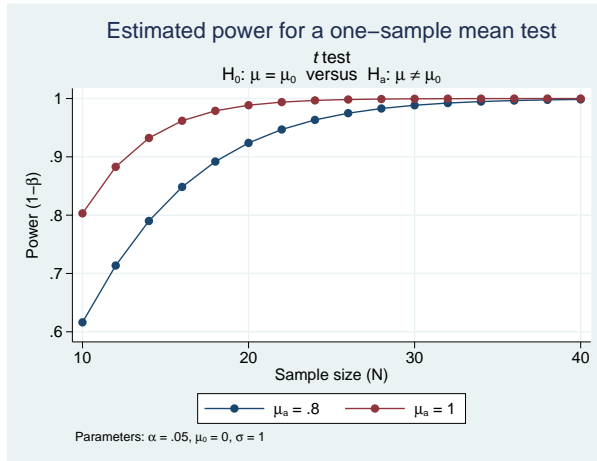


Figure 16.

◀

### ► Example 7: Plot options

We can use the `plotopts()` and `plot#opts()` suboptions within `graph()` to modify the default look of the plotted lines. If there are multiple curves, the `plotopts()` suboption will apply changes to all curves. Use the corresponding `plot#opts()` suboption to change the look of only the specific, #th, curve.

Here are a few examples of using these suboptions.

```
. power onemean 0 (0.1(0.1)1), graph(plotopts(mlabel(N) mlabpos(1)))
```

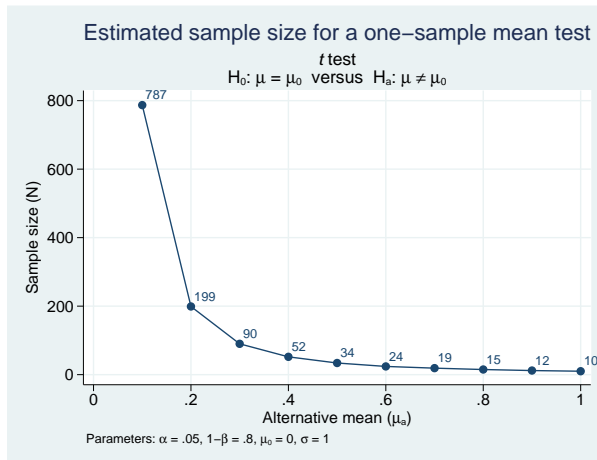


Figure 17.

We specified `mlabel()` within the `plotopts()` suboption to label each data point on the graph with its corresponding sample-size value. `mlabpos()` was used to place the marker labels at the one o'clock position.

For plots containing multiple curves such as in figure 3, the `plotopts()` suboption controls the look of all curves. For example, we can change the marker symbol from the default solid circle to a solid triangle.

```
. power onemean 0 (0.8 1), n(10(2)40) sd(1 1.5) graph(plotopts(msymbol(T)))
```

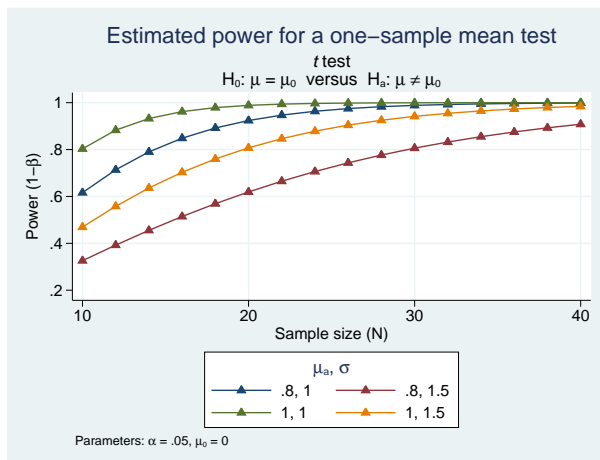


Figure 18.

To control the look of each curve, we can use multiple `plot#opts()` suboptions. For example, we can request that the curves corresponding to the same standard deviation be plotted using the same color:

```
. power onemean 0 (0.8 1), n(10(2)40) sd(1 1.5)
> graph(plot3opts(color(navy)) plot4opts(color(maroon)))
```

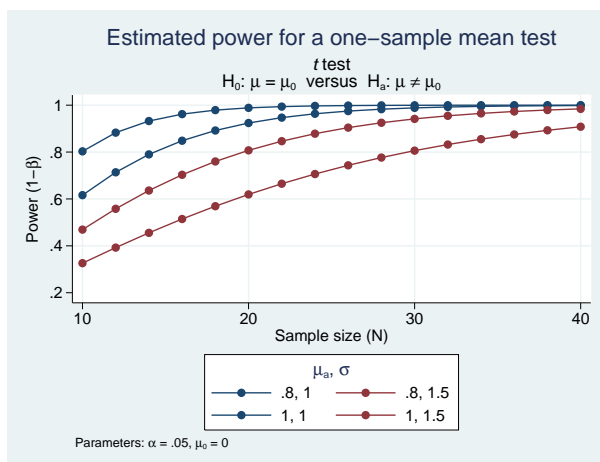


Figure 19.

### Example 8: Modifying the look of by graphs

The look of by graphs is controlled by the `byopts()` suboption specified within `power`'s `graph()` option.

For example, in [figure 10](#), we can specify `yrescale` within the `byopts()` suboption to allow the scales of the two by graphs to differ. We use the alternative means of 0.5 and 1 instead of 0.8 and 1 to demonstrate differences between scales.

```
. power onemean 0 (0.5 1), n(10(2)40) sd(1 1.5) graph(by(ma) byopts(yrescale))
```

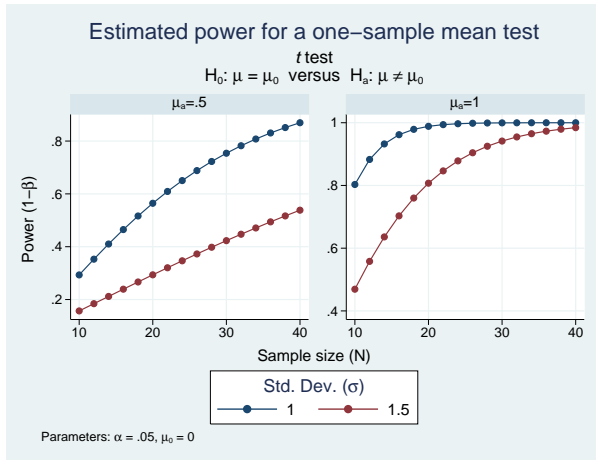


Figure 20.

We can use `byopts()` to change the overall graph title and subtitle.

```
. power onemean 0 (0.5 1), n(10(2)40) sd(1 1.5) graph(by(ma) byopts(yrescale  
> title("Power vs sample size") subtitle("")))
```

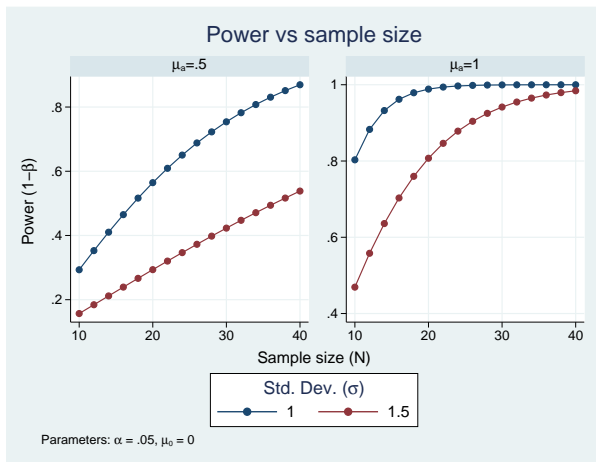


Figure 21.

Note that if you use `title()` and `subtitle()` outside `byopts()`, you will change the title and subtitle of the individual by graphs and not the overall graph.

## Parallel plots

Sometimes, you may be interested in comparing powers of parallel sets of parameters, that is, parameters that vary in parallel instead of being nested. In this situation, the results represent a collection of data points rather than a curve and are displayed on the graph as a scatterplot without connecting points.

For such parallel plots, the default display of the results on the  $y$  axis may be cumbersome. A more appealing look may be a graph that swaps the  $y$  and  $x$  axes, the horizontal graph. Such a look may be achieved by specifying the `horizontal` suboption within `graph()`.

```
. power onemean 0 (0.1(0.1)0.9), sd(1(0.1)1.9) parallel
> graph(x(ma sd) horizontal nosimple ytitle(""))
```

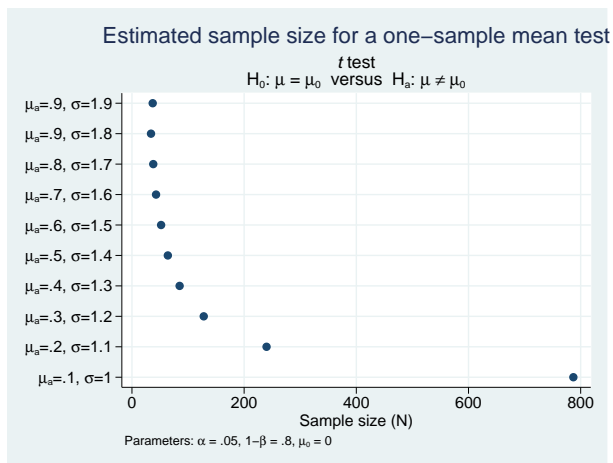


Figure 22.

To improve the look of the horizontal graph, we requested that the labels on the  $y$  axis include the parameter symbol by specifying the `nosimplelabels` suboption, and we also suppressed the  $y$ -axis title.

## Also see

[PSS] **power** — Power and sample-size analysis for hypothesis tests

[PSS] **power, table** — Produce table of results from the power command

# Title

**power, table** — Produce table of results from the power command

[Description](#)      [Menu](#)      [Syntax](#)      [Suboptions](#)  
[Remarks and examples](#)      [Stored results](#)      [Also see](#)

## Description

`power`, `table` displays results in a tabular format. `table` is implied if any of the `power` command's arguments or options contain more than one element. The `table` option is useful if you are producing graphs and would like to see the table as well or if you are producing results one case at a time using a loop and wish to display results in a table. The `notable` option suppresses table results; it is implied with the graphical output of `power`, `graph`; see [\[PSS\] power, graph](#).

## Menu

Statistics > Power and sample size

## Syntax

*Produce default table*

```
power ..., table ...
```

*Suppress table*

```
power ..., notable ...
```

*Produce custom table*

```
power ..., table( [colspec] [, tableopts] ) ...
```

where *colspec* is

```
column[:label] [column[:label] [...]]
```

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>tableopts</i>	Description
<code>add</code>	add <i>columns</i> to the default table
<code>labels(<i>labspec</i>)</code>	change default labels for specified columns; default labels are column names
<code>widths(<i>widthspec</i>)</code>	change default column widths; default is specific to each column
<code>formats(<i>fmtspec</i>)</code>	change default column formats; default is specific to each column
<code>noformat</code>	do not use default column formats
<code>separator(#)</code>	draw a horizontal separator line every # lines; default is <code>separator(0)</code> , meaning no separator lines
<code>divider</code>	draw divider lines between columns
<code>byrow</code>	display rows as computations are performed; seldom used
<code>noheader</code>	suppress table header; seldom used
<code>continue</code>	draw a continuation border in the table output; seldom used

`noheader` and `continue` are not shown in the dialog box.

<i>column</i>	Description
<code>alpha</code>	significance level
<code>power</code>	power
<code>beta</code>	type II error probability
<code>N</code>	total number of subjects
<code>N1</code>	number of subjects in the control group
<code>N2</code>	number of subjects in the experimental group
<code>nratio</code>	ratio of sample sizes, experimental to control
<code>delta</code>	effect size
<code>target</code>	target parameter
<code>_all</code>	display all supported columns
<code>method_columns</code>	columns specific to the <code>method</code> specified with <code>power</code>

By default, the following columns are displayed:

- `alpha`, `power`, and `N` are always displayed;
- `N1` and `N2` are displayed for two-sample methods;
- `delta` is displayed when defined by the method;
- additional columns specific to each `power method` may be displayed.

## Suboptions

The following are suboptions within the `table()` option of the `power` command.

### Table

`add` requests that the columns specified in *colspec* be added to the default table. The columns are added to the end of the table.

`labels(labspec)` specifies the labels to be used in the table for the specified columns. *labspec* is `column "label" [column "label" [...]]`

`labels()` takes precedence over the specification of column labels in *colspec*.

`widths` (*widthspec*) specifies column widths. The default values are the widths of the default column formats plus one. If the `noformat` option is used, the default for each column is nine. The column widths are adjusted to accommodate longer column labels and larger format widths. *widthspec* is either a list of values including missing values (*numlist*) or

```
column # [column # [...]]
```

For the value-list specification, the number of specified values may not exceed the number of columns in the table. A missing value (.) may be specified for any column to indicate the default width. If fewer widths are specified than the number of columns in the table, the last width specified is used for the remaining columns.

The alternative column-list specification provides a way to change widths of specific columns.

`formats` (*fntspec*) specifies column formats. The default is `%7.0g` for integer-valued columns and `%7.4g` for real-valued columns. *fntspec* is either a string value-list of `formats` that may include empty strings or a column list:

```
column "fmt" [column "fmt" [...]]
```

For the value-list specification, the number of specified values may not exceed the number of columns in the table. An empty string ("") may be specified for any column to indicate the default format. If fewer formats are specified than the number of columns in the table, the last format specified is used for the remaining columns.

The alternative column-list specification provides a way to change formats of specific columns.

`noformat` requests that the default formats not be applied to the column values. If this suboption is specified, the column values are based on the column width.

`separator` (#) specifies how often separator lines should be drawn between rows of the table. The default is `separator(0)`, meaning that no separator lines should be displayed.

`divider` specifies that divider lines be drawn between columns. The default is no dividers.

`byrow` specifies that table rows be displayed as computations are performed. By default, the table is displayed after all computations are performed. This suboption may be useful when the computation of each row of the table takes a long time.

The following suboptions are available but are not shown in the dialog box:

`noheader` prevents the table header from displaying. This suboption is useful when the command is issued repeatedly, such as within a loop.

`continue` draws a continuation border at the bottom of the table. This suboption is useful when the command is issued repeatedly, such as within a loop.

## Remarks and examples

Remarks are presented under the following headings:

- Using power, table*
- Default tables*
- Modifying default tables*
- Custom tables*

`power, table` displays results from the `power` command in a table. This is useful for *sensitivity analysis*, which investigates the effect of varying study parameters on power, sample size, or other components of the study. The true values of study parameters are usually unknown. PSS analysis uses best guesses for these values. It is important to evaluate the sensitivity of the computed power or

sample size to the chosen values of study parameters. For example, to evaluate variability of power values, you can compute powers for various ranges of values for the parameters of interest and display the resulting powers in a table or plot them on a graph (see [PSS] **power, graph**).

## Using power, table

If you specify the `table` option or include more than one element in command arguments or in options allowing multiple values, the `power` command displays results in a tabular form. If desired, you can suppress the table by specifying the `notable` option. The `table` option is useful if you are producing graphical output or if you are producing results one case at a time, such as within a loop, and wish to display results in a table; see [example 4](#) below.

Each method specified with the `power` command has its own default table. Among the columns that are always included in the default table are significance level (`alpha`), power (`power`), total sample size (`N`), and effect size (`delta`)—if effect size is defined by the method. For two-sample methods, the columns containing the sample sizes of the control and experimental groups, `N1` and `N2`, respectively, are also included after the total sample size.

You can build your own table by specifying the columns and, optionally, their labels in the `table()` option. You can also add columns to the default table by specifying `add` within `power`'s `table()` option. The columns are displayed in the order they are specified. Each method provides its own list of supported columns; see the description of the `table()` option for each method. You can further customize the table by specifying various suboptions within `power`'s `table()` option.

The default column labels are the column names. You can provide your own column labels in `colspec` or by specifying `table()`'s suboption `labels()`. Labels containing spaces should be enclosed in quotes, and labels containing quotes should be enclosed in compound quotes. The `labels()` suboption is useful for changing the labels of existing columns; see [example 2](#) below for details.

The default formats are `%7.4g` for real-valued columns and `%7.0g` for integer-valued columns. If the `noformat` suboption is specified, the default column widths are nine characters. You can use `formats()` to change the default column formats and `widths()` to change the default column widths. The `formats()` and `widths()` suboptions provide two alternative specifications, a value-list specification or a column-list specification. The value-list specification accepts a list of values—strings for formats and numbers for widths—corresponding to each column of the displayed table. Empty strings ("") for formats and missing values (.) for widths are allowed and denote the default values. It is an error to specify more values than the number of displayed columns. If fewer values are specified, then the last value specified is used for the remaining columns. The column-list specification includes a list of pairs containing a column name followed by the corresponding value of the format or width. This specification is useful if you want to modify the formats or the widths of only selected columns. For column labels or formats exceeding the default column width, the widths of the respective columns are adjusted to accommodate the column labels and the specified formats.

If you specify the `noformat` suboption, the default formats are ignored, and the format of a column is determined by the column width: if the column width is `#`, the displayed format is `%(# - 2).#g`. For example, if the column width is 9, the displayed format is `%7.0g`.

You may further customize the look of the table by using `separator(#)` to include separator lines after every `#` lines and by using the `divider` suboption to include divider lines between columns.

The `noheader` and `continue` suboptions are useful when you are building your own table within a loop; see [example 4](#) in *Custom tables*.

In what follows, we demonstrate the default and custom tables of the results from `power` and sample-size analysis for a two-sided 5%-level one-sample *t* test comparing the population mean with a hypothesized value; see [PSS] **power onemean**.



## Default tables

If there is only one set of results, the `power` command displays those results as text. When the `power` command has multiple sets of results, they are automatically displayed in a table. You can also specify the `table` option at any time to request that results be displayed in a table.

The displayed columns are specific to the chosen method of analysis and to the options specified with the command. The columns that always appear in the table include the significance level (`alpha`), power (`power`), and total sample size (`N`). If the concept of effect size is defined for the chosen method, the effect size (`delta`) is also displayed in the default table.

### ► Example 1: Default tables from `power onemean`

Suppose we want to explore the requirements on the sample size for a one-sample mean comparison test to detect means of different magnitudes. For simplicity, we consider only two target mean values, 1 and 2, and keep all other study parameters of `power onemean` at their default values: power at 80%, two-sided significance level at 0.05, and the estimate of the population standard deviation at 1. We use a zero null value of the mean. See [PSS] [power onemean](#) for details.

```
. power onemean 0 (1 2)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd
.05	.8	10	1	0	1	1
.05	.8	5	2	0	2	1

As we mentioned earlier, the `alpha`, `power`, `N`, and `delta` columns are usually displayed in the default table. The `power onemean` command additionally displays columns containing the null mean, the alternative mean, and the standard deviation.

If we need to account for finite population, we can specify the `fpc()` option with `power onemean`. The default table will contain an additional column, `fpc`.

```
. power onemean 0 (1 2), fpc(500)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd	fpc
.05	.8	10	1	0	1	1	500
.05	.8	5	2	0	2	1	500

If there is only one set of results, the `power` command displays those results as text. If desired, you can request a table by specifying the `table` option.

```
. power onemean 0 1, table
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd
.05	.8	10	1	0	1	1

◀

## Modifying default tables

We can modify labels, widths, and formats of the default columns by specifying the corresponding suboptions within the `table()` option. We can also add columns to the default table by using `table()`'s suboption `add`.

### ► Example 2: Modifying default tables from `power onemean`

We can change the default labels of all or selected columns by using the `labels()` suboption within `power`'s `table()` option. For example, we can change the labels of the sample-size columns and standard deviation columns of the first table in [example 1](#) to “Sample size” and “Std. Dev.”, respectively.

```
. power onemean 0 (1 2), table(, labels(N "Sample size" sd "Std. Dev."))
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	Sample size	delta	m0	ma	Std. Dev.
.05	.8	10	1	0	1	1
.05	.8	5	2	0	2	1

We can also change default column formats and widths by using the `formats()` and `widths()` suboptions.

```
. power onemean 0 (1 2), n(5) table(, labels(N "Sample size" sd "Std. Dev."))
> widths(N 14 sd 14) formats(power "%7.5f")
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	Sample size	delta	m0	ma	Std. Dev.
.05	0.40139	5	1	0	1	1
.05	0.90888	5	2	0	2	1

For this table, we switched from a sample-size determination to power determination by specifying the `n()` option. We changed the default column widths of the sample-size columns and standard deviation columns to 14. We also changed the default `%7.4g` format of the power column to `%7.5f`.

We can add columns to the default table by listing the names of the columns in the `table()` option and specifying its suboption `add`.

```
. power onemean 0 (1 2), table(diff, add)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd	diff
.05	.8	10	1	0	1	1	1
.05	.8	5	2	0	2	1	2

We added the column `diff`, which contains the difference between the alternative and hypothesized values of the mean, to the default table produced by `power onemean`.

◀

## Custom tables

We can use the `table()` option to build custom tables, which contain the columns you want in the order you want. You can also build a table within a `foreach` or `forvalues` loop as we demonstrate in [example 4](#) below. This is useful in the case when you want to obtain multiple sets of results over parameters of the `power` command that do not allow the `numlist` specification.

### ▶ Example 3: Producing custom tables

As an example of a custom table, we produce a table containing only four columns: significance level, power, sample size, and effect size.

```
. power onemean 0 (1 2), table(alpha:"Significance level"
> power:Power N:"Sample size" delta:"Effect size", widths(. 15))
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

Significance level	Power	Sample size	Effect size
.05	.8	10	1
.05	.8	5	2

To improve the look of the table, we also specified the `widths(. 15)` suboption to increase the column widths of the last 3 columns to 15, leaving the width of the first column, the significance level, at its default value.

We can use the `_all` qualifier to request that all table columns supported by the method be displayed.

```
. power onemean 0 (1 2), table(_all)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	beta	N	delta	m0	ma	diff	sd
.05	.8	.2	10	1	0	1	1	1
.05	.8	.2	5	2	0	2	2	1

◀

#### ► Example 4: Building table using a loop

Some options of `power` commands may not allow the *numlist* specification. In this case, you can build a table manually by using a loop such as `foreach` (see [P] **foreach**) or `forvalues` (see [P] **forvalues**) loop. One way to do this is to write a program that loops over parameters of interest. We demonstrate a program that loops over varying values of the alternative mean of `power onemean`. You can easily adapt this program to meet your needs.

```
program dotable
  args ma
  numlist "ma" // expand the numeric list in macro ma
  local ma "r(numlist)"
  local nvals : list sizeof ma
  local i 1
  foreach val of local ma { // loop over numeric values in ma
    if ('i'==1) {
      power onemean 0 'val', table(, continue)
    }
    else if ('i'<'nvals') {
      power onemean 0 'val', table(, noheader continue) notitle
    }
    else {
      power onemean 0 'val', table(, noheader) notitle
    }
    local ++i
  }
end
```

The `dotable` program accepts one argument, `ma`, which may contain one or more numeric values of the alternative mean specified as a *numlist*. The program uses combinations of `continue`, `noheader`, and `notitle` to display a table. The first call to `power onemean` requests that the table be displayed without the bottom line by specifying the `continue` suboption within `table()`. The subsequent calls (except the last) specify the `continue` suboption, the `notitle` option with `power onemean`, and `noheader` within the `table()` option to request that neither the output before the table nor the table header be displayed. The last call omits the `continue` suboption so that the bottom line is displayed.

As a result, we obtain the following table:

```
. dotable "1(1)4"
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd
.05	.8	10	1	0	1	1
.05	.8	5	2	0	2	1
.05	.8	4	3	0	3	1
.05	.8	3	4	0	4	1

◀

## Stored results

`power`, `table` stores the following in `r()` in addition to other results stored by `power`:

### Scalars

`r(separator)` number of lines between separator lines in the table  
`r(divider)` 1 if `divider` is requested in the table; 0 otherwise

### Macros

`r(columns)` displayed table columns  
`r(labels)` table column labels  
`r(widths)` table column widths  
`r(formats)` table column formats

### Matrices

`r(pss_table)` table of results

## Also see

[PSS] [power](#) — Power and sample-size analysis for hypothesis tests

[PSS] [power, graph](#) — Graph results from the `power` command

# Title

**power onemean** — Power analysis for a one-sample mean test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power onemean` computes sample size, power, or target mean for a one-sample mean test. By default, it computes sample size for given power and the values of the mean parameters under the null and alternative hypotheses. Alternatively, it can compute power for given sample size and values of the null and alternative means or the target mean for given sample size, power, and the null mean. Also see [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for two-sided test of  $H_0: \mu = 10$  versus  $H_a: \mu \neq 10$  with null mean  $m_0 = 10$ , alternative mean  $m_a = 15$ , and standard deviation of 12 using default power of 0.8 and significance level  $\alpha = 0.05$

```
power onemean 10 15, sd(12)
```

As above, but for a one-sided test with power of 0.9

```
power onemean 10 15, sd(12) power(.9) onesided
```

Same as above, but specified as  $m_0$  and difference  $m_a - m_0 = 5$

```
power onemean 10, sd(12) power(.9) onesided diff(5)
```

Power for a sample size of 75

```
power onemean 10 15, sd(12) n(75)
```

Power for sample sizes of 50, 60, 70, and 80

```
power onemean 10 15, sd(12) n(50(10)80)
```

As above, but display results in a graph of power versus sample size

```
power onemean 10 15, sd(12) n(50(10)80) graph
```

Effect size and target mean for  $m_0 = 10$  with standard deviation of 4, for a sample size of 40, power of 0.9, and  $\alpha = 0.01$

```
power onemean 10, sd(4) n(40) power(.9) alpha(.01)
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power onemean  $m_0$   $m_a$  [ , power(numlist) options ]
```

Compute power

```
power onemean  $m_0$   $m_a$  , n(numlist) [options ]
```

Compute effect size and target mean

```
power onemean  $m_0$  , n(numlist) power(numlist) [options ]
```

where  $m_0$  is the null (hypothesized) mean or the value of the mean under the null hypothesis, and  $m_a$  is the alternative (target) mean or the value of the mean under the alternative hypothesis.  $m_0$  and  $m_a$  may each be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 **numlist**).

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample size
* <u>diff</u> ( <i>numlist</i> )	difference between the alternative mean and the null mean, $m_a - m_0$ ; specify instead of the alternative mean $m_a$
* <u>sd</u> ( <i>numlist</i> )	standard deviation; default is <code>sd(1)</code>
<u>knownsd</u>	request computation assuming a known standard deviation; default is to assume an unknown standard deviation
* <u>fpc</u> ( <i>numlist</i> )	finite population correction (FPC) as a sampling rate or as a population size
<u>direction</u> ( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [ , <u>replace</u> ])	save the table data to <i>filename</i> ; use <u>replace</u> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
Iteration	
<u>init</u> (#)	initial value for sample size or mean; default is to use normal approximation
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
[ <u>no</u> ] <u>log</u>	suppress or display iteration log
[ <u>no</u> ] <u>dots</u>	suppress or display iterations as dots
<u>notitle</u>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] **11.1.8 numlist**.

`notitle` does not appear in the dialog box.

where *tablespec* is

`column` [ :*label* ] [ `column` [ :*label* ] [ ... ] ] [ , *tableopts* ]

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).



<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	number of subjects	$N$
<code>delta</code>	effect size	$\delta$
<code>m0</code>	null mean	$\mu_0$
<code>ma</code>	alternative mean	$\mu_a$
<code>diff</code>	difference between the alternative and null means	$\mu_a - \mu_0$
<code>sd</code>	standard deviation	$\sigma$
<code>fpc</code>	FPC as population size	$N_{\text{pop}}$
	FPC as sampling rate	$\gamma$
<code>target</code>	target parameter; synonym for <code>ma</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `diff` and `fpc` are shown in the default table if specified.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] [power](#). The `nfractional` option is allowed only for sample-size determination.

`diff(numlist)` specifies the difference between the alternative mean and the null mean,  $m_a - m_0$ . You can specify either the alternative mean  $m_a$  as a command argument or the difference between the two means in `diff()`. If you specify `diff(#)`, the alternative mean is computed as  $m_a = m_0 + \#$ . This option is not allowed with the effect-size determination.

`sd(numlist)` specifies the sample standard deviation or the population standard deviation. The default is `sd(1)`. By default, `sd()` specifies the sample standard deviation. If `knownsd` is specified, `sd()` specifies the population standard deviation.

`knownsd` requests that the standard deviation be treated as known in the computation. By default, the standard deviation is treated as unknown, and the computation is based on a  $t$  test, which uses a Student's  $t$  distribution as a sampling distribution of the test statistic. If `knownsd` is specified, the computation is based on a  $z$  test, which uses a normal distribution as the sampling distribution of the test statistic.

`fpc(numlist)` requests that a finite population correction be used in the computation. If `fpc()` has values between 0 and 1, it is interpreted as a sampling rate,  $n/N$ , where  $N$  is the total number of units in the population. If `fpc()` has values greater than  $n$ , it is interpreted as a population size. It is an error for `fpc()` to have values between 1 and  $n$  or to have a mixture of sampling rates and population sizes.

`direction()`, `onesided`, `parallel`; see [PSS] [power](#).

### Table

`table`, `table()`, `notable`; see [PSS] [power](#), [table](#).

`saving()`; see [PSS] [power](#).

## Graph

`graph`, `graph()`; see [PSS] **power**, **graph**. Also see the *column* table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value of the sample size for the sample-size determination or the initial value of the mean for the effect-size determination. The default is to use a closed-form normal approximation to compute an initial value of the sample size or mean.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power onemean` but is not shown in the dialog box:

`notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

- Introduction*
- Using power onemean*
- Computing sample size*
- Computing power*
- Computing effect size and target mean*
- Performing hypothesis tests on mean*
- Video examples*

This entry describes the `power onemean` command and the methodology for power and sample-size analysis for a one-sample mean test. See [PSS] **intro** for a general introduction to power and sample-size analysis, and see [PSS] **power** for a general introduction to the `power` command using hypothesis tests.

## Introduction

There are many examples of studies where a researcher would like to compare an observed mean with a hypothesized mean. A company that provides preparatory classes for a standardized exam would like to see if the mean score of students who take its classes is higher than the national average. A fitness center would like to know if its average clients' weight loss is greater than zero after six months. Or a government agency would like to know if a job training program results in higher wages than the national average.

This entry describes power and sample-size analysis for the inference about the population mean performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: \mu = \mu_0$  versus the two-sided alternative hypothesis  $H_a: \mu \neq \mu_0$ , the upper one-sided alternative  $H_a: \mu > \mu_0$ , or the lower one-sided alternative  $H_a: \mu < \mu_0$ .

The considered one-sample mean tests rely on the assumption that a random sample is normally distributed or that the sample size is large. Different test statistics can be based on whether the variance of the sampling process is known a priori. In the case of a known variance, the test statistic follows a standard normal distribution under the null hypothesis, and the corresponding test is known as a  $z$  test. In the case of an unknown variance, an estimate of the population variance is used to form a test statistic, which follows a Student's  $t$  distribution under the null hypothesis, and the corresponding test is known as a  $t$  test.

The random sample is typically drawn from an infinite population. When the sample is drawn from a population of a fixed size, sampling variability must be adjusted for a finite population size.

The `power onemean` command provides power and sample-size analysis for the comparison of a mean with a reference value using a  $t$  test or a  $z$  test.

## Using power onemean

`power onemean` computes sample size, power, or target mean for a one-sample mean test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

By default, all computations are based on a  $t$  test, which assumes an unknown standard deviation, and use the default value of 1 as the estimate of the standard deviation. You may specify other values for the standard deviation in the `sd()` option. For a known standard deviation, you can specify the `knownsd` option to request a  $z$  test.

To compute sample size, you must specify the means under the null and alternative hypotheses,  $m_0$  and  $m_a$ , respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the sample size in the `n()` option and the means under the null and alternative hypotheses,  $m_0$  and  $m_a$ , respectively.

Instead of the alternative mean,  $m_a$ , you may specify the difference  $m_a - m_0$  between the alternative mean and the null mean in the `diff()` option when computing sample size or power.

To compute effect size, the standardized difference between the alternative and null means, and the corresponding target mean, you must specify the sample size in the `n()` option, the power in the `power()` option, the null mean  $m_0$ , and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the target mean is assumed to be larger than the specified null mean value. This is also equivalent to the assumption of a positive effect size. You can change the direction to lower, which means that the target mean is assumed to be smaller than the specified null value, by specifying the `direction(lower)` option. This is equivalent to assuming a negative effect size.

By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see [Fractional sample sizes](#) in [PSS] [unbalanced designs](#) for an example. The `nfractional` option is allowed only for sample-size determination.

Some of `power onemean`'s computations require iteration. For example, when the standard deviation is unknown, computations use a noncentral Student's  $t$  distribution. Its degrees of freedom depends on the sample size, and the noncentrality parameter depends on the sample size and effect size. Therefore, the sample-size and effect-size determinations require iteration. The default initial values of the estimated parameters are obtained by using a closed-form normal approximation. They may be changed by specifying the `init()` option. See [PSS] [power](#) for the descriptions of other options that control the iteration procedure.

All computations assume an infinite population. For a finite population, use the `fpc()` option to specify a sampling rate or a population size. When this option is specified, a finite population correction is applied to the population standard deviation. The correction factor depends on the sample size; therefore, computing sample size for a finite population requires iteration even for a known standard deviation. The initial value for the sample size is based on the corresponding sample size assuming an infinite population.

In the following sections, we describe the use of `power onemean` accompanied by examples for computing sample size, power, and target mean.

## Computing sample size

To compute sample size, you must specify the means under the null and alternative hypotheses,  $m_0$  and  $m_a$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a one-sample mean test

Consider an example from [Tamhane and Dunlop \(2000, 209\)](#) that discusses the effectiveness of coaching programs in improving the verbal part of SAT scores. Previous studies found that students retaking the SAT exams without any coaching program improve their scores by 15 points on average with a standard deviation of about 40 points. A new coaching program claims to improve the SAT scores by 40 points above the average. The changes in scores are assumed to be approximately normally distributed. The parameter of interest in this example is the mean change in the test scores. To test the claim, investigators wish to conduct another study and compute the sample size that is required to detect a mean change in scores of 40 points with 80% power using a 5%-level two-sided test. We assume that the true population standard deviation is unknown and use its estimate from previous studies to compute the sample size:

```
. power onemean 15 40, sd(40)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.6250
      m0 =    15.0000
      ma =    40.0000
      sd =    40.0000
Estimated sample size:
      N =          23
```

We find that a sample of 23 subjects is required to detect a shift of 40 points in average SAT scores given the standard deviation of 40 points with 80% power using a 5%-level two-sided test.

As we mentioned in [Using `power onemean`](#) and as is also indicated in the output, sample-size computation requires iteration when the standard deviation is unknown. The iteration log is suppressed by default, but you can display it by specifying the `log` option.

◀

### ► Example 2: Specifying difference between means

Instead of the alternative mean change of 40 as in [example 1](#), we can specify the difference of 25 between the mean changes in scores under the alternative and null hypotheses in the `diff()` option and obtain the same results.

```

. power onemean 15, diff(25) sd(40)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.6250
    m0 =     15.0000
    ma =     40.0000
    diff =    25.0000
    sd =     40.0000
Estimated sample size:
    N =         23

```

When we specify the `diff()` option, the difference between the alternative and null values is also reported in the output.



### ▷ Example 3: Known variance

If we know the population standard deviation, we can use the `knownsd` option to request a  $z$  test.

```

. power onemean 15 40, sd(40) knownsd
Performing iteration ...
Estimated sample size for a one-sample mean test
z test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.6250
    m0 =     15.0000
    ma =     40.0000
    sd =     40.0000
Estimated sample size:
    N =         21

```

The output now indicates that the computation is based on a  $z$  test instead of a  $t$  test. We find that a smaller sample of 21 subjects is required to detect the same effect size as in [example 1](#) when the standard deviation is known.



## Computing power

To compute power, you must specify the sample size in the `n()` option and the means under the null and alternative hypotheses,  $m_0$  and  $m_a$ , respectively.

### ▷ Example 4: Power of a one-sample mean test

Continuing [example 1](#), we will suppose that we are designing a new study and anticipate to obtain a sample of 30 subjects. To compute the power corresponding to this sample size given the study parameters from example 1, we specify the sample size of 30 in the `n()` option:

```
. power onemean 15 40, n(30) sd(40)
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha =    0.0500
      N    =     30
    delta =    0.6250
     m0    =   15.0000
     ma    =   40.0000
     sd    =   40.0000
Estimated power:
    power =    0.9112
```

With a larger sample size, the power of the test increases to about 91.12%.



### ▷ Example 5: Multiple values of study parameters

To investigate the effect of a finite population size on power, we can specify a list of population sizes in the `fpc()` option:

```
. power onemean 15 40, n(30) sd(40) fpc(100 500 1000)
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
```

alpha	power	N	delta	m0	ma	sd	fpc
.05	.9769	30	.625	15	40	40	100
.05	.9267	30	.625	15	40	40	500
.05	.919	30	.625	15	40	40	1000

As expected, when the population size increases, the power tends to get closer to that obtained by assuming an infinite population size.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).



### ▷ Example 6: Reproducing published results from a text book

We can also reproduce the example from [Tamhane and Dunlop \(2000, 213–214\)](#). The authors consider a one-sided test with a 0.132 significance level and a known standard deviation and compute the power to be 95.3%. We can replicate their example by typing

```
. power onemean 15 40, n(20) sd(40) alpha(0.132) onesided knownsd
Estimated power for a one-sample mean test
z test
Ho: m = m0 versus Ha: m > m0
Study parameters:
    alpha =    0.1320
      N =      20
    delta =    0.6250
      m0 =    15.0000
      ma =    40.0000
      sd =    40.0000
Estimated power:
    power =    0.9533
```

◀

## Computing effect size and target mean

Effect size  $\delta$  for a one-sample mean test is defined as the ratio of the difference between the alternative and null values of the mean to the standard deviation,  $\delta = (\mu_a - \mu_0)/\sigma$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding alternative or target mean that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and null mean must be specified. In addition, you must also decide on the direction of the effect: upper, which means  $\mu_a > \mu_0$ , or lower, which means  $\mu_a < \mu_0$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ▶ Example 7: Minimum detectable value of the mean change in SAT scores

Continuing with [example 4](#), we may also be interested to find the smallest mean change in SAT scores that can be detected with a power of 80% given a sample of 30 subjects. To compute this, we specify only the null value of 15 as the command argument and also specify the sample size and power in the `n(30)` and `power(0.8)` options, respectively. We use the same value of 40 for the standard deviation as in [example 4](#).

```
. power onemean 15, n(30) power(0.8) sd(40)
Performing iteration ...
Estimated target mean for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0; ma > m0
Study parameters:
    alpha =    0.0500
    power =    0.8000
      N =      30
      m0 =    15.0000
      sd =    40.0000
Estimated effect size and target mean:
    delta =    0.5292
      ma =    36.1694
```

The estimated smallest mean change in SAT scores is 36.17, which corresponds to the effect size of 0.53. Compared with [example 1](#), for the same power of 80%, this example shows a smaller difference between the mean SAT scores of the two programs for a larger sample of 30 subjects.

In the above, we assumed the effect to be in the upper direction. By symmetry, the effect size in the lower direction will be  $-0.53$ , which can also be obtained by specifying `direction(lower)` in the above example.

◀

## Performing hypothesis tests on mean

In this section, we briefly demonstrate the use of the `ttest` command for testing hypotheses about means; see [R] [ttest](#) for details. Suppose we wish to test the hypothesis that the mean is different from a reference value on the collected data. We can use the `ttest` command to do this. Below we demonstrate the use of this command for the analysis of `sat.dta`.

### ▶ Example 8: Testing for mean

Suppose that we wish to test whether the mean verbal SAT score is equal to 600. We use the `ttest` command to do this as follows:

```
. use http://www.stata-press.com/data/r14/sat
(Artificial SAT data)
```

```
. ttest score == 600
```

```
One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
score	75	504.8	15.24616	132.0356	474.4214	535.1786

```
mean = mean(score)                                t = -6.2442
Ho: mean = 600                                    degrees of freedom = 74
Ha: mean < 600                                    Ha: mean != 600      Ha: mean > 600
Pr(T < t) = 0.0000                                Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000
```

We find statistical evidence to reject the null hypothesis of  $H_0: \mu_{SAT} = 600$  versus a two-sided alternative  $H_a: \mu_{SAT} \neq 600$  at the 5% significance level; the  $p$ -value  $< 0.0000$ .

We use the estimates based on this study to perform a sample-size analysis we would have conducted before the study.

```
. power onemean 600 505, sd(132)
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
alpha = 0.0500
power = 0.8000
delta = -0.7197
m0 = 600.0000
ma = 505.0000
sd = 132.0000
Estimated sample size:
N = 18
```

We find that the sample size required to detect a mean score of 505 with 80% power using a 5%-level two-sided test is only 18. The current sample contains 75 subjects, which would allow us to detect a potentially smaller (in absolute value) difference between the alternative and null means.

◀



## Video examples

Sample-size calculation for comparing a sample mean to a reference value

Power calculation for comparing a sample mean to a reference value

Minimum detectable effect size for comparing a sample mean to a reference value

## Stored results

`power onemean` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(m0)</code>	mean under the null hypothesis
<code>r(ma)</code>	mean under the alternative hypothesis
<code>r(diff)</code>	difference between the alternative and null means
<code>r(sd)</code>	standard deviation
<code>r(knownsd)</code>	1 if option <code>knownsd</code> is specified; 0 otherwise
<code>r(fpc)</code>	finite population correction (if specified)
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or mean
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>onemean</code>
<code>r(direction)</code>	<code>upper</code> or <code>lower</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Let  $x_1, \dots, x_n$  be a sequence of  $n$  independent and identically distributed random variables drawn from a normal population with mean  $\mu$  and variance  $\sigma^2$ . Let

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

denote the sample mean and the sample variance, respectively. Let  $\mu_0$  and  $\mu_a$  denote the null and alternative values of the mean parameter, respectively.

A one-sample mean test involves testing the null hypothesis  $H_0: \mu = \mu_0$  versus the two-sided alternative hypothesis  $H_a: \mu \neq \mu_0$ , the upper one-sided alternative  $H_a: \mu > \mu_0$ , or the lower one-sided alternative  $H_a: \mu < \mu_0$ .

If the `nfractional` option is not specified, the computed sample size is rounded up.

The following formulas are based on [Chow, Shao, and Wang \(2008\)](#).

Methods and formulas are presented under the following headings:

- Known standard deviation*
- Unknown standard deviation*
- Finite population size*

## Known standard deviation

In the case of a known standard deviation, the sampling distribution of the test statistic  $z = \sqrt{n}(\bar{x} - \mu_0)/\sigma$  under the null hypothesis follows the standard normal distribution, and the corresponding test is known as a  $z$  test.

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha}$  and  $z_\beta$  be the  $(1 - \alpha)$ th and the  $\beta$ th quantiles of the standard normal distribution.

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi(\sqrt{n}\delta - z_{1-\alpha}) & \text{for an upper one-sided test} \\ \Phi(-\sqrt{n}\delta - z_{1-\alpha}) & \text{for a lower one-sided test} \\ \Phi(\sqrt{n}\delta - z_{1-\alpha/2}) + \Phi(-\sqrt{n}\delta - z_{1-\alpha/2}) & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $\Phi(\cdot)$  is the cdf of the standard normal distribution and  $\delta = (\mu_a - \mu_0)/\sigma$  is the effect size.

The sample size  $n$  for a one-sided test is computed by inverting a one-sided power equation from (1):

$$n = \left( \frac{z_{1-\alpha} - z_\beta}{\delta} \right)^2 \quad (2)$$

Similarly, the absolute value of the effect size for a one-sided test is computed as follows:

$$|\delta| = \frac{(z_{1-\alpha} - z_\beta)}{\sqrt{n}} \quad (3)$$

Note that the magnitude of the effect size is the same regardless of the direction of the test.

The minimum detectable value of the mean for a one-sided test is computed as  $\mu_a = \mu_0 + (z_{1-\alpha} - z_\beta)\sigma/\sqrt{n}$  when  $\mu_a > \mu_0$  and as  $\mu_a = \mu_0 - (z_{1-\alpha} - z_\beta)\sigma/\sqrt{n}$  when  $\mu_a < \mu_0$ .

Sample size and minimum detectable value of the mean for a two-sided test are computed iteratively using the two-sided power equation from (1). The initial values are obtained from (2) and (3), correspondingly, with  $\alpha/2$  in place of  $\alpha$ .

## Unknown standard deviation

In the case of an unknown standard deviation, an unbiased estimator  $s$  is used in place of  $\sigma$  in the definition of a  $z$  test statistic. The sampling distribution of the test statistic  $t = \sqrt{n}(\bar{x} - \mu_0)/s$  under the null hypothesis follows a Student's  $t$  distribution with  $n - 1$  degrees of freedom, and the corresponding test is known as a  $t$  test.

Let  $t_{n-1,\alpha}$  denote the  $\alpha$ th quantile of a Student's  $t$  distribution with  $n - 1$  degrees of freedom. Under the alternative hypothesis, the test statistic follows a noncentral Student's  $t$  distribution, and the power is computed using

$$\pi = \begin{cases} 1 - T_{n-1,\lambda}(t_{n-1,1-\alpha}) & \text{for an upper one-sided test} \\ T_{n-1,\lambda}(-t_{n-1,1-\alpha}) & \text{for a lower one-sided test} \\ 1 - T_{n-1,\lambda}(t_{n-1,1-\alpha/2}) + T_{n-1,\lambda}(-t_{n-1,1-\alpha/2}) & \text{for a two-sided test} \end{cases} \quad (4)$$

where  $T_{n-1,\lambda}(\cdot)$  is the cumulative noncentral Student's  $t$  distribution with a noncentrality parameter  $\lambda = \sqrt{n}\delta$ .

Sample size and minimum detectable value of the mean are obtained by iteratively solving nonlinear equations in (4), for  $n$  and  $\delta$ , respectively. The default initial values for the iterative procedure are calculated from (2) and (3), respectively, assuming a normal distribution.

## Finite population size

The above formulas assume that the random sample is drawn from an infinite population. In cases when the size of the population is known, we need to make the following adjustment to the standard deviation,

$$\sigma_{\text{fpc}} = \sigma \sqrt{\left(1 - \frac{n}{N}\right)}$$

where  $\sigma_{\text{fpc}}$  is the population standard deviation adjusted for finite population size. The correction factor depends on the sample size; therefore, computing sample size in this case requires iteration. The initial value for the sample size is based on the corresponding normal approximation with infinite population size.

## References

- Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.
- Tamhane, A. C., and D. D. Dunlop. 2000. *Statistics and Data Analysis: From Elementary to Intermediate*. Upper Saddle River, NJ: Prentice Hall.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **ttest** —  $t$  tests (mean-comparison tests)

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power twomeans` computes sample size, power, or the experimental-group mean for a two-sample means test. By default, it computes sample size for the given power and the values of the control-group and experimental-group means. Alternatively, it can compute power for given sample size and values of the control-group and experimental-group means or the experimental-group mean for given sample size, power, and the control-group mean. Also see [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \mu_1 = \mu_2$  versus  $H_a: \mu_1 \neq \mu_2$  given alternative control-group mean  $m_1 = 8$  and alternative experimental-group mean  $m_2 = 12$  with shared standard deviation of 9 using default power of 0.8 and significance level  $\alpha = 0.05$

```
power twomeans 8 12, sd(9)
```

As above, but for  $m_2$  equal to 10, 11, 12, 13, and 14

```
power twomeans 8 (10(1)14), sd(9)
```

As above, but display results in a graph of sample size versus  $m_2$

```
power twomeans 8 (10(1)14), sd(9) graph
```

As above, but specify different standard deviations  $s_1 = 7$  and  $s_2 = 10$

```
power twomeans 8 (10(1)14), sd1(7) sd2(10) graph
```

Sample size for one-sided test with power of 0.9

```
power twomeans 8 12, sd(9) power(.9) onesided
```

Same as above, specified as  $\mu_1$  and difference between means  $m_2 - m_1 = 4$

```
power twomeans 8, sd(9) power(.9) onesided diff(4)
```

Power for a total sample size of 74 with balanced group sizes

```
power twomeans 8 12, sd(9) n(74)
```

As above, but for sample sizes of 45 and 30 in groups 1 and 2, respectively

```
power twomeans 8 12, sd(9) n1(45) n2(30)
```

Effect size and target mean difference for a sample size of 200 with power of 0.8

```
power twomeans 8, sd(9) power(.8) n(200)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

```
power twomeans  $m_1$   $m_2$  [ , power(numlist) options ]
```

*Compute power*

```
power twomeans  $m_1$   $m_2$  , n(numlist) [options ]
```

*Compute effect size and experimental-group mean*

```
power twomeans  $m_1$  , n(numlist) power(numlist) [options ]
```

where  $m_1$  is the mean of the control (reference) group, and  $m_2$  is the mean of the experimental (comparison) group.  $m_1$  and  $m_2$  may each be specified either as one number or as a list of values in parentheses (see [U] [11.1.8 numlist](#)).

<i>options</i>	Description
Main	
* <u>a</u> lpha( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>p</u> ower( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>b</u> eta( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
* <u>n</u> 1( <i>numlist</i> )	sample size of the control group
* <u>n</u> 2( <i>numlist</i> )	sample size of the experimental group
* <u>n</u> ratio( <i>numlist</i> )	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<u>compute</u> ( <u>n</u> 1   <u>n</u> 2)	solve for $N1$ given $N2$ or for $N2$ given $N1$
<u>n</u> fractional	allow fractional sample sizes
* <u>d</u> iff( <i>numlist</i> )	difference between the experimental-group mean and the control-group mean, $m_2 - m_1$ ; specify instead of the experimental-group mean $m_2$
* <u>s</u> d( <i>numlist</i> )	common standard deviation of the control and the experimental groups assuming equal standard deviations in both groups; default is <code>sd(1)</code>
* <u>s</u> d1( <i>numlist</i> )	standard deviation of the control group; requires <code>sd2()</code>
* <u>s</u> d2( <i>numlist</i> )	standard deviation of the experimental group; requires <code>sd1()</code>
<u>known</u> sds	request computation assuming known standard deviations for both groups; default is to assume unknown standard deviations
<u>d</u> irection( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>one</u> sided	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>s</u> aving( <i>filename</i> [ , <u>replace</u> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
Iteration	
<u>init</u> (#)	initial value for sample sizes or experimental-group mean

---

<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no] log</code>	suppress or display iteration log
<code>[no] dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

---

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 [numlist](#).

`notitle` does not appear in the dialog box.

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N1</code>	number of subjects in the control group	$N_1$
<code>N2</code>	number of subjects in the experimental group	$N_2$
<code>nratio</code>	ratio of sample sizes, experimental to control	$N_2/N_1$
<code>delta</code>	effect size	$\delta$
<code>m1</code>	control-group mean	$\mu_1$
<code>m2</code>	experimental-group mean	$\mu_2$
<code>diff</code>	difference between the control-group mean and the experimental-group mean	$\mu_2 - \mu_1$
<code>sd</code>	common standard deviation	$\sigma$
<code>sd1</code>	control-group standard deviation	$\sigma_1$
<code>sd2</code>	experimental-group standard deviation	$\sigma_2$
<code>target</code>	target parameter; synonym for <code>m2</code>	
<code>_all</code>	display all supported columns	

---

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `nratio`, `diff`, `sd`, `sd1`, and `sd2` are shown in the default table if specified.

## Options

Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `compute()`, `nfractional`; see [PSS] [power](#).

`diff` ([numlist](#)) specifies the difference between the experimental-group mean and the control-group mean,  $m_2 - m_1$ . You can specify either the experimental-group mean  $m_2$  as a command argument

or the difference between the two means in `diff()`. If you specify `diff(#)`, the experimental-group mean is computed as  $m_2 = m_1 + \#$ . This option is not allowed with the effect-size determination.

`sd(numlist)` specifies the common standard deviation of the control and the experimental groups assuming equal standard deviations in both groups. The default is `sd(1)`.

`sd1(numlist)` specifies the standard deviation of the control group. If you specify `sd1()`, you must also specify `sd2()`.

`sd2(numlist)` specifies the standard deviation of the experimental group. If you specify `sd2()`, you must also specify `sd1()`.

`knownsds` requests that standard deviations of each group be treated as known in the computations.

By default, standard deviations are treated as unknown, and the computations are based on a two-sample  $t$  test, which uses a Student's  $t$  distribution as a sampling distribution of the test statistic. If `knownsds` is specified, the computation is based on a two-sample  $z$  test, which uses a normal distribution as the sampling distribution of the test statistic.

`direction()`, `onesided`, `parallel`; see [PSS] **power**.

#### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

#### Graph

`graph`, `graph()`; see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

#### Iteration

`init(#)` specifies the initial value for the estimated parameter. For sample-size determination, the estimated parameter is either the control-group size  $n_1$  or, if `compute(n2)` is specified, the experimental-group size  $n_2$ . For the effect-size determination, the estimated parameter is the experimental-group mean  $m_2$ . The default initial values for a two-sided test are obtained as a closed-form solution for the corresponding one-sided test with the significance level  $\alpha/2$ . The default initial values for the  $t$  test computations are based on the corresponding large-sample normal approximation.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power twomeans` but is not shown in the dialog box:

`notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power twomeans*

*Computing sample size*

*Computing power*

*Computing effect size and experimental-group mean*

*Testing a hypothesis about two independent means*



This entry describes the `power twomeans` command and the methodology for power and sample-size analysis for a two-sample means test. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

The analysis of means is one of the most commonly used approaches in a wide variety of statistical studies. Many applications lead to the study of two independent means, such as studies comparing the average mileage of foreign and domestic cars, the average SAT scores obtained from two different coaching classes, the average yields of a crop due to a certain fertilizer, and so on. The two populations of interest are assumed to be independent.

This entry describes power and sample-size analysis for the inference about two population means performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: \mu_2 = \mu_1$  versus the two-sided alternative hypothesis  $H_a: \mu_2 \neq \mu_1$ , the upper one-sided alternative  $H_a: \mu_2 > \mu_1$ , or the lower one-sided alternative  $H_a: \mu_2 < \mu_1$ .

The considered two-sample tests rely on the assumption that the two random samples are normally distributed or that the sample size is large. Suppose that the two samples are normally distributed. If variances of the considered populations are known a priori, the test statistic has a standard normal distribution under the null hypothesis, and the corresponding test is referred to as a two-sample  $z$  test. If variances of the two populations are not known, then the null sampling distribution of the test statistic depends on whether the two variances are assumed to be equal. If the two variances are assumed to be equal, the test statistic has an exact Student's  $t$  distribution under the null hypothesis. The corresponding test is referred to as a two-sample  $t$  test. If the two variances are not equal, then the distribution can only be approximated by a Student's  $t$  distribution; the degrees of freedom is approximated using Satterthwaite's method. We refer to this test as Satterthwaite's  $t$  test. For a large sample, the distribution of the test statistic is approximately normal, and the corresponding test is a large-sample  $z$  test.

The `power twomeans` command provides power and sample-size analysis for the above tests.

## Using power twomeans

`power twomeans` computes sample size, power, or experimental-group mean for a two-sample means test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced- or equal-allocation design; see [PSS] [unbalanced designs](#) for a description of how to specify an unbalanced design.

By default, all computations are for a two-sample  $t$  test, which assumes equal and unknown standard deviations. By default, the common standard deviation is set to one but may be changed by specifying the `sd()` option. To specify different standard deviations, use the respective `sd1()` and `sd2()` options. These options must be specified together and may not be used in combination with `sd()`. When `sd1()` and `sd2()` are specified, the computations are based on Satterthwaite's  $t$  test, which assumes unequal and unknown standard deviations. If standard deviations are known, use the `knownsds` option to request that computations be based on a two-sample  $z$  test.

To compute the total sample size, you must specify the control-group mean  $m_1$ , the experimental-group mean  $m_2$ , and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

Instead of the total sample size, you can compute one of the group sizes given the other one. To compute the control-group sample size, you must specify the `compute(n1)` option and the sample size of the experimental group in the `n2()` option. Likewise, to compute the experimental-group sample size, you must specify the `compute(n2)` option and the sample size of the control group in the `n1()` option.

To compute power, you must specify the total sample size in the `n()` option, the control-group mean  $m_1$ , and the experimental-group mean  $m_2$ .

Instead of the experimental-group mean  $m_2$ , you may specify the difference  $m_2 - m_1$  between the experimental-group mean and the control-group mean in the `diff()` option when computing sample size or power.

To compute effect size, the difference between the experimental-group mean and the null mean, and the experimental-group mean, you must specify the total sample size in the `n()` option, the power in the `power()` option, the control-group mean  $m_1$ , and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the experimental-group mean is assumed to be larger than the specified control-group value. You can change the direction to be lower, which means that the experimental-group mean is assumed to be smaller than the specified control-group value, by specifying the `direction(lower)` option.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see [Two samples](#) in [PSS] **unbalanced designs** for more details.

In the following sections, we describe the use of `power twomeans` accompanied by examples for computing sample size, power, and experimental-group mean.

## Computing sample size

To compute sample size, you must specify the control-group mean  $m_1$ , the experimental-group mean  $m_2$ , and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-sample means test

Consider a study investigating the effects of smoking on lung function of males. The response variable is forced expiratory volume (FEV), measured in liters (L), where better lung function implies higher values of FEV. We wish to test the null hypothesis  $H_0: \mu_1 = \mu_2$  versus a two-sided alternative hypothesis  $H_a: \mu_1 \neq \mu_2$ , where  $\mu_1$  and  $\mu_2$  are the mean FEV for nonsmokers and smokers, respectively.

Suppose that the mean FEV from previous studies was reported to be 3 L for nonsmokers and 2.7 L for smokers. We are designing a new study and wish to find out how many subjects we need to enroll so that the power of a 5%-level two-sided test to detect the specified difference between means is at least 80%. We assume equal numbers of subjects in each group and a common standard deviation of 1.

```
. power twomeans 3 2.7
Performing iteration ...
Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = -0.3000
    m1 = 3.0000
    m2 = 2.7000
    sd = 1.0000
Estimated sample sizes:
    N = 352
    N per group = 176
```

We need a total sample of 352 subjects, 176 per group, to detect the specified mean difference between the smoking and nonsmoking groups with 80% power using a two-sided 5%-level test.

The default computation is for the case of equal and unknown standard deviations, as indicated by the output. You can specify the `knownsds` option to request the computation assuming known standard deviations.

◀

## ▶ Example 2: Sample size assuming unequal standard deviations

Instead of assuming equal standard deviations as in [example 1](#), we use the estimates of the standard deviations from previous studies as our hypothetical values. The standard deviation of FEV for the nonsmoking group was reported to be 0.8 L and that for the smoking group was reported to be 0.7 L. We specify standard deviations in the `sd1()` and `sd2()` options.

```
. power twomeans 3 2.7, sd1(0.8) sd2(0.7)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = -0.3000
    m1 = 3.0000
    m2 = 2.7000
    sd1 = 0.8000
    sd2 = 0.7000
Estimated sample sizes:
    N = 200
    N per group = 100
```

The specified standard deviations are smaller than one, so we obtain a smaller required total sample size of 200 compared with [example 1](#).

◀

### ▶ Example 3: Specifying difference between means

Instead of the mean FEV of 2.7 for the smoking group as in [example 2](#), we can specify the difference between the two means of  $2.7 - 3 = -0.3$  in the `diff()` option.

```
. power twomeans 3, sd1(0.8) sd2(0.7) diff(-0.3)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.3000
     m1 =    3.0000
     m2 =    2.7000
    diff =   -0.3000
     sd1 =    0.8000
     sd2 =    0.7000
Estimated sample sizes:
      N =      200
  N per group =    100
```

We obtain the same results as in [example 2](#). The difference between means is now also reported in the output following the individual means.

◀

### ▶ Example 4: Computing one of the group sizes

Suppose we anticipate a sample of 120 nonsmoking subjects. We wish to compute the required number of subjects in the smoking group, keeping all other study parameters as in [example 2](#). We specify the number of subjects in the nonsmoking group in the `n1()` option and specify the `compute(n2)` option.

```
. power twomeans 3 2.7, sd1(0.8) sd2(0.7) n1(120) compute(n2)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.3000
     m1 =    3.0000
     m2 =    2.7000
     sd1 =    0.8000
     sd2 =    0.7000
     N1 =     120
Estimated sample sizes:
      N =     202
     N2 =     82
```

We need a sample of 82 smoking subjects given a sample of 120 nonsmoking subjects.

◀

### ▷ Example 5: Unbalanced design

By default, `power twomeans` computes sample size for a balanced- or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size for an unbalanced design by specifying the `nratio()` option.

Continuing with [example 2](#), we will suppose that we anticipate to recruit twice as many smokers than nonsmokers; that is,  $n_2/n_1 = 2$ . We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```
. power twomeans 3 2.7, sd1(0.8) sd2(0.7) nratio(2)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.3000
     m1 =    3.0000
     m2 =    2.7000
    sd1 =    0.8000
    sd2 =    0.7000
    N2/N1 =    2.0000
Estimated sample sizes:
     N =    237
    N1 =    79
    N2 =   158
```

We need a total sample size of 237 subjects, which is larger than the required total sample size for the corresponding balanced design from [example 2](#).

Also see [Two samples](#) in [PSS] [unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.

◀

## Computing power

To compute power, you must specify the total sample size in the `n()` option, the control-group mean  $m_1$ , and the experimental-group mean  $m_2$ .

### ▷ Example 6: Power of a two-sample means test

Continuing with [example 1](#), we will suppose that we have resources to enroll a total of only 250 subjects, assuming equal-sized groups. To compute the power corresponding to this sample size given the study parameters from [example 1](#), we specify the total sample size in `n()`:

```
. power twomeans 3 2.7, n(250)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N =      250
N per group =    125
    delta =   -0.3000
      m1 =    3.0000
      m2 =    2.7000
      sd =    1.0000
Estimated power:
    power =    0.6564
```

With a total sample of 250 subjects, we obtain a power of only 65.64%.

◀

### ▶ Example 7: Multiple values of study parameters

In this example, we assess the effect of varying the common standard deviation (assuming equal standard deviations in both groups) of FEV on the power of our study.

Continuing with [example 6](#), we compute powers for a range of common standard deviations between 0.5 and 1.5 with the step size of 0.1. We specify the corresponding `numlist` in the `sd()` option.

```
. power twomeans 3 2.7, sd(0.5(0.1)1.5) n(250)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
```

alpha	power	N	N1	N2	delta	m1	m2	sd
.05	.9972	250	125	125	-.3	3	2.7	.5
.05	.976	250	125	125	-.3	3	2.7	.6
.05	.9215	250	125	125	-.3	3	2.7	.7
.05	.8397	250	125	125	-.3	3	2.7	.8
.05	.747	250	125	125	-.3	3	2.7	.9
.05	.6564	250	125	125	-.3	3	2.7	1
.05	.5745	250	125	125	-.3	3	2.7	1.1
.05	.5036	250	125	125	-.3	3	2.7	1.2
.05	.4434	250	125	125	-.3	3	2.7	1.3
.05	.3928	250	125	125	-.3	3	2.7	1.4
.05	.3503	250	125	125	-.3	3	2.7	1.5

The power decreases from 99.7% to 35.0% as the common standard deviation increases from 0.5 to 1.5 L.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

◀

## Computing effect size and experimental-group mean

Effect size  $\delta$  for a two-sample means test is defined as the difference between the experimental-group mean and the control-group mean  $\delta = \mu_2 - \mu_1$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding experimental-group mean that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and control-group mean must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $m_2 > m_1$ , or lower, meaning  $m_2 < m_1$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ► Example 8: Minimum detectable change in the experimental-group mean

Continuing with [example 6](#), we compute the smallest change in the mean of the smoking group that can be detected given a total sample of 250 subjects and 80% power, assuming equal-group allocation. To solve for the mean FEV of the smoking group, after the command name, we specify the nonsmoking-group mean of 3, total sample size `n(250)`, and power `power(0.8)`.

Because our initial study was based on the hypothesis that FEV for the smoking group is lower than that of the nonsmoking group, we specify the `direction(lower)` option to compute the smoking-group mean that is lower than the specified nonsmoking-group mean.

```
. power twomeans 3, n(250) power(0.8) direction(lower)
Performing iteration ...
Estimated experimental-group mean for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1; m2 < m1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      250
N per group =    125
         m1 =    3.0000
         sd =    1.0000
Estimated effect size and experimental-group mean:
      delta =   -0.3558
         m2 =    2.6442
```

We find that the minimum detectable value of the effect size is  $-0.36$ , which corresponds to the mean FEV of 2.64 for the smoking group.

◀

## Testing a hypothesis about two independent means

After data are collected, we can use the `ttest` command to test the equality of two independent means using a  $t$  test; see [\[R\] ttest](#) for details. In this section, we demonstrate the use of `ttesti`, the immediate form of the `test` command, which can be used to test a hypothesis using summary statistics instead of the actual data values.

▷ Example 9: Two-sample  $t$  test

Consider an example from [van Belle et al. \(2004, 129\)](#), where newborn infants were divided into two groups: a treatment group, where infants received daily “walking stimulus” for eight weeks, and a control group, where no stimulus was provided. The goal of this study was to test whether receiving the walking stimulus during stages of infancy induces the walking ability to develop sooner.

The average number of months before the infants started walking was recorded for both groups. The authors provide estimates of the average of 10.125 months for the treatment group with estimated standard deviation of 1.447 months and 12.35 months for the control group with estimated standard deviation of 0.9618 months. The sample sizes for treatment and control groups were 6 and 5, respectively. We supply these estimates to the `ttesti` command and use the `unequal` option to perform a  $t$  test assuming unequal variances.

```
. ttesti 6 10.125 1.447 5 12.35 0.9618, unequal
```

Two-sample t test with unequal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	6	10.125	.5907353	1.447	8.606467	11.64353
y	5	12.35	.43013	.9618	11.15577	13.54423
combined	11	11.13636	.501552	1.66346	10.01884	12.25389
diff		-2.225	.7307394		-3.887894	-.562106

```
diff = mean(x) - mean(y)                                t = -3.0449
Ho: diff = 0                                           Satterthwaite's degrees of freedom = 8.66326
Ha: diff < 0                                           Ha: diff != 0                                Ha: diff > 0
Pr(T < t) = 0.0073                                     Pr(|T| > |t|) = 0.0145                       Pr(T > t) = 0.9927
```

We reject the null hypothesis of  $H_0: \mu_C = \mu_T$  against the two-sided alternative  $H_a: \mu_C \neq \mu_T$  at the 5% significance level; the  $p$ -value = 0.0145.

We use the estimates of this study to perform a sample-size analysis we would have conducted before a new study. In our analysis, we assume equal-group allocation.

```
. power twomeans 10.125 12.35, power(0.8) sd1(1.447) sd2(0.9618)
```

Performing iteration ...

Estimated sample sizes for a two-sample means test

Satterthwaite's t test assuming unequal variances

Ho: m2 = m1 versus Ha: m2 != m1

Study parameters:

```
alpha = 0.0500
power = 0.8000
delta = 2.2250
m1 = 10.1250
m2 = 12.3500
sd1 = 1.4470
sd2 = 0.9618
```

Estimated sample sizes:

```
N = 14
N per group = 7
```

We find that the sample size required to detect a difference of 2.225 ( $12.35 - 10.125 = 2.225$ ) given the control-group standard deviation of 1.447 and the experimental-group standard deviation of 0.9618 using a 5%-level two-sided test is 7 in each group.



## Stored results

power twomeans stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(m1)</code>	control-group mean
<code>r(m2)</code>	experimental-group mean
<code>r(diff)</code>	difference between the experimental- and control-group means
<code>r(sd)</code>	common standard deviation of the control and experimental groups
<code>r(sd1)</code>	standard deviation of the control group
<code>r(sd2)</code>	standard deviation of the experimental group
<code>r(knownsds)</code>	1 if option <code>knownsds</code> is specified; 0 otherwise
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for sample sizes or experimental-group mean
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	twomeans
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Consider two independent samples with  $n_1$  subjects in the control group and  $n_2$  subjects in the experimental group. Let  $x_{11}, \dots, x_{1n_1}$  be a random sample of size  $n_1$  from a normal population with mean  $\mu_1$  and variance  $\sigma_1^2$ . Let  $x_{21}, \dots, x_{2n_2}$  be a random sample of size  $n_2$  from a normal population with mean  $\mu_2$  and variance  $\sigma_2^2$ . Let effect size  $\delta$  be the difference between the experimental-group mean and the control-group mean,  $\delta = \mu_2 - \mu_1$ . The sample means and variances for the two independent samples are

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \quad \text{and} \quad s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i} \quad \text{and} \quad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2$$

where  $\bar{x}_j$  and  $s_j^2$  are the respective sample means and sample variances of the two samples.

A two-sample means test involves testing the null hypothesis  $H_0: \mu_2 = \mu_1$  versus the two-sided alternative hypothesis  $H_a: \mu_2 \neq \mu_1$ , the upper one-sided alternative  $H_a: \mu_2 > \mu_1$ , or the lower one-sided alternative  $H_a: \mu_2 < \mu_1$ .

The two-sample means test can be performed under four different assumptions: 1) population variances are known and not equal; 2) population variances are known and equal; 3) population variances are unknown and not equal; and 4) population variances are unknown and equal.

Let  $\sigma_D$  denote the standard deviation of the difference between the two sample means. The test statistic of the form

$$TS = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sigma_D} \quad (1)$$

is used in each of the four cases described above. Each case, however, determines the functional form of  $\sigma_D$  and the sampling distribution of the test statistic (1) under the null hypothesis.

Let  $R = n_2/n_1$  denote the allocation ratio. Then  $n_2 = R \times n_1$  and power can be viewed as a function of  $n_1$ . Therefore, for sample-size determination, the control-group sample size  $n_1$  is computed first. The experimental-group size  $n_2$  is then computed as  $R \times n_1$ , and the total sample size is computed as  $n = n_1 + n_2$ . By default, sample sizes are rounded to integer values; see *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

The following formulas are based on Armitage, Berry, and Matthews (2002); Chow, Shao, and Wang (2008); and Dixon and Massey (1983).

Methods and formulas are presented under the following headings:

- Known standard deviations*
- Unknown standard deviations*
- Unequal standard deviations*
- Equal standard deviations*

## Known standard deviations

Below we present formulas for the computations that assume unequal standard deviations. When standard deviations are equal, the corresponding formulas are special cases of the formulas below with  $\sigma_1 = \sigma_2 = \sigma$ .

When the standard deviations of the control and the experimental groups are known, the test statistic in (1) is a  $z$  test statistic

$$z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

with  $\sigma_D = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ . The sampling distribution of this test statistic under the null hypothesis is standard normal. The corresponding test is referred to as a  $z$  test.

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha}$  and  $z_\beta$  be the  $(1 - \alpha)$ th and the  $\beta$ th quantiles of a standard normal distribution.

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi\left(\frac{\delta}{\sigma_D} - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\ \Phi\left(-\frac{\delta}{\sigma_D} - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\ \Phi\left(\frac{\delta}{\sigma_D} - z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta}{\sigma_D} - z_{1-\alpha/2}\right) & \text{for a two-sided test} \end{cases} \quad (2)$$

where  $\Phi(\cdot)$  is the cdf of a standard normal distribution.

For a one-sided test, the control-group sample size  $n_1$  is computed as follows:

$$n_1 = \left(\frac{z_{1-\alpha} - z_\beta}{\mu_2 - \mu_1}\right)^2 \left(\sigma_1^2 + \frac{\sigma_2^2}{R}\right) \quad (3)$$

For a one-sided test, if one of the group sizes is known, the other one is computed using the following formula. For example, to compute  $n_1$  given  $n_2$ , we use the following formula:

$$n_1 = \frac{\sigma_1^2}{\left(\frac{\mu_2 - \mu_1}{z_{1-\alpha} - z_\beta}\right)^2 - \frac{\sigma_2^2}{n_2}} \quad (4)$$

For a two-sided test, sample sizes are computed by iteratively solving the two-sided power equation in (2). The default initial values for the iterative procedure are calculated from the respective equations (3) and (4), with  $\alpha$  replaced with  $\alpha/2$ .

The absolute value of the effect size for a one-sided test is obtained by inverting the corresponding one-sided power equation in (2):

$$|\delta| = \sigma_D(z_{1-\alpha} - z_\beta)$$

Note that the magnitude of the effect size is the same regardless of the direction of the test.

The experimental-group mean for a one-sided test is then computed as

$$\mu_2 = \begin{cases} \mu_1 + (z_{1-\alpha} - z_\beta)\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} & \text{when } \mu_2 > \mu_1 \\ \mu_1 - (z_{1-\alpha} - z_\beta)\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} & \text{when } \mu_2 < \mu_1 \end{cases}$$

For a two-sided test, the experimental-group mean is computed by iteratively solving the two-sided power equation in (2) for  $\mu_2$ . The default initial value is obtained from the corresponding one-sided computation with  $\alpha/2$ .

## Unknown standard deviations

When the standard deviations of the control group and the experimental group are unknown, the test statistic in (1) is a  $t$  test statistic

$$t = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{s_D}$$

where  $s_D$  is the estimated standard deviation of the sample mean difference. The sampling distribution of this test statistic under the null hypothesis is (approximately) a Student's  $t$  distribution with  $\nu$  degrees of freedom. Parameters  $\nu$  and  $s_D$  are defined below, separately for the case of equal and unequal standard deviations.

Let  $t_{\nu,\alpha}$  denote the  $\alpha$ th quantile of a Student's  $t$  distribution with  $\nu$  degrees of freedom. Under the alternative hypothesis, the test statistic follows a noncentral Student's  $t$  distribution with  $\nu$  degrees of freedom and noncentrality parameter  $\lambda$ .

The power is computed from the following equations:

$$\pi = \begin{cases} 1 - T_{\nu,\lambda}(t_{\nu,1-\alpha}) & \text{for an upper one-sided test} \\ T_{\nu,\lambda}(-t_{\nu,1-\alpha}) & \text{for a lower one-sided test} \\ 1 - T_{\nu,\lambda}(t_{\nu,1-\alpha/2}) + T_{\nu,\lambda}(-t_{\nu,1-\alpha/2}) & \text{for a two-sided test} \end{cases} \quad (5)$$

In the equations above,  $\lambda = |\mu_2 - \mu_1|/s_D$ .

Sample sizes and the experimental-group mean are obtained by iteratively solving the nonlinear equation (5) for  $n_1$ ,  $n_2$ , and  $\mu_2$ , respectively. For sample-size and effect-size computations, the default initial values for the iterative procedure are calculated using the corresponding formulas assuming known standard deviations from the [previous subsection](#).

## Unequal standard deviations

In the case of unequal standard deviations,

$$s_D = \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

and the degrees of freedom  $\nu$  of the test statistic is obtained by Satterthwaite's formula:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

The sampling distribution of the test statistic under the null hypothesis is an approximate Student's  $t$  distribution. We refer to the corresponding test as Satterthwaite's  $t$  test.

## Equal standard deviations

In the case of equal standard deviations,

$$s_D = s_p \sqrt{1/n_1 + 1/n_2}$$

where  $s_p = \left\{ \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2 \right\} / (n_1 + n_2 - 2)$  is the pooled-sample standard deviation.

The degrees of freedom  $\nu$  is

$$\nu = n_1 + n_2 - 2$$

The sampling distribution of the test statistic under the null hypothesis is exactly a Student's  $t$  distribution. We refer to the corresponding test as a two-sample  $t$  test.

## References

- Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.
- Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.
- Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw-Hill.
- van Belle, G., L. D. Fisher, P. J. Heagerty, and T. S. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences*. 2nd ed. New York: Wiley.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power oneway** — Power analysis for one-way analysis of variance
- [PSS] **power twoway** — Power analysis for two-way analysis of variance
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **ttest** —  $t$  tests (mean-comparison tests)

# Title

**power pairedmeans** — Power analysis for a two-sample paired-means test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power pairedmeans` computes sample size, power, or target mean difference for a two-sample paired-means test. By default, it computes sample size for given power and the values of the null and alternative mean differences. Alternatively, it can compute power for given sample size and the values of the null and alternative mean differences or the target mean difference for given sample size, power, and the null mean difference. Also see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \mu_2 - \mu_1 = d = 0$  versus  $H_a: d \neq 0$  given alternative pretreatment mean  $m_{a1} = 73$  and alternative posttreatment mean  $m_{a2} = 57$  with standard deviation of the differences  $\sigma_d = 36$  using default power of 0.8 and significance level  $\alpha = 0.05$

```
power pairedmeans 73 57, sddiff(36)
```

Same as above, specified using the difference between means of  $-16$

```
power pairedmeans, altdiff(-16) sddiff(36)
```

As above, but instead of standard deviation of the differences, specify correlation between paired observations of 0.5 with pretreatment standard deviation of 29 and posttreatment standard deviation of 40

```
power pairedmeans 73 57, corr(.5) sd1(29) sd2(40)
```

For differences in means of  $-20$ ,  $-18$ ,  $-16$ ,  $-14$ ,  $-12$ , and  $-10$

```
power pairedmeans, altdiff(-20(2)-10) sddiff(36)
```

Power for a sample size of 23

```
power pairedmeans 73 57, sddiff(36) n(23)
```

Effect size and target mean difference for sample sizes 20, 30, and 40 with power of 0.85

```
power pairedmeans 73, sddiff(36) power(.85) n(20(10)40)
```

As above, but display results as a graph of target mean difference versus sample size

```
power pairedmeans 73, sddiff(36) power(.85) n(20(10)40) graph
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power pairedmeans  $m_{a1}$   $m_{a2}$ , corrspec [power(numlist) options]
```

Compute power

```
power pairedmeans  $m_{a1}$   $m_{a2}$ , corrspec n(numlist) [options]
```

Compute effect size and target mean difference

```
power pairedmeans [ $m_{a1}$ ], corrspec n(numlist) power(numlist) [options]
```

where *corrspec* is one of

```
sddiff()
```

```
corr() [sd()]
```

```
corr() [sd1() sd2()]
```

$m_{a1}$  is the alternative pretreatment mean or the pretreatment mean under the alternative hypothesis, and  $m_{a2}$  is the alternative posttreatment mean or the value of the posttreatment mean under the alternative hypothesis.  $m_{a1}$  and  $m_{a2}$  may each be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 **numlist**).

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample size
* <u>nulldiff</u> ( <i>numlist</i> )	null difference, the difference between the posttreatment mean and the pretreatment mean under the null hypothesis; default is <code>nulldiff(0)</code>
* <u>altdiff</u> ( <i>numlist</i> )	alternative difference $d_a = m_{a2} - m_{a1}$ , the difference between the posttreatment mean and the pretreatment mean under the alternative hypothesis
* <u>sddiff</u> ( <i>numlist</i> )	standard deviation $\sigma_d$ of the differences; may not be combined with <code>corr()</code>
* <u>corr</u> ( <i>numlist</i> )	correlation between paired observations; required unless <code>sddiff()</code> is specified
* <u>sd</u> ( <i>numlist</i> )	common standard deviation; default is <code>sd(1)</code> and requires <code>corr()</code>
* <u>sd1</u> ( <i>numlist</i> )	standard deviation of the pretreatment group; requires <code>corr()</code>
* <u>sd2</u> ( <i>numlist</i> )	standard deviation of the posttreatment group; requires <code>corr()</code>
<u>knownsd</u>	request computation assuming a known standard deviation $\sigma_d$ ; default is to assume an unknown standard deviation
* <u>fpc</u> ( <i>numlist</i> )	finite population correction (FPC) as a sampling rate or population size
<u>direction</u> ( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<u>[no]</u> <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [ , <code>replace</code> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>



iteration	
<code>init(#)</code>	initial value for sample size or mean difference; default is to use normal approximation
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 [numlist](#).

`notitle` does not appear in the dialog box.

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	number of subjects	$N$
<code>delta</code>	effect size	$\delta$
<code>d0</code>	null mean difference	$d_0$
<code>da</code>	alternative mean difference	$d_a$
<code>ma1</code>	alternative pretreatment mean	$\mu_{a1}$
<code>ma2</code>	alternative posttreatment mean	$\mu_{a2}$
<code>sd_d</code>	standard deviation of the differences	$\sigma_d$
<code>sd</code>	common standard deviation	$\sigma$
<code>sd1</code>	standard deviation of the pretreatment group	$\sigma_1$
<code>sd2</code>	standard deviation of the posttreatment group	$\sigma_2$
<code>corr</code>	correlation between paired observations	$\rho$
<code>fpc</code>	FPC as a population size	$N_{\text{pop}}$
	FPC as a sampling rate	$\gamma$
<code>target</code>	target parameter; synonym for <code>da</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `ma1`, `ma2`, `sd`, `sd1`, `sd2`, `corr`, and `fpc` are shown in the default table if specified.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**. The `nfractional` option is allowed only for sample-size determination.

`nulldiff(numlist)` specifies the difference between the posttreatment mean and the pretreatment mean under the null hypothesis. The default is `nulldiff(0)`, which means that the pretreatment mean equals the posttreatment mean under the null hypothesis.

`altdiff(numlist)` specifies the alternative difference  $d_a = m_{a2} - m_{a1}$ , the difference between the posttreatment mean and the pretreatment mean under the alternative hypothesis. This option is the alternative to specifying the alternative means  $m_{a1}$  and  $m_{a2}$ . If  $m_{a1}$  is specified in combination with `altdiff(#)`, then  $m_{a2} = \# + m_{a1}$ .

`sddiff(numlist)` specifies the standard deviation  $\sigma_d$  of the differences. Either `sddiff()` or `corr()` must be specified.

`corr(numlist)` specifies the correlation between paired, pretreatment and posttreatment, observations. This option along with `sd1()` and `sd2()` or `sd()` is used to compute the standard deviation of the differences unless that standard deviation is supplied directly in the `sddiff()` option. Either `corr()` or `sddiff()` must be specified.

`sd(numlist)` specifies the common standard deviation of the pretreatment and posttreatment groups. Specifying `sd(#)` implies that both `sd1()` and `sd2()` are equal to `#`. Options `corr()` and `sd()` are used to compute the standard deviation of the differences unless that standard deviation is supplied directly with the `sddiff()` option. The default is `sd(1)`.

`sd1(numlist)` specifies the standard deviation of the pretreatment group. Options `corr()`, `sd1()`, and `sd2()` are used to compute the standard deviation of the differences unless that standard deviation is supplied directly with the `sddiff()` option.

`sd2(numlist)` specifies the standard deviation of the posttreatment group. Options `corr()`, `sd1()`, and `sd2()` are used to compute the standard deviation of the differences unless that standard deviation is supplied directly with the `sddiff()` option.

`knownsd` requests that the standard deviation of the differences  $\sigma_d$  be treated as known in the computations. By default, the standard deviation is treated as unknown, and the computations are based on a paired  $t$  test, which uses a Student's  $t$  distribution as a sampling distribution of the test statistic. If `knownsd` is specified, the computation is based on a paired  $z$  test, which uses a normal distribution as the sampling distribution of the test statistic.

`fpc(numlist)` requests that a finite population correction be used in the computation. If `fpc()` has values between 0 and 1, it is interpreted as a sampling rate,  $n/N$ , where  $N$  is the total number of units in the population. If `fpc()` has values greater than  $n$ , it is interpreted as a population size. It is an error for `fpc()` to have values between 1 and  $n$  or to have a mixture of sampling rates and population sizes.

`direction()`, `onesided`, `parallel`; see [PSS] **power**.

### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

## Graph

`graph`, `graph()`; see [PSS] [power](#), [graph](#). Also see the *column* table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value of the sample size for the sample-size determination or the initial value of the mean difference for the effect-size determination. The default is to use a closed-form normal approximation to compute an initial value of the sample size or mean difference.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power pairedmeans` but is not shown in the dialog box: `notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

[Introduction](#)

[Using power pairedmeans](#)

[Computing sample size](#)

[Computing power](#)

[Computing effect size and target mean difference](#)

[Testing a hypothesis about two correlated means](#)

[Video examples](#)

This entry describes the `power pairedmeans` command and the methodology for power and sample-size analysis for a two-sample paired-means test. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

The analysis of paired means is commonly used in settings such as repeated-measures designs with before and after measurements on the same individual or cross-sectional studies of paired measurements from twins. For example, a company might initiate a voluntary exercise program and would like to test that the average weight loss of participants from beginning to six months is greater than zero. Or a school district might design an intensive remedial program for students with low math scores and would like to know if the students' math scores improve from the pretest to the posttest. For paired data, the inference is made on the mean difference accounting for the dependence between the two groups.

This entry describes power and sample-size analysis for the inference about the population mean difference performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: d = d_0$  versus the two-sided alternative hypothesis  $H_a: d \neq d_0$ , the upper one-sided alternative  $H_a: d > d_0$ , or the lower one-sided alternative  $H_a: d < d_0$ . The parameter  $d$  is the mean difference between the posttreatment mean  $\mu_2$  and pretreatment mean  $\mu_1$ .

A two-sample paired-means test assumes that the two correlated samples are drawn from two normal populations or that the sample size is large. When the population variances are known, the sampling distribution of the test statistic under the null hypothesis is standard normal, and the corresponding test is known as a paired  $z$  test. If the population variances are unknown, the sampling distribution of the test statistic under the null hypothesis is Student's  $t$ , and the corresponding test is known as a paired  $t$  test.

The random sample is typically drawn from an infinite population. When the sample is drawn from a population of a fixed size, sampling variability must be adjusted for a finite population size.

The `power pairedmeans` command provides power and sample-size analysis for the comparison of two correlated means using a paired  $t$  test or a paired  $z$  test.

## Using `power pairedmeans`

`power pairedmeans` computes sample size, power, or target mean difference for a two-sample paired-means test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

By default, all computations are based on a paired  $t$  test, which assumes an unknown standard deviation of the differences. For a known standard deviation, you can specify the `knownsd` option to request a paired  $z$  test.

For all computations, you must specify either the standard deviation of the differences in the `sddiff()` option or the correlation between the paired observations in the `corr()` option. If you specify the `corr()` option, then individual standard deviations of the pretreatment and posttreatment groups may also be specified in the respective `sd1()` and `sd2()` options. By default, their values are set to 1. When the two standard deviations are equal, you may specify the common standard deviation in the `sd()` option instead of specifying them individually.

To compute sample size, you must specify the pretreatment and posttreatment means under the alternative hypothesis,  $m_{a1}$  and  $m_{a2}$ , respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the sample size in the `n()` option and the pretreatment and posttreatment means under the alternative hypothesis,  $m_{a1}$  and  $m_{a2}$ , respectively.

Instead of the alternative means  $m_{a1}$  and  $m_{a2}$ , you can specify the difference  $m_{a2} - m_{a1}$  between the alternative posttreatment mean and the alternative pretreatment mean in the `altdiff()` option when computing sample size or power.

By default, the difference between the posttreatment mean and the pretreatment mean under the null hypothesis is set to zero. You may specify other values in the `nulldiff()` option.

To compute effect size, the standardized difference between the alternative and null mean differences, and target mean difference, you must specify the sample size in the `n()` option, the power in the `power()` option, and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the target mean difference is assumed to be larger than the specified null value. This is also equivalent to the assumption of a positive effect size. You can change the direction to be lower, which means that the target mean difference is assumed to be smaller than the specified null value, by specifying the `direction(lower)` option. This is equivalent to assuming a negative effect size.

By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS] **unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

Some of `power pairedmeans`'s computations require iteration. For example, when the standard deviation of the differences is unknown, computations use a noncentral Student's  $t$  distribution. Its degrees of freedom depends on the sample size, and the noncentrality parameter depends on the sample size and effect size. Therefore, the sample-size and effect-size determinations require iteration. The default initial values of the estimated parameters are obtained by using a closed-form normal

approximation. They may be changed by specifying the `init()` option. See [PSS] [power](#) for the descriptions of other options that control the iteration procedure.

All computations assume an infinite population. For a finite population, use the `fpc()` option to specify a sampling rate or a population size. When this option is specified, a finite population correction is applied to the standard deviation of the differences. The correction factor depends on the sample size; therefore, computing sample size in this case requires iteration. The initial value for sample-size determination in this case is based on the corresponding normal approximation with a finite population size.

In the following sections, we describe the use of `power pairedmeans` accompanied by examples for computing sample size, power, and target mean difference.

## Computing sample size

To compute sample size, you must specify the pretreatment and posttreatment means under the alternative hypothesis,  $m_{a1}$  and  $m_{a2}$ , respectively, or the difference between them in `altdiff()` and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-sample paired-means test

Consider a study of low birthweight (LBW) infants as in [Howell \(2002, 186\)](#). The variable of interest is the Bayley mental development index (MDI) of infants when they are 6-, 12-, and 24-months old. Previous research suggested that the MDI scores for LBW children might decline significantly between 6 and 24 months of age. Suppose we would like to conduct a similar study where the null hypothesis of interest is no difference between 6-month and 24-month MDI scores,  $H_0: d = 0$ , and the two-sided alternative is  $H_a: d \neq 0$ , implying the existence of a difference.

In this example, we use the estimates from [Howell \(2002, 193\)](#) as our study parameters. The mean MDI score of a 6-month group was estimated to be 111. We want to obtain the minimum sample size that is required to detect the mean MDI score of 106.71 in a 24-month group with a power of 80% using a 5%-level two-sided test. The standard deviation of the differences was previously estimated to be 16.04. To compute the sample size, we specify the alternative means after the command name and standard deviation of the differences in `sddiff()`.

```

. power pairedmeans 111 106.71, sddiff(16.04)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha =    0.0500          ma1 = 111.0000
      power =    0.8000          ma2 = 106.7100
      delta =   -0.2675
      d0 =      0.0000
      da =     -4.2900
      sd_d =    16.0400
Estimated sample size:
      N =          112

```

As we mentioned in the [previous section](#), sample-size determination requires iteration in the case of an unknown standard deviation. By default, `power pairedmeans` suppresses the iteration log, which may be displayed by specifying the `log` option.

A sample of 112 subjects is required for the test to detect the resulting difference of  $-4.29$  with a power of 80%.

Study parameters are divided into two columns. The parameters that are always displayed are listed in the first column, and the parameters that are displayed only if they are specified are listed in the second column.

In this example, we specified optional command arguments containing the alternative pretreatment mean `ma1` and the alternative posttreatment mean `ma2`. Because these arguments are optional, they are listed in the second column.

◀

## ▷ Example 2: Specifying mean differences

Instead of the individual alternative means, we can specify their difference,  $106.71 - 111 = -4.29$ , in the `altdiff()` option.

```

. power pairedmeans, altdiff(-4.29) sddiff(16.04)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.2675
      d0 =      0.0000
      da =     -4.2900
      sd_d =    16.0400
Estimated sample size:
      N =          112

```

We obtain the same results as in [example 1](#).

◀

### ▷ Example 3: Specifying individual standard deviations

Howell (2002) also reported the group-specific standard deviations: 13.85 in the 6-month group and 12.95 in the 24-month group. Using the values of individual standard deviations and the standard deviation of the differences from the [previous example](#), we obtain the correlation between the 6-month group and the 24-month group to be  $(13.85^2 + 12.95^2 - 16.04^2)/(2 \times 13.85 \times 12.95) = 0.285$ . To compute the sample size, we specify the group-specific standard deviations in `sd1()` and `sd2()` and the correlation in `corr()`.

```
. power pairedmeans 111 106.71, corr(0.285) sd1(13.85) sd2(12.95)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha = 0.0500          ma1 = 111.0000
      power = 0.8000          ma2 = 106.7100
      delta = -0.2675         sd1 = 13.8500
      d0 = 0.0000            sd2 = 12.9500
      da = -4.2900           corr = 0.2850
      sd_d = 16.0403
Estimated sample size:
      N = 112
```

We obtain the same sample size as in [example 1](#).

The correlation and standard deviations are reported in the second column.

◀

### ▷ Example 4: Specifying common standard deviation

If standard deviations in both groups are equal, we may specify the common standard deviation in option `sd()`. As a demonstration, we use the average of the individual standard deviations  $(13.85 + 12.95)/2 = 13.4$  as our common standard deviation.

```
. power pairedmeans 111 106.71, corr(0.285) sd(13.4)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test assuming sd1 = sd2 = sd
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha = 0.0500          ma1 = 111.0000
      power = 0.8000          ma2 = 106.7100
      delta = -0.2677         sd = 13.4000
      d0 = 0.0000            corr = 0.2850
      da = -4.2900
      sd_d = 16.0241
Estimated sample size:
      N = 112
```

The resulting standard deviation of the differences of 16.0241 is close to our earlier estimate of 16.04, so the computed sample size is the same as the sample size in [example 1](#).

◀

### ▷ Example 5: Nonzero null

In all the previous examples, we assumed that the difference between the 6-month and 24-month means is zero under the null hypothesis. For a nonzero null hypothesis, you can specify the corresponding null value in the `nulldiff()` option.

Continuing with [example 2](#), we will suppose that we are testing the nonzero null hypothesis of  $H_0: d = d_0 = -1$ . We compute the sample size as follows:

```
. power pairedmeans, nulldiff(-1) altdiff(-4.29) sddiff(16.04)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.2051
    d0 =     -1.0000
    da =     -4.2900
    sd_d =    16.0400
Estimated sample size:
    N =        189
```

Compared with [example 2](#), the absolute value of the effect size `delta` decreases to 0.2051, and thus a larger sample of 189 subjects is required to detect this smaller effect.

◀

## Computing power

To compute power, you must specify the sample size in the `n()` option and the pretreatment and posttreatment means under the alternative hypothesis,  $m_{a1}$  and  $m_{a2}$ , respectively, or the difference between them in the `altdiff()` option.

### ▷ Example 6: Power of a two-sample paired-means test

Continuing with [example 1](#), we will suppose that because of limited resources, we anticipate to obtain a sample of only 100 subjects. To compute power, we specify the sample size in the `n()` option:

```
. power pairedmeans 111 106.71, n(100) sddiff(16.04)
Estimated power for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
    alpha =    0.0500          ma1 =   111.0000
    N =        100           ma2 =   106.7100
    delta =   -0.2675
    d0 =      0.0000
    da =     -4.2900
    sd_d =    16.0400
Estimated power:
    power =    0.7545
```

Compared with [example 1](#), the power decreases to 75.45%.

◀



### ► Example 7: Known standard deviation

In the case of a known standard deviation  $\sigma_d$ , you can specify the `knownsd` option to request a paired  $z$  test. Using the same study parameters as in [example 6](#), we can compute the power as follows:

```
. power pairedmeans 111 106.71, n(100) sddiff(16.04) knownsd
Estimated power for a two-sample paired-means test
Paired z test
Ho: d = d0 versus Ha: d != d0
Study parameters:
      alpha =    0.0500          ma1 = 111.0000
      N      =    100           ma2 = 106.7100
      delta  =   -0.2675
      d0     =    0.0000
      da     =   -4.2900
      sd_d   =   16.0400
Estimated power:
      power  =    0.7626
```

The power of 76.26% of a paired  $z$  test is close to the power of 75.45% of a paired  $t$  test obtained in [example 6](#).

◀

### ► Example 8: Multiple values of study parameters

Continuing with [example 3](#), we will suppose that we would like to assess the effect of varying correlation on the power of our study. The standard deviation of the MDI scores for infants aged 6 months is 13.85 and that for infants aged 24 months is 12.95, which are obtained from [Howell \(2002, 193\)](#). We believe the data on pairs to be positively correlated because we expect a 6-month-old infant with a high score to have a high score at 24 months of age as well. We specify a range of correlations between 0.1 and 0.9 with the step size of 0.1 in the `corr()` option:

```
. power pairedmeans 111 106.71, n(100) sd1(13.85) sd2(12.95) corr(0.1(0.1)0.9)
> table(alpha N power corr sd_d delta)
Estimated power for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
```

alpha	N	power	corr	sd_d	delta
.05	100	.656	.1	17.99	-.2385
.05	100	.7069	.2	16.96	-.2529
.05	100	.7632	.3	15.87	-.2703
.05	100	.8239	.4	14.7	-.2919
.05	100	.8859	.5	13.42	-.3196
.05	100	.9425	.6	12.01	-.3571
.05	100	.983	.7	10.41	-.412
.05	100	.9988	.8	8.518	-.5037
.05	100	1	.9	6.057	-.7083

As the correlation increases, the power also increases. This is because the standard deviation of the differences is negatively related to correlation when the correlation is positive. As the correlation increases, the standard deviation of the differences decreases, thus resulting in higher power. Likewise, the opposite is true when the correlation is negative.

For multiple values of parameters, the results are automatically displayed in a table. In the above, we use the `table()` option to build a custom table. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## Computing effect size and target mean difference

Effect size  $\delta$  for a two-sample paired-means test is defined as a standardized difference between the alternative mean difference  $d_a$  and the null mean difference  $d_0$ ,  $\delta = (d_a - d_0)/\sigma_d$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding mean difference that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and the alternative pretreatment mean must be specified. By default, the null mean difference is set to 0. In addition, you must also decide on the direction of the effect: upper, meaning  $d_a > d_0$ , or lower, meaning  $d_a < d_0$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ▶ Example 9: Minimum detectable value of the effect size

Continuing with [example 6](#), we may be interested to find the minimum effect size with a power of 80% given a sample of 100 subjects. To compute the smallest effect size and the corresponding target mean difference, we specify the sample size `n(100)`, power `power(0.8)`, and the standard deviation of the differences `sddiff(16.04)`:

```
. power pairedmeans 111, n(100) power(0.8) sddiff(16.04)
Performing iteration ...
Estimated target parameters for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0; da > d0
Study parameters:
      alpha =    0.0500          ma1 = 111.0000
      power =    0.8000
      N      =    100
      d0     =    0.0000
      sd_d   =   16.0400
Estimated effect size and target parameters:
      delta =    0.2829
      da    =    4.5379
      ma2   =   115.5379
```

The smallest detectable value of the effect size is 0.28, which corresponds to the alternative mean difference of 4.54. Compared with [example 1](#), for the same power of 80%, the target mean difference increased to 4.54 when the sample size was reduced to 100 subjects.

◀

## Testing a hypothesis about two correlated means

In this section, we demonstrate the use of the `tttest` command for testing hypotheses about paired means. Suppose we wish to test the hypothesis that the means of the paired samples are the same. We can use the `tttest` command to do this. We demonstrate the use of this command using the fictional `bpwide` dataset; see [R] [tttest](#) for details.

## ▷ Example 10: Testing means from paired data

Suppose that we have a sample of 120 patients. We are interested in investigating whether a certain drug induces a change in the systolic blood pressure. We record blood pressures for each patient before and after the drug is administered. In this case, each patient serves as his or her own control. We wish to test whether the mean difference between the posttreatment and pretreatment systolic blood pressures are significantly different from zero.

```
. use http://www.stata-press.com/data/r14/bpwide
(fictional blood-pressure data)
```

```
. ttest bp_before == bp_after
```

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
bp_bef~e	120	156.45	1.039746	11.38985	154.3912	158.5088
bp_after	120	151.3583	1.294234	14.17762	148.7956	153.921
diff	120	5.091667	1.525736	16.7136	2.070557	8.112776

```
mean(diff) = mean(bp_before - bp_after)          t = 3.3372
Ho: mean(diff) = 0                               degrees of freedom = 119
Ha: mean(diff) < 0                               Ha: mean(diff) != 0          Ha: mean(diff) > 0
Pr(T < t) = 0.9994                               Pr(|T| > |t|) = 0.0011      Pr(T > t) = 0.0006
```

We find statistical evidence to reject the null hypothesis of  $H_0: d = 0$  versus the two-sided alternative  $H_a: d \neq 0$  at the 5% significance level; the  $p$ -value = 0.0011.

We use the estimates of this study to perform a sample-size analysis we would have conducted before the study.

```
. power pairedmeans, altdiff(5.09) sddiff(16.71)
```

Performing iteration ...

Estimated sample size for a two-sample paired-means test

Paired t test

Ho:  $d = d_0$  versus Ha:  $d \neq d_0$

Study parameters:

```
alpha = 0.0500
power = 0.8000
delta = 0.3046
d0 = 0.0000
da = 5.0900
sd_d = 16.7100
```

Estimated sample size:

```
N = 87
```

We find that the sample size required to detect a mean difference of 5.09 for given standard deviation of the differences of 16.71 with 80% power using a 5%-level two-sided test is 87.

## Video examples

[Sample-size calculation for comparing sample means from two paired samples](#)

[Power calculation for comparing sample means from two paired samples](#)

[Minimum detectable effect size for comparing sample means from two paired samples](#)

## Stored results

`power pairedmeans` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(d0)</code>	difference between the posttreatment and pretreatment means under the null hypothesis
<code>r(da)</code>	difference between the posttreatment and pretreatment means under the alternative hypothesis
<code>r(ma1)</code>	pretreatment mean under the alternative hypothesis
<code>r(ma2)</code>	posttreatment mean under the alternative hypothesis
<code>r(sd_d)</code>	standard deviation of the differences
<code>r(corr)</code>	correlation between paired observations
<code>r(sd1)</code>	standard deviation of the pretreatment group
<code>r(sd2)</code>	standard deviation of the posttreatment group
<code>r(sd)</code>	common standard deviation
<code>r(knownsd)</code>	1 if option <code>knownsd</code> is specified; 0 otherwise
<code>r(fpc)</code>	finite population correction
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or target mean difference
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	<code>pairedmeans</code>
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Consider a sequence of  $n$  paired observations denoted by  $X_{ij}$  for  $i = 1, \dots, n$  and groups  $j = 1, 2$ . Individual observation corresponds to the pair  $(X_{i1}, X_{i2})$ , and inference is made on the differences within the pairs. Let  $d = \mu_2 - \mu_1$  denote the mean difference, where  $\mu_j$  is the population mean of group  $j$ , and  $D_i = X_{i2} - X_{i1}$  denote the difference between individual observations. Let  $d_0$  and  $d_a$  denote the null and alternative values of the mean difference  $d$ . Let  $\bar{d} = \sum_{i=1}^n D_i/n$  denote the sample mean difference.

Unlike a two-sample means test where we consider two independent samples, a paired-means test allows the two groups to be dependent. As a result, the standard deviation of the differences is given by  $\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$ , where  $\sigma_1$  and  $\sigma_2$  are the pretreatment and posttreatment group standard deviations, respectively, and  $\rho$  is the correlation between the paired measurements.

Power, sample-size, and effect-size determination for a paired-means test is analogous to a one-sample mean test where the sample of differences  $D_i$ 's is treated as a single sample. See *Methods and formulas* in [PSS] **power onemean**.

Also see Armitage, Berry, and Matthews (2002); Dixon and Massey (1983); and Chow, Shao, and Wang (2008) for more details.

## References

- Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.
- Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.
- Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw-Hill.
- Howell, D. C. 2002. *Statistical Methods for Psychology*. 5th ed. Belmont, CA: Wadsworth.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power repeated** — Power analysis for repeated-measures analysis of variance
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **ttest** —  $t$  tests (mean-comparison tests)

# Title

**power oneproportion** — Power analysis for a one-sample proportion test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power oneproportion` computes sample size, power, or target proportion for a one-sample proportion test. By default, it computes sample size for given power and the values of the proportion parameters under the null and alternative hypotheses. Alternatively, it can compute power for given sample size and values of the null and alternative proportions or the target proportion for given sample size, power, and the null proportion. Also see [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \pi = 0.2$  versus  $H_a: \pi \neq 0.2$  with null proportion  $p_0 = 0.2$ , alternative proportion  $p_a = 0.1$ , default power of 0.8, and significance level  $\alpha = 0.05$

```
power oneproportion .2 .1
```

As above, but for  $p_a$  equal to 0.05, 0.075, 0.1, 0.125, and 0.15

```
power oneproportion .2 (.05(.025).15)
```

As above, but display results as a graph of sample size versus alternative proportion

```
power oneproportion .2 (.05(.025).15), graph
```

Sample size for one-sided test with power of 0.9

```
power oneproportion .2 .1, power(.9) onesided
```

Sample size for a Wald test instead of the default score test

```
power oneproportion .2 .1, test(wald)
```

Power for a sample size of 120

```
power oneproportion .2 .1, n(120)
```

As above, but for sample sizes of 110, 120, 130, and 140

```
power oneproportion .2 .1, n(110(10)140)
```

Effect size and target proportion for  $p_0 = 0.3$ , sample size of 75, and power of 0.8

```
power oneproportion .3, n(75) power(.8)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

```
power oneproportion  $p_0$   $p_a$  [ , power(numlist) options ]
```

*Compute power*

```
power oneproportion  $p_0$   $p_a$  , n(numlist) [ options ]
```

*Compute effect size and target proportion*

```
power oneproportion  $p_0$  , n(numlist) power(numlist) [ options ]
```

where  $p_0$  is the null (hypothesized) proportion or the value of the proportion under the null hypothesis, and  $p_a$  is the alternative (target) proportion or the value of the proportion under the alternative hypothesis.  $p_0$  and  $p_a$  may each be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 **numlist**).

<i>options</i>	Description
<code>test(test)</code>	specify the type of test; default is <code>test(score)</code>
Main	
* <code>alpha(numlist)</code>	significance level; default is <code>alpha(0.05)</code>
* <code>power(numlist)</code>	power; default is <code>power(0.8)</code>
* <code>beta(numlist)</code>	probability of type II error; default is <code>beta(0.2)</code>
* <code>n(numlist)</code>	sample size; required to compute power or effect size
<code>nfractional</code>	allow fractional sample size
* <code>diff(numlist)</code>	difference between the alternative proportion and the null proportion, $p_a - p_0$ ; specify instead of the alternative proportion $p_a$
<code>critvalues</code>	show critical values for the binomial test
<code>continuity</code>	apply continuity correction to the normal approximation of the discrete distribution
<code>direction(upper lower)</code>	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<code>onesided</code>	one-sided test; default is two sided
<code>parallel</code>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<code>[no]table[ (tablespec) ]</code>	suppress table or display results as a table; see [PSS] <b>power, table</b>
<code>saving(filename [, replace])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph[ (graphopts) ]</code>	graph results; see [PSS] <b>power, graph</b>
Iteration	
<code>init(#)</code>	initial value for the sample size or proportion
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] **11.1.8 numlist**.

`notitle` does not appear in the dialog box.



<i>test</i>	Description
<code>score</code>	score test; the default
<code>wald</code>	Wald test
<code>binomial</code>	binomial test

`test()` does not appear in the dialog box. The dialog box selected is determined by the `test()` specification.

where *tablespec* is

```
column[ :label ] [ column[ :label ] [ ... ] ] [ , tableopts ]
```

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>alpha_a</code>	observed significance level	$\alpha_a$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	number of subjects	$N$
<code>delta</code>	effect size	$\delta$
<code>p0</code>	null proportion	$p_0$
<code>pa</code>	alternative proportion	$p_a$
<code>diff</code>	difference between the alternative and null proportions	$p_a - p_0$
<code>C_l</code>	lower critical value	$C_l$
<code>C_u</code>	upper critical value	$C_u$
<code>target</code>	target parameter; synonym for <code>pa</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Column `diff` is shown in the default table if specified.

Columns `alpha_a`, `C_l`, and `C_u` are available when the `test(binomial)` option is specified.

Columns `C_l` and `C_u` are shown in the default table, if the `critvalues` option is specified.

## Options

`test(test)` specifies the type of the test for power and sample-size computations. *test* is one of `score`, `wald`, or `binomial`.

`score` requests computations for the score test. This is the default test.

`wald` requests computations for the Wald test. This corresponds to computations using the value of the alternative proportion instead of the default null proportion in the formula for the standard error of the estimator of the proportion.

`binomial` requests computations for the binomial test. The computation using the binomial distribution is not available for sample-size and effect-size determinations; see [example 7](#) for details. Iteration options are not allowed with this test.

## Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**. The `nfractional` option is allowed only for sample-size determination.

`diff(numlist)` specifies the difference between the alternative proportion and the null proportion,  $p_a - p_0$ . You can specify either the alternative proportion  $p_a$  as a command argument or the difference between the two proportions in `diff()`. If you specify `diff(#)`, the alternative proportion is computed as  $p_a = p_0 + \#$ . This option is not allowed with the effect-size determination.

`critvalues` requests that the critical values be reported when the computation is based on the binomial distribution.

`continuity` requests that continuity correction be applied to the normal approximation of the discrete distribution. `continuity` cannot be specified with `test(binomial)`.

`direction()`, `onesided`, `parallel`; see [PSS] **power**.

## Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

## Graph

`graph`, `graph()`; see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value of the sample size for the sample-size determination or the initial value of the proportion for the effect-size determination.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power oneproportion` but is not shown in the dialog box: `notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

- Introduction*
- Using power oneproportion*
- Computing sample size*
- Computing power*
- Computing effect size and target proportion*
- Performing hypothesis tests on proportion*
- Video examples*

This entry describes the `power oneproportion` command and the methodology for power and sample-size analysis for a one-sample proportion test. See [PSS] **intro** for a general introduction to power and sample-size analysis and [PSS] **power** for a general introduction to the `power` command using hypothesis tests.

## Introduction

There are many examples of studies where a researcher would like to compare an observed proportion with a hypothesized proportion. A political campaign might like to know if the proportion of a country's population that supports a new legislative initiative is greater than 50%. A veterinary drug manufacturer might test a new topical treatment to kill fleas on dogs. It would like to know the sample size necessary to demonstrate that the treatment is effective in ridding at least 80% of the test dogs of fleas. The Nevada Gaming Control Board might test a Las Vegas casino's slot machines to verify that it meets the statutory minimum payout percentage of 75%. The board would like to know the number of "pulls" necessary to reject the one-sided null hypothesis that the payout percentage is less than 75%.

The analysis of proportions is carried out in experiments or observational studies where the response variable is binary. Each observation is an outcome from a Bernoulli trial with a fixed probability  $p$  of observing an event of interest in a population. Hypothesis testing of binomial outcomes relies on a set of assumptions: 1) Bernoulli outcome is observed a fixed number of times; 2) the probability  $p$  is fixed across all trials; and 3) individual trials are independent.

This entry describes power and sample-size analysis for the inference about the population proportion performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: p = p_0$  versus the two-sided alternative hypothesis  $H_a: p \neq p_0$ , the upper one-sided alternative  $H_a: p > p_0$ , or the lower one-sided alternative  $H_a: p < p_0$ .

Two common hypothesis tests for a one-sample proportion are the small-sample binomial test and the asymptotic (large-sample) normal test. The binomial test is based on the binomial distribution, the exact sampling distribution, of the test statistic and is commonly known as an *exact binomial* test. The asymptotic normal test is based on the large-sample normal approximation of the sampling distribution of the test statistic and is often referred to as a  $z$  test.

`power oneproportion` provides power and sample-size analysis for both the binomial and a large-sample  $z$  test of a one-sample proportion.

## Using power oneproportion

`power oneproportion` computes sample size, power, or target proportion for a one-sample proportion test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

`power oneproportion` performs power analysis for three different tests, which can be specified within the `test()` option. The default is a large-sample score test (`test(score)`), which approximates the sampling distribution of the test statistic by the standard normal distribution. You may instead request computations based on a large-sample Wald test by specifying the `test(wald)` option. For power determination, you can also request the small-sample binomial test by specifying the `test(binomial)` option. The binomial test is not available for the sample-size and effect-size determinations; see [example 7](#) for details.

To compute sample size, you must specify the proportions under the null and alternative hypotheses,  $p_0$  and  $p_a$ , respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the sample size in the `n()` option and the proportions under the null and alternative hypotheses,  $p_0$  and  $p_a$ , respectively.

Instead of the alternative proportion  $p_a$ , you may specify the difference  $p_a - p_0$  between the alternative proportion and the null proportion in the `diff()` option when computing sample size or power.

To compute effect size, the difference between the alternative and null proportions, and target proportion, you must specify the sample size in the `n()` option, the power in the `power()` option, the null proportion  $p_0$ , and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the target proportion is assumed to be larger than the specified null value. You can change the direction to lower, which means that the target proportion is assumed to be smaller than the specified null value, by specifying the `direction(lower)` option.

By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS] **unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

Some of `power oneproportion`'s computations require iteration. For example, for a large-sample  $z$  test, sample size for a two-sided test is obtained by iteratively solving a nonlinear power equation. The default initial value for the sample size for the iteration procedure is obtained using a closed-form one-sided formula. If desired, it may be changed by specifying the `init()` option. See [PSS] **power** for the descriptions of other options that control the iteration procedure.

In the following sections, we describe the use of `power oneproportion` accompanied with examples for computing sample size, power, and target proportion.

## Computing sample size

To compute sample size, you must specify the proportions under the null and alternative hypotheses,  $p_0$  and  $p_a$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a one-sample proportion test

Consider a study of osteoporosis in postmenopausal women from [Chow, Shao, and Wang \(2008, 55\)](#). The term “osteoporosis” refers to the decrease in bone mass that is most prevalent in postmenopausal women. Females diagnosed with osteoporosis have vertebral bone density more than 10% below the average bone density of women with similar demographic characteristics such as age, height, weight, and race.

The World Health Organization (WHO) defines osteoporosis as having the bone density value that is smaller than 2.5 standard deviations below the peak bone mass levels in young women. Suppose investigators wish to assess the effect of a new treatment on increasing the bone density for women diagnosed with osteoporosis. The treatment is deemed successful if a subject's bone density improves by more than one standard deviation of her measured bone density.

Suppose that previous studies have reported a response rate of 30% for women with increased bone density after treatment. Investigators expect the new treatment to generate a higher response rate of roughly 50%. The goal is to obtain the minimum required sample size to detect an alternative proportion of 0.5 using the test of  $H_0: p = 0.3$  versus  $H_a: p \neq 0.3$  with 80% power and 5% significance level. To compute sample size, we specify the null and alternative proportions after the command name:

```

. power oneproportion 0.3 0.5
Performing iteration ...
Estimated sample size for a one-sample proportion test
Score z test
Ho: p = p0 versus Ha: p != p0
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.2000
    p0 = 0.3000
    pa = 0.5000
Estimated sample size:
    N = 44

```

We find that at least 44 subjects are needed to detect a change in proportion from 0.3 to 0.5 with 80% power using a 5%-level two-sided test.

◀

### ► Example 2: Specifying the difference between proportions

Instead of the alternative proportion, we can specify the difference of  $0.05 - 0.03 = 0.2$  between the alternative proportion and the null proportion in the `diff()` option and obtain the same results:

```

. power oneproportion 0.3, diff(0.2)
Performing iteration ...
Estimated sample size for a one-sample proportion test
Score z test
Ho: p = p0 versus Ha: p != p0
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.2000
    p0 = 0.3000
    pa = 0.5000
    diff = 0.2000
Estimated sample size:
    N = 44

```

The difference between proportions is now also displayed in the output.

◀

### ► Example 3: Wald test

The default computation is based on a score test and thus uses the null proportion as the estimate of the true proportion in the formula for the standard error. We can request the computation based on a Wald test by specifying the `test(wald)` option. In this case, the alternative proportion will be used as an estimate of the true proportion in the formula for the standard error.

```

. power oneproportion 0.3 0.5, test(wald)
Performing iteration ...
Estimated sample size for a one-sample proportion test
Wald z test
Ho: p = p0 versus Ha: p != p0
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.2000
      p0 = 0.3000
      pa = 0.5000
Estimated sample size:
      N = 50

```

We find that the required sample size increases to 50 subjects.

◀

## Computing power

To compute power, you must specify the sample size in the `n()` option and the proportions under the null and alternative hypotheses,  $p_0$  and  $p_a$ , respectively.

### ▷ Example 4: Power of a one-sample proportion test

Continuing with [example 1](#), we will suppose that we are designing a new study and anticipate to obtain a sample of 30 subjects. To compute the power corresponding to this sample size given the study parameters from example 1, we specify the sample size of 30 in the `n()` option:

```

. power oneproportion 0.3 0.5, n(30)
Estimated power for a one-sample proportion test
Score z test
Ho: p = p0 versus Ha: p != p0
Study parameters:
      alpha = 0.0500
      N = 30
      delta = 0.2000
      p0 = 0.3000
      pa = 0.5000
Estimated power:
      power = 0.6534

```

As expected, with a smaller sample size, we achieve a lower power (only 65.34%).

◀

### ► Example 5: Multiple values of study parameters

To see the effect of sample size on power, we can specify a range of sample sizes in the `n()` option.

```
. power oneproportion 0.3 0.5, n(40(1)50)
Estimated power for a one-sample proportion test
Score z test
Ho: p = p0 versus Ha: p != p0
```

alpha	power	N	delta	p0	pa
.05	.7684	40	.2	.3	.5
.05	.7778	41	.2	.3	.5
.05	.787	42	.2	.3	.5
.05	.7958	43	.2	.3	.5
.05	.8043	44	.2	.3	.5
.05	.8124	45	.2	.3	.5
.05	.8203	46	.2	.3	.5
.05	.8279	47	.2	.3	.5
.05	.8352	48	.2	.3	.5
.05	.8422	49	.2	.3	.5
.05	.849	50	.2	.3	.5

As expected, power is an increasing function of the sample size.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).



### ► Example 6: Sign test

We can use `power oneproportion` to perform power and sample-size analysis for a nonparametric sign test comparing the median of a sample with a reference value. The sign test for comparing a median is simply a test of a binomial proportion with the reference (null) value of 0.5,  $H_0: p = 0.5$ .

For example, consider a study similar to the one described in [example 1](#). Suppose we want to test whether the median bone density exceeds a threshold value in a population of females who received a certain treatment. This is equivalent to testing whether the proportion  $p$  of bone-density values exceeding the threshold is greater than 0.5, that is,  $H_0: p = 0.5$  versus  $H_a: p > 0.5$ . Suppose that from previous studies such proportion was estimated to be 0.7. We anticipate to enroll 30 subjects and would like to compute the corresponding power of an upper one-sided small-sample binomial test to detect the change in proportion from 0.5 to 0.7.

```

. power oneproportion 0.5 0.7, n(30) test(binomial) onesided
Estimated power for a one-sample proportion test
Binomial test
Ho: p = p0 versus Ha: p > p0
Study parameters:
    alpha = 0.0500
      N = 30
    delta = 0.2000
      p0 = 0.5000
      pa = 0.7000
Estimated power and alpha:
    power = 0.7304
  actual alpha = 0.0494

```

For a sample size of 30 subjects, we obtain a power of 73% to detect the difference of 0.2 between the alternative and null values. In addition to power, `power oneproportion` also displays the actual (observed) significance level, which is 0.0494 in our example and is very close to the specified significance level of 0.05.

When the sampling distribution of the test statistic is discrete such as for the binomial test, the specified nominal significance level may not be possible to precisely achieve, because the space of the observed significance levels is discrete. As such, `power oneproportion` also displays the observed significance level given the specified sample size, power, and other study parameters. Also see [example 7](#).

◀

## ▶ Example 7: Saw-toothed power function

In [example 6](#), we briefly described one issue arising with power and sample-size analysis for the binomial test. The observed significance levels are discrete because the binomial sampling distribution of the test statistic is discrete. Another related issue arising because of the discrete nature of the sampling distribution is the nonmonotonic relationship between power and sample size—as the sample size increases, the corresponding power may not necessarily increase. The power function may have a so-called *saw-toothed* shape ([Chernick and Liu 2002](#)), where it increases initially, then drops, then increases again, and so on. See [figure 1](#) below for an example.

To demonstrate the issue, we return to [example 5](#) and plot powers for a range of sample size values between 45 and 60. We specify the `graph()` option to produce a graph and the `table()` option to produce a table; see [\[PSS\] power, graph](#) and [\[PSS\] power, table](#) for more details about the graphical and tabular outputs from `power`. Within `graph()`, we request that the reference line be plotted on the  $y$  axis at a power of 0.8 and that the data points be labeled with the corresponding sample sizes. Within `table()`, we specify the `formats()` suboption to display only three digits after the decimal point for the `power` and `alpha_a` columns. We also specify the `critvalues` option to display columns containing critical values in the table.



```
. power oneprop 0.3 0.5, n(45(1)60) test(binomial) critvalues
> table(, formats(alpha_a "%7.3f" power "%7.3f"))
> graph(ylines(0.8) plotopts(mlabel(N)))
```

Estimated power for a one-sample proportion test  
 Binomial test  
 Ho: p = p0 versus Ha: p != p0

alpha	alpha_a	power	N	delta	p0	pa	C_l	C_u
.05	0.034	0.724	45	.2	.3	.5	7	21
.05	0.035	0.769	46	.2	.3	.5	7	21
.05	0.037	0.809	47	.2	.3	.5	7	21
.05	0.026	0.765	48	.2	.3	.5	7	22
.05	0.042	0.804	49	.2	.3	.5	8	22
.05	0.031	0.760	50	.2	.3	.5	8	23
.05	0.031	0.799	51	.2	.3	.5	8	23
.05	0.033	0.834	52	.2	.3	.5	8	23
.05	0.037	0.795	53	.2	.3	.5	9	24
.05	0.037	0.830	54	.2	.3	.5	9	24
.05	0.038	0.860	55	.2	.3	.5	9	24
.05	0.028	0.825	56	.2	.3	.5	9	25
.05	0.043	0.855	57	.2	.3	.5	10	25
.05	0.044	0.881	58	.2	.3	.5	10	25
.05	0.032	0.851	59	.2	.3	.5	10	26
.05	0.033	0.877	60	.2	.3	.5	10	26

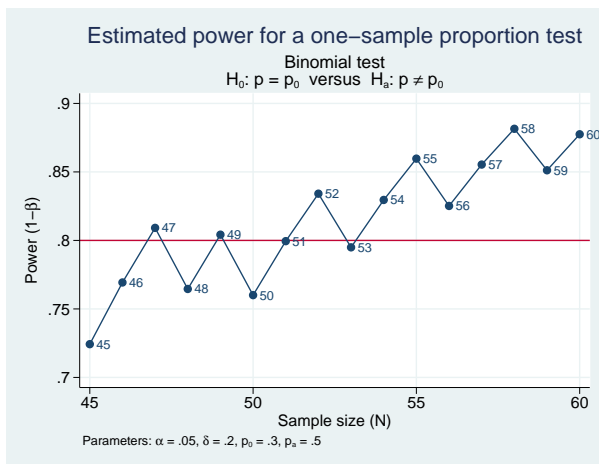


Figure 1. Saw-toothed power function

The power is not a monotonic function of the sample size. Also from the table, we can see that all the observed significance levels are smaller than the specified level of 0.05.

To better understand what is going on, we will walk through the steps of power determination. First, the critical values are determined as the minimum value  $C_l$  and the maximum value  $C_u$  between 0 and  $n$  that satisfy the following inequalities,

$$\Pr(X \leq C_l | p = p_0) \leq \alpha/2 \text{ and } \Pr(X \geq C_u | p = p_0) \leq \alpha/2$$

where the number of successes  $X$  has a binomial distribution with the total number of trials  $n$  and a probability of a success in a single trial  $p$ ,  $X \sim \text{Bin}(n, p)$ . The power is then computed as the sum of the above two probabilities with  $p = p_a$ .

For example, let's compute the power for the first row of the table. The sample size is 45, the lower critical value is 7, and the upper critical value is 21. We use the probability functions `binomial()` and `binomialtail()` to compute the respective lower- and upper-tailed probabilities of the binomial distribution.

```
. di "Lower tail: " binomial(45,7,0.3)
Lower tail: .0208653
. di "Upper tail: " binomialtail(45,21,0.3)
Upper tail: .01352273
. di "Obs. level: " binomial(45,7,0.3) + binomialtail(45,21,0.3)
Obs. level: .03438804
. di "Power:      " binomial(45,7,0.5) + binomialtail(45,21,0.5)
Power:        .7242594
```

Each of the tails is less than 0.025 ( $\alpha/2 = 0.05/2 = 0.025$ ). The observed significance level and power match the results from the first row of the table.

Now let's increase the lower critical value by one,  $C_l = 8$ , and decrease the upper critical value by one,  $C_u = 20$ :

```
. di "Lower tail: " binomial(45,8,0.3)
Lower tail: .04711667
. di "Upper tail: " binomialtail(45,20,0.3)
Upper tail: .02834511
```

Each of the tail probabilities now exceeds 0.025. If we could use values between 7 and 8 and between 20 and 21, we could match the tails exactly to 0.025, and then the monotonicity of the power function would be preserved. This is impossible for the binomial distribution (or any discrete distribution) because the number of successes must be integers.

Because of the saw-toothed nature of the power curve, obtaining an optimal sample size becomes tricky. If we wish to have power of 80%, then from the above table and graph, we see that potential sample sizes are 47, 49, 52, 54, and so on. One may be tempted to choose the smallest sample size for which the power is at least 80%. This, however, would not guarantee that the power is at least 80% for any larger sample size. Instead, [Chernick and Liu \(2002\)](#) suggest selecting the smallest sample size after which the troughs of the power curve do not go below the desired power. Following this recommendation in our example, we would pick a sample size of 54, which corresponds to the observed significance level of 0.037 and power of 0.83.

In the above, we showed the power curve for the sample sizes between 45 and 60. It may be a good idea to also look at the power plot for larger sample sizes to verify that the power continues to increase and does not drop below the desired power level.

◀

## Computing effect size and target proportion

In an analysis of a one-sample proportion, the effect size  $\delta$  is often defined as the difference between the alternative proportion and the null proportion,  $\delta = p_a - p_0$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding alternative or target proportion that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and null proportion must be specified. In addition, you must also

decide on the direction of the effect: upper, meaning  $p_a > p_0$ , or lower, meaning  $p_a < p_0$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ▷ Example 8: Minimum detectable value of the proportion

Continuing with [example 4](#), we may also be interested to find the minimum value of the proportion that can be detected with a power of 80% given a sample of 30 subjects. To compute this, after the command name, we specify the null proportion of 0.3, sample size `n(30)`, and power `power(0.8)`:

```
. power oneproportion 0.3, n(30) power(0.8)
Performing iteration ...
Estimated target proportion for a one-sample proportion test
Score z test
Ho: p = p0 versus Ha: p != p0; pa > p0
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      30
        p0 =    0.3000
Estimated effect size and target proportion:
      delta =    0.2406
         pa =    0.5406
```

The smallest detectable value of the proportion is 0.54.

In the above, we assumed the effect to be in the upper direction,  $p_a > p_0$ . We can obtain the results in the lower direction by specifying the `direction(lower)` option.

◀

## Performing hypothesis tests on proportion

In this section, we briefly demonstrate how you can test hypotheses about proportions; see [\[R\] prtest](#) and [\[R\] bitest](#) for details. Suppose we wish to test the hypothesis that the proportion is different from a reference value on the collected data. We can use the `prtest` command or the `bitest` command to do this.

### ▷ Example 9: Testing for proportion

We use `lbw.dta`, which contains data on birthweights of infants from a sample of 189 females. One of the variables in the dataset is variable `ui`, which records the presence or absence of uterine irritability. Although the real objective of this study is different, suppose we wish to test the null hypothesis that the proportion of women in a sample who experience uterine irritability is equal to 0.20. We can use the `prtest` command to perform a large-sample test of a single proportion.

```

. use http://www.stata-press.com/data/r14/lbw
(Hosmer & Lemeshow data)
. prtest ui==0.2
One-sample test of proportion                               ui: Number of obs =    189
-----+-----
Variable |           Mean   Std. Err.   [95% Conf. Interval]
-----+-----
      ui |   .1481481   .0258404   .0975019   .1987944
-----+-----
      p = proportion(ui)                                z =  -1.7821
Ho: p = 0.2
      Ha: p < 0.2                Ha: p != 0.2                Ha: p > 0.2
Pr(Z < z) = 0.0374             Pr(|Z| > |z|) = 0.0747             Pr(Z > z) = 0.9626

```

We do not have statistical evidence to reject the null hypothesis of  $H_0: p = 0.2$  versus a two-sided alternative  $H_a: p \neq 0.2$  at least at the 5% significance level; the  $p$ -value = 0.0747 > 0.05.

If our true objective were to study uterine irritability in the population of females, we would have performed the corresponding power and sample-size analysis before collecting the data. For example, using the estimates of `lbw.dta`, we can use `power oneproportion` to compute the required sample size for a 5%-level two-sided large-sample  $z$  test to detect the change in proportion from the reference value of 0.2 to approximately 0.148 with a power of, say, 80%:

```

. power oneproportion 0.2 0.148
Performing iteration ...
Estimated sample size for a one-sample proportion test
Score z test
Ho: p = p0 versus Ha: p != p0
Study parameters:
      alpha =   0.0500
      power =   0.8000
      delta =  -0.0520
      p0 =    0.2000
      pa =    0.1480
Estimated sample size:
      N =      434

```

We find that we need 434 subjects, many more than the current sample of 189, to detect the specified change in proportions.

◀

## Video examples

[Sample-size calculation for comparing a sample proportion to a reference value](#)

[Power calculation for comparing a sample proportion to a reference value](#)

[Minimum detectable effect size for comparing a sample proportion to a reference value using Stata](#)

## Stored results

power oneproportion stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(alpha_a)</code>	actual significance level of the binomial method
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(p0)</code>	proportion under the null hypothesis
<code>r(pa)</code>	proportion under the alternative hypothesis
<code>r(diff)</code>	difference between the alternative and null proportions
<code>r(C_l)</code>	lower critical value of the binomial distribution
<code>r(C_u)</code>	upper critical value of the binomial distribution
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or proportion
<code>r(continuity)</code>	1 if continuity correction is used; 0 otherwise
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	oneproportion
<code>r(test)</code>	score, wald, or binomial
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Let  $x_1, \dots, x_n$  be a sequence of  $n$  independent and identically distributed Bernoulli random variates. Let  $x_i = 1$  denote a success and  $x_i = 0$  denote a failure. Let  $P(x_i = 1) = p$  denote the probability of a success in the population. Each individual observation is a Bernoulli trial with a success probability  $p$ , which implies that the sum  $X = \sum_{i=1}^n x_i$  has a binomial distribution with mean  $np$  and standard deviation  $\sqrt{np(1-p)}$ . Let

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \text{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

denote the sample proportion and its standard error, respectively. Let  $p_0$  and  $p_a$  denote the null and alternative values of the proportion parameter, respectively.

A one-sample proportion test involves testing the null hypothesis  $H_0: p = p_0$  versus the two-sided alternative hypothesis  $H_a: p \neq p_0$ , the upper one-sided alternative  $H_a: p > p_0$ , or the lower one-sided alternative  $H_a: p < p_0$ .

If the `nfractional` option is not specified, the computed sample size is rounded up.

The following formulas are based on [Chow, Shao, and Wang \(2008\)](#).

Methods and formulas are presented under the following headings:

*Large-sample normal approximation*  
*Binomial test*

## Large-sample normal approximation

For a large sample, the distribution of the sample proportion  $\hat{p}$  may be approximated by the normal distribution with mean  $p$  and variance  $p(1-p)/n$ . Two test statistics are considered: the score test statistic  $z = (\hat{p} - p_0) / \sqrt{p_0(1-p_0)/n}$  and the Wald test statistic  $z = (\hat{p} - p_0) / \sqrt{\hat{p}(1-\hat{p})/n}$ . The score test statistic uses the null value of the proportion to construct the standard error, which leads to its sampling distribution being closer to the standard normal distribution than if the Wald statistic were used ([Agresti \[2013, 13\]](#)).

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha}$  and  $z_\beta$  be the  $(1-\alpha)$ th and the  $\beta$ th quantiles of the standard normal distribution. Also let  $\eta = \sqrt{\{p_0(1-p_0)\} / \{p_a(1-p_a)\}}$ .

The power  $\pi = 1 - \beta$  of the score  $z$  test is computed using

$$\pi = \begin{cases} \Phi\left(\frac{\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha}\eta\right) & \text{upper one sided} \\ \Phi\left(\frac{-\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha}\eta\right) & \text{lower one sided} \\ \Phi\left(\frac{\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha/2}\eta\right) + \Phi\left(\frac{-\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha/2}\eta\right) & \text{two sided} \end{cases} \quad (1)$$

where  $\Phi(\cdot)$  is the cdf of the standard normal distribution, and  $c$  is the normal-approximation continuity correction:  $c = 1/(2\sqrt{n})$  if the `continuity` option is specified, and  $c = 0$  otherwise.

The power of the Wald  $z$  test can be obtained from (1) by replacing the term  $p_0(1-p_0)$  in  $\eta$  with  $p_a(1-p_a)$  so that  $\eta = 1$ .

The sample size  $n$  for a one-sided test is computed using

$$n = \left\{ \frac{z_{1-\alpha}\sqrt{p_0(1-p_0)} + z_{1-\beta}\sqrt{p_a(1-p_a)}}{\delta} \right\}^2$$

If the `continuity` option is specified, the sample size  $n_c$  for a one-sided test is computed as

$$n_c = \frac{n}{4} \left( 1 + \sqrt{1 + \frac{2}{n|p_a - p_0|}} \right)^2$$

where  $n$  is the sample size computed without the correction ([Fleiss, Levin, and Paik 2003](#); [Levin and Chen 1999](#)).

The sample size for a two-sided test and minimum detectable value of the proportion are computed iteratively using the corresponding power equation from (1).

## Binomial test

Power of the binomial test is computed using the binomial (exact) sampling distribution of the test statistic. Consider a one-sided test given by

$$H_0: p = p_0 \quad \text{versus} \quad H_a: p > p_0$$

Let  $X$  denote the number of successes in the sample. The null hypothesis is rejected if  $X$  is greater than a critical value  $k$  such that the resulting  $p$ -value is less than or equal to the significance level  $\alpha$ .

The  $p$ -value for testing the above one-sided hypothesis can be obtained from the following equation:

$$P(X \geq k; n, p_0) = \sum_{i=k}^n \binom{n}{i} p_0^i (1 - p_0)^{n-i}$$

The  $p$ -value for testing the two-sided hypothesis  $H_a: p \neq p_0$  is given by

$$2 \times \min \{P(X \geq k; n, p_0), P(X \leq k; n, p_0)\}$$

For a one-sided test, the power of the test is computed from the following nonlinear equation:

$$\pi = 1 - \beta = \sum_{k=0}^n \binom{n}{k} p_a^k (1 - p_a)^{n-k} I \{P(X \geq k; n, p_0) \leq \alpha\}$$

Power for a two-sided test can be obtained by replacing the indicator function above with  $I [2 \times \min \{P(X \geq k; n, p_0), P(X \leq k; n, p_0)\} \leq \alpha]$ .

The computational details may be found in [Krishnamoorthy and Peng \(2007\)](#).

## References

- Agresti, A. 2013. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- Chernick, M. R., and C. Y. Liu. 2002. The saw-toothed behavior of power versus sample size and software solutions: Single binomial proportion using exact methods. *American Statistician* 56: 149–155.
- Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley.
- Krishnamoorthy, K., and J. Peng. 2007. Some properties of the exact and score methods for binomial proportion and sample size calculation. *Communications in Statistics—Simulation and Computation* 36: 1171–1186.
- Levin, B., and X. Chen. 1999. Is the one-half continuity correction used once or twice to derive a well-known approximate sample size formula to compare two independent binomial distributions? *American Statistician* 53: 62–66.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **bitest** — Binomial probability test
- [R] **prtest** — Tests of proportions

# Title

**power twoproportions** — Power analysis for a two-sample proportions test

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power twoproportions` computes sample size, power, or the experimental-group proportion for a two-sample proportions test. By default, it computes sample size for given power and the values of the control-group and experimental-group proportions. Alternatively, it can compute power for given sample size and values of the control-group and experimental-group proportions or the experimental-group proportion for given sample size, power, and the control-group proportion. Also see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \pi_1 = \pi_2$  vs  $H_a: \pi_1 \neq \pi_2$  given alternative control-group proportion  $p_1 = 0.8$ , alternative experimental-group proportion  $p_2 = 0.65$  with default power of 0.8 and significance level  $\alpha = 0.05$

```
power twoproportions .8 .65
```

Same as above, but specified as  $p_1$  and difference between proportions  $p_2 - p_1 = -0.15$

```
power twoproportions .8, diff(-.15)
```

Same as above, but specified as  $p_1 = 0.8$  and ratio  $p_2/p_1 = 0.8125$

```
power twoproportions .8, ratio(.8125)
```

Same as above, but specified as  $p_1 = 0.8$  and odds ratio  $\{p_2/(1-p_2)\}/\{p_1/(1-p_1)\} = 0.464$

```
power twoproportions .8, oratio(.464)
```

Power for sample sizes of 50 and 80 in groups 1 and 2, respectively

```
power twoproportions 0.8 0.65, n1(50) n2(80)
```

Power for total sample sizes of 150, 170, 190, 210, and 230

```
power twoproportions .8 .65, n(150(20)230)
```

As above, but display results in a graph of power versus sample size

```
power twoproportions .8 .65, n(150(20)230) graph
```

As above, but with difference equal to 0.1, 0.15, and 0.2

```
power twoproportions .65, n(150(20)200) graph diff(.1(.05).2)
```

Display results in a table showing total sample size, difference, and power

```
power twoproportions .65, n(150(20)200) table(N diff power)
```

Effect size and target  $p_2$  for  $p_1 = 0.6$  with sample size of 200, power of 0.8, and  $\alpha = 0.01$

```
power twoproportions .6, n(200) power(.8) alpha(.01)
```

## Menu

Statistics > Power and sample size



## Syntax

*Compute sample size*

```
power twoproportions  $p_1$   $p_2$  [ , power(numlist) options ]
```

*Compute power*

```
power twoproportions  $p_1$   $p_2$  , n(numlist) [options ]
```

*Compute effect size and experimental-group proportion*

```
power twoproportions  $p_1$  , n(numlist) power(numlist) [options ]
```

where  $p_1$  is the proportion in the control (reference) group, and  $p_2$  is the proportion in the experimental (comparison) group.  $p_1$  and  $p_2$  may each be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 **numlist**).

<i>options</i>	Description
<code>test(test)</code>	specify the type of test; default is <code>test(chi2)</code>
Main	
* <code>alpha(numlist)</code>	significance level; default is <code>alpha(0.05)</code>
* <code>power(numlist)</code>	power; default is <code>power(0.8)</code>
* <code>beta(numlist)</code>	probability of type II error; default is <code>beta(0.2)</code>
* <code>n(numlist)</code>	total sample size; required to compute power or effect size
* <code>n1(numlist)</code>	sample size of the control group
* <code>n2(numlist)</code>	sample size of the experimental group
* <code>nratio(numlist)</code>	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<code>compute(n1   n2)</code>	solve for $N1$ given $N2$ or for $N2$ given $N1$
<code>nfractional</code>	allow fractional sample sizes
* <code>diff(numlist)</code>	difference between the experimental-group and control-group proportions, $p_2 - p_1$ ; specify instead of the experimental-group proportion $p_2$
* <code>ratio(numlist)</code>	ratio of the experimental-group proportion to the control-group proportion, $p_2/p_1$ ; specify instead of the experimental-group proportion $p_2$
* <code>rdiff(numlist)</code>	risk difference, $p_2 - p_1$ ; synonym for <code>diff()</code>
* <code>rrisk(numlist)</code>	relative risk, $p_2/p_1$ ; synonym for <code>ratio()</code>
* <code>oratio(numlist)</code>	odds ratio, $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$
<code>effect(effect)</code>	specify the type of effect to display; default is <code>effect(diff)</code>
<code>continuity</code>	apply continuity correction to the normal approximation of the discrete distribution
<code>direction(upper lower)</code>	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<code>onesided</code>	one-sided test; default is two sided
<code>parallel</code>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<code>[no]table[ (tablespec) ]</code>	suppress table or display results as a table; see [PSS] <b>power, table</b>
<code>saving(filename [ , replace ])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph[ (graphopts) ]</code>	graph results; see [PSS] <b>power, graph</b>

Iteration

<code>init(#)</code>	initial value for sample sizes or experimental-group proportion
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no] log</code>	suppress or display iteration log
<code>[no] dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 [numlist](#).

`notitle` does not appear in the dialog box.

<i>test</i>	Description
<code>chi2</code>	Pearson's chi-squared test; the default
<code>lrchi2</code>	likelihood-ratio test
<code>fisher</code>	Fisher–Irwin's exact conditional test

`test()` does not appear in the dialog box. The dialog box selected is determined by the `test()` specification.

<i>effect</i>	Description
<code>diff</code>	difference between proportions, $p_2 - p_1$ ; the default
<code>ratio</code>	ratio of proportions, $p_2/p_1$
<code>rdiff</code>	risk difference, $p_2 - p_1$
<code>rrisk</code>	relative risk, $p_2/p_1$
<code>oratio</code>	odds ratio, $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>alpha_a</code>	observed significance level	$\alpha_a$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N1</code>	number of subjects in the control group	$N_1$
<code>N2</code>	number of subjects in the experimental group	$N_2$
<code>nratio</code>	ratio of sample sizes, experimental to control	$N_2/N_1$
<code>delta</code>	effect size	$\delta$
<code>p1</code>	control-group proportion	$p_1$
<code>p2</code>	experimental-group proportion	$p_2$
<code>diff</code>	difference between the experimental-group proportion and the control-group proportion	$p_2 - p_1$
<code>ratio</code>	ratio of the experimental-group proportion to the control-group proportion	$p_2/p_1$
<code>rdiff</code>	risk difference	$p_2 - p_1$
<code>rrisk</code>	relative risk	$p_2/p_1$
<code>oratio</code>	odds ratio	$\theta$
<code>target</code>	target parameter; synonym for <code>p2</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Column `alpha_a` is available when the `test(fisher)` option is specified.

Columns `nratio`, `diff`, `ratio`, `rdiff`, `rrisk`, and `oratio` are shown in the default table if specified.

## Options

`test(test)` specifies the type of the test for power and sample-size computations. *test* is one of `chi2`, `lrchi2`, or `fisher`.

`chi2` requests computations for the Pearson's  $\chi^2$  test. This is the default test.

`lrchi2` requests computations for the likelihood-ratio test.

`fisher` requests computations for Fisher–Irwin's exact conditional test. Iteration options are not allowed with this test.

Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `compute()`, `nfractional`; see [PSS] [power](#).

`diff(numlist)` specifies the difference between the experimental-group proportion and the control-group proportion,  $p_2 - p_1$ . You can specify either the experimental-group proportion  $p_2$  as a command argument or the difference between the two proportions in `diff()`. If you specify `diff(#)`, the experimental-group proportion is computed as  $p_2 = p_1 + \#$ . This option is not allowed with the effect-size determination and may not be combined with `ratio()`, `rdiff()`, `rrisk()`, or `oratio()`.

`ratio(numlist)` specifies the ratio of the experimental-group proportion to the control-group proportion,  $p_2/p_1$ . You can specify either the experimental-group proportion  $p_2$  as a command argument or

the ratio of the two proportions in `ratio()`. If you specify `ratio(#)`, the experimental-group proportion is computed as  $p_2 = p_1 \times \#$ . This option is not allowed with the effect-size determination and may not be combined with `diff()`, `rdiff()`, `rrisk()`, or `oratio()`.

`rdiff(numlist)` specifies the risk difference  $p_2 - p_1$ . This is a synonym for the `diff()` option, except the results are labeled as risk differences. This option is not allowed with the effect-size determination and may not be combined with `diff()`, `ratio()`, `rrisk()`, or `oratio()`.

`rrisk(numlist)` specifies the relative risk or risk ratio  $p_2 - p_1$ . This is a synonym for the `ratio()` option, except the results are labeled as relative risks. This option is not allowed with the effect-size determination and may not be combined with `diff()`, `ratio()`, `rdiff()`, or `oratio()`.

`oratio(numlist)` specifies the odds ratio  $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$ . You can specify either the experimental-group proportion  $p_2$  as a command argument or the odds ratio in `oratio()`. If you specify `oratio(#)`, the experimental-group proportion is computed as  $p_2 = 1/\{1 + (1 - p_1)/(p_1 \times \#)\}$ . This option is not allowed with the effect-size determination and may not be combined with `diff()`, `ratio()`, `rdiff()`, or `rrisk()`.

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. *effect* is one of `diff`, `ratio`, `rdiff`, `rrisk`, or `oratio`. By default, the effect size `delta` is the difference between proportions. If `diff()`, `ratio()`, `rdiff()`, `rrisk()`, or `oratio()` is specified, the effect size `delta` will contain the effect corresponding to the specified option. For example, if `oratio()` is specified, `delta` will contain the odds ratio.

`continuity` requests that continuity correction be applied to the normal approximation of the discrete distribution. `continuity` cannot be specified with `test(fisher)` or `test(lrchi2)`.

`direction()`, `onesided`, `parallel`; see [PSS] [power](#).

#### Table

`table`, `table()`, `notable`; see [PSS] [power, table](#).

`saving()`; see [PSS] [power](#).

#### Graph

`graph`, `graph()`; see [PSS] [power, graph](#). Also see the *column* table for a list of symbols used by the graphs.

#### Iteration

`init(#)` specifies the initial value for the estimated parameter. For sample-size determination, the estimated parameter is either the control-group size  $n_1$  or, if `compute(n2)` is specified, the experimental-group size  $n_2$ . For the effect-size determination, the estimated parameter is the experimental-group proportion  $p_2$ . The default initial values for sample sizes for a two-sided test are based on the corresponding one-sided large-sample  $z$  test with the significance level  $\alpha/2$ . The default initial value for the experimental-group proportion is computed using the bisection method.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power twoproportions` but is not shown in the dialog box: `notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power twoproportions*

*Alternative ways of specifying effect*

*Computing sample size*

*Computing power*

*Computing effect size and experimental-group proportion*

*Testing a hypothesis about two independent proportions*

*Video examples*

This entry describes the `power twoproportions` command and the methodology for power and sample-size analysis for a two-sample proportions test. See [\[PSS\] intro](#) for a general introduction to power and sample-size analysis and [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

The comparison of two independent proportions arises in studies involving two independent binomial populations. There are many examples of studies where a researcher would like to compare two independent proportions. A pediatrician might be interested in the relationship between low birthweight and the mothers' use of a particular drug during pregnancy. He or she would like to test the null hypothesis that there is no difference in the proportion of low-birthweight babies for mothers who took the drug and mothers who did not. A drug manufacturer may want to test the developed new topical treatment for a foot fungus by testing the null hypothesis that the proportion of successfully treated patients is the same in the treatment and placebo groups.

Hypothesis testing of binomial outcomes relies on a set of assumptions: 1) a Bernoulli outcome is observed a fixed number of times; 2) the probability  $p$  of observing an event of interest in one trial is fixed across all trials; and 3) individual trials are independent. Each of the two populations must conform to the assumptions of a binomial distribution.

This entry describes power and sample-size analysis for the inference about two population proportions performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: p_2 = p_1$  versus the two-sided alternative hypothesis  $H_a: p_2 \neq p_1$ , the upper one-sided alternative  $H_a: p_2 > p_1$ , or the lower one-sided alternative  $H_a: p_2 < p_1$ .

The large-sample Pearson's  $\chi^2$  and likelihood-ratio tests are commonly used to test hypotheses about two independent proportions. The test of [Fisher \(1935\)](#) and [Irwin \(1935\)](#) is commonly used to compare the two proportions in small samples.

The `power twoproportions` command provides power and sample-size analysis for these three tests. For Fisher's exact test, the direct computation is available only for the power of the test. Estimates of the sample size and effect size for Fisher's exact test are difficult to compute directly because of the discrete nature of the sampling distribution of the test statistic. They can, however, be obtained indirectly on the basis of the power computation; see [example 8](#) for details.

## Using power twoproportions

`power twoproportions` computes sample size, power, or experimental-group proportion for a two-sample proportions test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default,

all computations assume a balanced or equal-allocation design; see [PSS] [unbalanced designs](#) for a description of how to specify an unbalanced design.

`power twoproportions` performs power analysis for three different tests, which can be specified within the `test()` option. The default is Pearson's  $\chi^2$  test (`test(chi2)`), which approximates the sampling distribution of the test statistic by the standard normal distribution. You may instead request computations based on the likelihood-ratio test by specifying the `test(lrchi2)` option. To request Fisher's exact conditional test based on the hypergeometric distribution, you can specify `test(fisher)`. The `fisher` method is not available for computing sample size or effect size; see [example 8](#) for details.

To compute the total sample size, you must specify the control-group proportion  $p_1$ , the experimental-group proportion  $p_2$ , and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

Instead of the total sample size, you can compute one of the group sizes given the other one. To compute the control-group sample size, you must specify the `compute(n1)` option and the sample size of the experimental group in the `n2()` option. Likewise, to compute the experimental-group sample size, you must specify the `compute(n2)` option and the sample size of the control group in the `n1()` option.

To compute power, you must specify the total sample size in the `n()` option, the control-group proportion  $p_1$ , and the experimental-group proportion  $p_2$ .

Instead of the experimental-group proportion  $p_2$ , you can specify other alternative measures of effect when computing sample size or power; see [Alternative ways of specifying effect](#) below.

To compute effect size and the experimental-group proportion, you must specify the total sample size in the `n()` option, the power in the `power()` option, the control-group proportion  $p_1$ , and optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the experimental-group proportion is assumed to be larger than the specified control-group value. You can change the direction to be lower, which means that the experimental-group proportion is assumed to be smaller than the specified control-group value, by specifying the `direction(lower)` option.

There are multiple definitions of the effect size for a two-sample proportions test. The `effect()` option specifies what definition `power twoproportions` should use when reporting the effect size, which is labeled as `delta` in the output of the `power` command. The available definitions are the difference between the experimental-group proportion and the control-group proportion (`diff`), the ratio of the experimental-group proportion to the control-group proportion (`ratio`), the risk difference  $p_2 - p_1$  (`rdiff`), the relative risk  $p_2/p_1$  (`rrisk`), and the odds ratio  $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$  (`oratio`). When `effect()` is specified, the effect size `delta` contains the estimate of the corresponding effect and is labeled accordingly. By default, `delta` corresponds to the difference between proportions. If any of the options `diff()`, `ratio()`, `rdiff()`, `rrisk()`, or `oratio()` are specified and `effect()` is not specified, `delta` will contain the effect size corresponding to the specified option.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see [Two samples](#) in [PSS] [unbalanced designs](#) for more details.

## Alternative ways of specifying effect

As we mentioned above, `power twoproportions` provides a number of ways to specify the disparity between the control-group and experimental-group proportions for sample-size and power determinations.

You can specify the control-group proportion  $p_1$  and the experimental-group proportion  $p_2$  directly, after the command name:

```
power twoproportions p1 p2, ...
```

For this specification, the default effect size `delta` displayed by the `power` command is the difference  $p_2 - p_1$  between the proportions. You can use the `effect()` option to request another type of effect. For example, if you specify `effect(oratio)`,

```
power twoproportions p1 p2, effect(oratio) ...
```

the effect size `delta` will correspond to the odds ratio.

Alternatively, you can specify the control-group proportion  $p_1$  and one of the options `diff()`, `ratio()`, `rdiff()`, `rrisk()`, or `oratio()`. For these specifications, the effect size `delta` will contain the effect corresponding to the option. If desired, you can change this by specifying the `effect()` option.

Specify difference  $p_2 - p_1$  between the two proportions:

```
power twoproportions p1, diff(numlist) ...
```

Specify risk difference  $p_2 - p_1$ :

```
power twoproportions p1, rdiff(numlist) ...
```

Specify ratio  $p_2/p_1$  of the two proportions:

```
power twoproportions p1, ratio(numlist) ...
```

Specify relative risk or risk ratio  $p_2/p_1$ :

```
power twoproportions p1, rrisk(numlist) ...
```

Specify odds ratio  $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$ :

```
power twoproportions p1, oratio(numlist) ...
```

In the following sections, we describe the use of `power twoproportions` accompanied by examples for computing sample size, power, and experimental-group proportions.

## Computing sample size

To compute sample size, you must specify the control-group proportion  $p_1$ , the experimental-group proportion  $p_2$ , and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-sample proportions test

Consider a study investigating the effectiveness of aspirin in reducing the mortality rate due to myocardial infarction (heart attacks). Let  $p_A$  denote the proportion of deaths for aspirin users in the population and  $p_N$  denote the corresponding proportion for nonusers. We are interested in testing the null hypothesis  $H_0: p_A - p_N = 0$  against the two-sided alternative hypothesis  $H_a: p_A - p_N \neq 0$ .



Previous studies indicate that the proportion of deaths due to heart attacks is 0.015 for nonusers and 0.001 for users. Investigators wish to determine the minimum sample size required to detect an absolute difference of  $|0.001 - 0.015| = 0.014$  with 80% power using a two-sided 5%-level test.

To compute the required sample size, we specify the values 0.015 and 0.001 as the control- and experimental-group proportions after the command name. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults.

```
. power twoproportions 0.015 0.001
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.0140 (difference)
      p1 =     0.0150
      p2 =     0.0010
Estimated sample sizes:
      N =      1270
      N per group =    635
```

A total sample of 1,270 individuals, 635 individuals per group, must be obtained to detect an absolute difference of 0.014 between proportions of aspirin users and nonusers with 80% power using a two-sided 5%-level Pearson's  $\chi^2$  test.

◀

## ► Example 2: Alternative ways of specifying effect

The displayed effect size `delta` in [example 1](#) is the difference between the experimental-group proportion and the control-group proportion. We can redefine the effect size to be, for example, the odds ratio by specifying the `effect(oratio)` option.

```
. power twoproportions 0.015 0.001, effect(oratio)
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.0657 (odds ratio)
      p1 =     0.0150
      p2 =     0.0010
Estimated sample sizes:
      N =      1270
      N per group =    635
```

The effect size `delta` now contains the estimated odds ratio and is labeled correspondingly.

Instead of the estimate of the proportion in the experimental group, we may have an estimate of the odds ratio  $\frac{p_2(1-p_1)}{p_1(1-p_2)}$ . For example, the estimate of the odds ratio in this example is 0.0657. We can specify the value of the odds ratio in the `oratio()` option instead of specifying the experimental-group proportion 0.001:

```

. power twoproportions 0.015, oratio(0.0657)
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.0657 (odds ratio)
      p1 =     0.0150
      p2 =     0.0010
      odds ratio = 0.0657
Estimated sample sizes:
      N =      1270
      N per group = 635

```

The results are identical to the prior results. The estimate of the odds ratio is now displayed in the output, and the effect size `delta` now corresponds to the odds ratio.

We can also specify the following measures as input parameters: difference between proportions in the `diff()` option, risk difference in the `rdiff()` option, ratio of the proportions in the `ratio()` option, or relative risk in the `rrisk()` option.

◀

### ▷ Example 3: Likelihood-ratio test

Instead of the Pearson's  $\chi^2$  test as in [example 1](#), we can compute sample size for the likelihood-ratio test by specifying the `test(lrchi2)` option.

```

. power twoproportions 0.015 0.001, test(lrchi2)
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Likelihood-ratio test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.0140 (difference)
      p1 =     0.0150
      p2 =     0.0010
Estimated sample sizes:
      N =      1062
      N per group = 531

```

The required total sample size of 1,062 is smaller than that for the Pearson's  $\chi^2$  test.

◀

### ▷ Example 4: Computing one of the group sizes

Suppose we anticipate a sample of 600 aspirin users and wish to compute the required number of nonusers given the study parameters from [example 1](#). We specify the number of aspirin users in `n2()`, and we also include `compute(n1)`:

```
. power twoproportions 0.015 0.001, n2(600) compute(n1)
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.0140 (difference)
      p1 =     0.0150
      p2 =     0.0010
      N2 =      600
Estimated sample sizes:
      N =     1317
      N1 =      717
```

We require a sample of 717 nonusers given 600 aspirin users for a total of 1,317 subjects. The total number of subjects is larger for this unbalanced design compared with the corresponding balanced design in [example 1](#).

◀

### ► Example 5: Unbalanced design

By default, `power twoproportions` computes sample size for a balanced or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size for an unbalanced design by specifying the `nratio()` option.

Continuing with [example 1](#), we will suppose that we anticipate to recruit twice as many aspirin users as nonusers; that is,  $n_2/n_1 = 2$ . We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```
. power twoproportions 0.015 0.001, nratio(2)
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.0140 (difference)
      p1 =     0.0150
      p2 =     0.0010
      N2/N1 =    2.0000
Estimated sample sizes:
      N =     1236
      N1 =      412
      N2 =      824
```

We need a total sample size of 1,236 subjects.

Also see *Two samples* in [PSS] [unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.

◀

## Computing power

To compute power, you must specify the total sample size in the `n()` option, the control-group proportion  $p_1$ , and the experimental-group proportion  $p_2$ .

### ► Example 6: Power of a two-sample proportions test

Continuing with [example 1](#), we will suppose that we anticipate a sample of only 1,100 subjects. To compute the power corresponding to this sample size given the study parameters from [example 1](#), we specify the sample size in `n()`:

```
. power twoproportions 0.015 0.001, n(1100)
Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
         N =     1100
  N per group =    550
      delta =   -0.0140 (difference)
         p1 =    0.0150
         p2 =    0.0010
Estimated power:
      power =    0.7416
```

With a smaller sample of 1,100 subjects, we obtain a lower power of 74% compared with [example 1](#). ◀

### ► Example 7: Multiple values of study parameters

In this example, we would like to assess the effect of varying the proportion of aspirin users on the power of our study. Suppose that the total sample size is 1,100 with equal allocation between groups, and the value of the proportion in the nonusing group is 0.015. We specify a range of proportions for aspirin users from 0.001 to 0.009 with a step size of 0.001 as a *numlist* in parentheses as the second argument of the command:

```
. power twoproportions 0.015 (0.001(0.001)0.009), n(1100)
Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
```

alpha	power	N	N1	N2	delta	p1	p2
.05	.7416	1100	550	550	-.014	.015	.001
.05	.6515	1100	550	550	-.013	.015	.002
.05	.5586	1100	550	550	-.012	.015	.003
.05	.4683	1100	550	550	-.011	.015	.004
.05	.3846	1100	550	550	-.01	.015	.005
.05	.3102	1100	550	550	-.009	.015	.006
.05	.2462	1100	550	550	-.008	.015	.007
.05	.1928	1100	550	550	-.007	.015	.008
.05	.1497	1100	550	550	-.006	.015	.009

From the table, the power decreases from 74% to 15% as the proportion of deaths for aspirin users increases from 0.001 to 0.009 or the absolute value of the effect size (measured as the difference between the proportion of deaths for aspirin users and that for nonusers) decreases from 0.014 to 0.006.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## ▶ Example 8: Saw-toothed power function

We can also compute power for the small-sample Fisher's exact conditional test. The sampling distribution of the test statistic for this test is discrete. As such, Fisher's exact test shares the same issues arising with power and sample-size analysis as described in detail for the binomial one-sample proportion test in [example 7](#) of [PSS] [power oneproportion](#). In particular, the power function of Fisher's exact test has a saw-toothed shape as a function of the sample size. Here, we demonstrate the saw-toothed shape of the power function and refer you to [example 7](#) of [PSS] [power oneproportion](#) for details.

Let's plot powers of the Fisher's exact test for a range of experimental-group sizes between 50 and 65 given the control-group proportion of 0.6, the experimental-group proportion of 0.25, and the control-group size of 25. We specify the `graph()` option to produce a graph and the `table()` option to produce a table; see [PSS] [power, graph](#) and [PSS] [power, table](#) for more details about the graphical and tabular outputs from `power`. Within `graph()`, we specify options to request that the reference line be plotted on the  $y$  axis at a power of 0.8 and that the data points be labeled with the corresponding sample sizes. Within `table()`, we specify the `formats()` option to display only three digits after the decimal point for the power and `alpha_a` columns.

```
. power twoproportions 0.6 0.25, test(fisher) n1(25) n2(50(1)65)
> graph(yline(0.8) plotopts(mlabel(N)))
> table(, formats(alpha_a "%7.3f" power "%7.3f"))
Estimated power for a two-sample proportions test
Fisher's exact test
Ho: p2 = p1 versus Ha: p2 != p1
```

alpha	alpha_a	power	N	N1	N2	delta	p1	p2
.05	0.026	0.771	75	25	50	-.35	.6	.25
.05	0.025	0.793	76	25	51	-.35	.6	.25
.05	0.026	0.786	77	25	52	-.35	.6	.25
.05	0.026	0.782	78	25	53	-.35	.6	.25
.05	0.025	0.804	79	25	54	-.35	.6	.25
.05	0.026	0.793	80	25	55	-.35	.6	.25
.05	0.028	0.786	81	25	56	-.35	.6	.25
.05	0.025	0.814	82	25	57	-.35	.6	.25
.05	0.025	0.802	83	25	58	-.35	.6	.25
.05	0.025	0.797	84	25	59	-.35	.6	.25
.05	0.028	0.823	85	25	60	-.35	.6	.25
.05	0.026	0.813	86	25	61	-.35	.6	.25
.05	0.026	0.807	87	25	62	-.35	.6	.25
.05	0.025	0.819	88	25	63	-.35	.6	.25
.05	0.028	0.821	89	25	64	-.35	.6	.25
.05	0.027	0.816	90	25	65	-.35	.6	.25

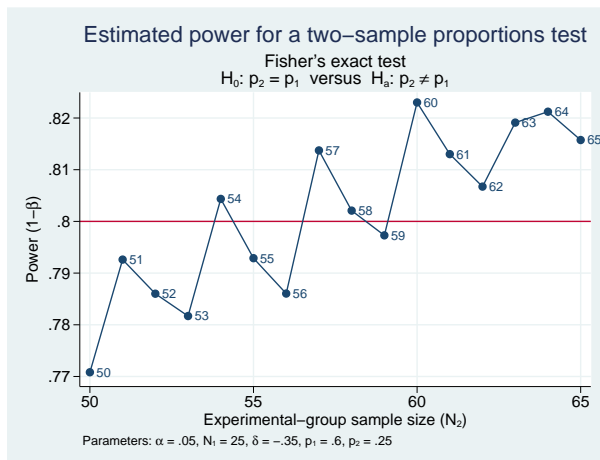


Figure 1. Saw-toothed power function

The power is not a monotonic function of the sample size. Also from the table, we see that all the observed significance levels are smaller than the specified level of 0.05.

Because of the saw-toothed nature of the power curve, obtaining an optimal sample size becomes tricky. For example, if we wish to have power of 80%, then from the above table and graph, we see that potential experimental-group sample sizes are 54, 57, 58, 60, and so on. One may be tempted to choose the smallest sample size for which the power is at least 80%. This, however, would not guarantee that the power is at least 80% for any larger sample size. Instead, [Chernick and Liu \(2002\)](#) suggest selecting the smallest sample size after which the troughs of the power curve do not go below the desired power. Following this recommendation in our example, we would pick a sample size of 60, which corresponds to the observed significance level of 0.028 and power of 0.823.

◀

## Computing effect size and experimental-group proportion

There are multiple definitions of the effect size for a two-sample proportions test. By default, effect size  $\delta$  is defined as the difference between the experimental-group proportion and the control-group proportion,  $\delta = p_2 - p_1$ , also known as a risk difference. Other available measures of the effect size are the ratio of the experimental-group proportion to the control-group proportion  $\delta = p_2/p_1$ , also known as a relative risk or risk ratio, and the odds ratio  $\delta = \{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding experimental-group proportion that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and control-group proportion must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $p_2 > p_1$ , or lower, meaning  $p_2 < p_1$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

The underlying computations solve the corresponding power equation for the value of the experimental-group proportion given power, sample size, and other study parameters. The effect size is then computed from the specified control-group proportion and the computed experimental-group proportion using the corresponding functional relationship. The difference between proportions is reported by default, but you can request other measures by specifying the `effect()` option.

### ▷ Example 9: Minimum detectable change in the experimental-group proportion

Continuing with [example 6](#), we will compute the smallest change in the proportion of deaths for aspirin users less than that for nonusers that can be detected given a total sample of 1,100 individuals and 80% power. To solve for the proportion of aspirin users in the experimental group, after the command name, we specify the control group (nonaspirin users), proportion of 0.015, total sample size `n(1100)`, and power `power(0.8)`:

```
. power twoproportions 0.015, n(1100) power(0.8) direction(lower)
Performing iteration ...
Estimated experimental-group proportion for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1; p2 < p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =     1100
  N per group =    550
         p1 =    0.0150
Estimated effect size and experimental-group proportion:
      delta =  -0.0147 (difference)
         p2 =    0.0003
```

We find that given the proportion of nonusers of 0.015, the smallest (in absolute value) difference between proportions that can be detected in this study is  $-0.0147$ , which corresponds to the proportion of aspirin users of 0.0003.

Although the difference between proportions is reported by default, we can request that another risk measure be reported by specifying the `effect()` option. For example, we can request that the odds ratio be reported instead:

```
. power twoproportions 0.015, n(1100) power(0.8) direction(lower) effect(oratio)
Performing iteration ...
Estimated experimental-group proportion for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1; p2 < p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =     1100
  N per group =    550
         p1 =    0.0150
Estimated effect size and experimental-group proportion:
      delta =    0.0195 (odds ratio)
         p2 =    0.0003
```

The corresponding value of the odds ratio in this example is 0.0195.

In these examples, we computed the experimental-group proportion assuming a lower direction,  $p_2 < p_1$ , which required you to specify the `direction(lower)` option. By default, experimental-group proportion is computed for an upper direction, meaning that the proportion is greater than the specified value of the control-group proportion.

**Testing a hypothesis about two independent proportions**

After the initial planning, we collect data and wish to test the hypothesis that the proportions from two independent populations are the same. We can use the `prtest` command to perform such hypothesis tests; see [R] `prtest` for details.

▷ **Example 10: Two-sample proportions test**

Consider a  $2 \times 3$  contingency table provided in table 2.1 of [Agresti \(2013, 38\)](#). The table is obtained from a report by the Physicians' Health Study Research Group at Harvard Medical School that investigated the relationship between aspirin use and heart attacks.

The report presents summary data on fatal and nonfatal heart attacks. In the current example, we combine these two groups into a single group representing the total cases with heart attacks for aspirin users and nonusers. The estimated proportion of heart attacks in the control group, nonaspirin users, is  $189/11034 = 0.0171$  and in the experimental group, aspirin users, is  $104/11037 = 0.0094$ .

```
. prtesti 11034 0.0171 11037 0.0094
```

```
Two-sample test of proportions
```

```
x: Number of obs = 11034
y: Number of obs = 11037
```

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
x	.0171	.0012342			.014681 .019519
y	.0094	.0009185			.0075997 .0112003
diff	.0077	.0015385			.0046846 .0107154
	under Ho:	.0015393	5.00	0.000	

```
diff = prop(x) - prop(y)
```

```
z = 5.0023
```

```
Ho: diff = 0
```

```
Ha: diff < 0
```

```
Ha: diff != 0
```

```
Ha: diff > 0
```

```
Pr(Z < z) = 1.0000
```

```
Pr(|Z| < |z|) = 0.0000
```

```
Pr(Z > z) = 0.0000
```

Let  $p_A$  and  $p_N$  denote the proportions of heart attacks in the population for aspirin users and nonusers, respectively. From the above results, we find a statistically significant evidence to reject the null hypothesis  $H_0: p_A = p_N$  versus a two-sided alternative  $H_a: p_A \neq p_N$  at the 5% significance level; the  $p$ -value is very small.

We use the parameters of this study to perform a sample-size analysis we would have conducted before the study.

```
. power twoproportions 0.0171 0.0094
```

```
Performing iteration ...
```

```
Estimated sample sizes for a two-sample proportions test
```

```
Pearson's chi-squared test
```

```
Ho: p2 = p1 versus Ha: p2 != p1
```

```
Study parameters:
```

```
alpha = 0.0500
```

```
power = 0.8000
```

```
delta = -0.0077 (difference)
```

```
p1 = 0.0171
```

```
p2 = 0.0094
```

```
Estimated sample sizes:
```

```
N = 6922
```

```
N per group = 3461
```



We find that for Pearson's  $\chi^2$  test, a total sample size of 6,922, assuming a balanced design, is required to detect the difference between the control-group proportion of 0.0171 and the experimental-group proportion of 0.0094 with 80% power using a 5%-level two-sided test.

◀

## Video examples

[How to calculate sample size for two independent proportions](#)

[How to calculate power for two independent proportions](#)

[How to calculate minimum detectable effect size for two independent proportions](#)

## Stored results

`power twoproportions` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(alpha_a)</code>	actual significance level of the Fisher's exact test
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(p1)</code>	control-group proportion
<code>r(p2)</code>	experimental-group proportion
<code>r(diff)</code>	difference between the experimental- and control-group proportions
<code>r(ratio)</code>	ratio of the experimental-group proportion to the control-group proportion
<code>r(rdifff)</code>	risk difference
<code>r(rrisk)</code>	relative risk
<code>r(oratio)</code>	odds ratio
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for sample sizes or experimental-group proportion
<code>r(continuity)</code>	1 if continuity correction is used; 0 otherwise
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>twoproportions</code>
<code>r(test)</code>	<code>chi2</code> , <code>lrchi2</code> , or <code>fisher</code>
<code>r(effect)</code>	specified effect: <code>diff</code> , <code>ratio</code> , etc.
<code>r(direction)</code>	<code>upper</code> or <code>lower</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Consider two independent samples of sizes  $n_1$  and  $n_2$ . Let  $x_{11}, \dots, x_{1n_1}$  be a random sample from a binomial distribution with the success probability  $p_1$ . We refer to this sample as a control group. Let  $x_{21}, \dots, x_{2n_2}$  be a random sample from a binomial distribution with the success probability  $p_2$ , the experimental group. Let  $x_{ij} = 1$  denote a success and  $x_{ij} = 0$  denote a failure. The two samples are assumed to be independent.

The sample proportions and their respective standard errors are

$$\hat{p}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \quad \text{and} \quad \text{se}(\hat{p}_1) = \sqrt{n_1 \hat{p}_1 (1 - \hat{p}_1)}$$

$$\hat{p}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i} \quad \text{and} \quad \text{se}(\hat{p}_2) = \sqrt{n_2 \hat{p}_2 (1 - \hat{p}_2)}$$

A two-sample proportions test involves testing the null hypothesis  $H_0: p_2 = p_1$  versus the two-sided alternative hypothesis  $H_a: p_2 \neq p_1$ , the upper one-sided alternative  $H_a: p_2 > p_1$ , or the lower one-sided alternative  $H_a: p_2 < p_1$ .

Let  $R = n_2/n_1$  denote the allocation ratio. Then  $n_2 = R \times n_1$  and power can be viewed as a function of  $n_1$ . Therefore, for sample-size determination, the control-group sample size  $n_1$  is computed first. The experimental-group size  $n_2$  is then computed as  $R \times n_1$ , and the total sample size is computed as  $n = n_1 + n_2$ . By default, sample sizes are rounded to integer values; see *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

The formulas below are based on [Fleiss, Levin, and Paik \(2003\)](#) and [Agresti \(2013\)](#).

Methods and formulas are presented under the following headings:

*Effect size*  
*Pearson's  $\chi^2$  test*  
*Likelihood-ratio test*  
*Fisher's exact conditional test*

## Effect size

The measures of risk or effect size can be defined in a number of ways for the two-sample proportions test. By default, the effect size is the difference between the experimental-group proportion and the control-group proportion. Other available risk measures include risk difference, risk ratio or relative risk, and odds ratio.

By default or when `effect(diff)` or `effect(rdif)` is specified, the effect size is computed as

$$\delta = p_2 - p_1$$

When `effect(ratio)` or `effect(rrisk)` is specified, the effect size is computed as

$$\delta = p_2/p_1$$

When `effect(oration)` is specified, the effect size is computed as

$$\delta = \theta = \{p_2(1 - p_1)\} / \{p_1(1 - p_2)\}$$

If `diff()`, `rdiff()`, `ratio()`, `rrisk()`, or `oration()` is specified, the value of the experimental-group proportion  $p_2$  is computed using the corresponding formula from above.

## Pearson's $\chi^2$ test

For a large sample size, a binomial process can be approximated by a normal distribution. The asymptotic sampling distribution of the test statistic

$$z = \frac{(\widehat{p}_2 - \widehat{p}_1) - (p_2 - p_1)}{\sqrt{\widehat{p}(1 - \widehat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

is standard normal, where  $\bar{p} = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$  is the pooled proportion and  $\widehat{p}$  is its estimator. The square of this statistic,  $z^2$ , has an approximate  $\chi^2$  distribution with one degree of freedom, and the corresponding test is known as Pearson's  $\chi^2$  test.

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha}$  and  $z_\beta$  be the  $(1 - \alpha)$ th and the  $\beta$ th quantiles of the standard normal distribution.

Let  $\sigma_D = \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$  be the standard deviation of the difference between proportions and  $\sigma_p = \sqrt{\bar{p}(1 - \bar{p})(1/n_1 + 1/n_2)}$  be the pooled standard deviation.

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi \left\{ \frac{(p_2 - p_1) - c - z_{1-\alpha} \sigma_p}{\sigma_D} \right\} & \text{for an upper one-sided test} \\ \Phi \left\{ \frac{-(p_2 - p_1) - c - z_{1-\alpha} \sigma_p}{\sigma_D} \right\} & \text{for a lower one-sided test} \\ \Phi \left\{ \frac{(p_2 - p_1) - c - z_{1-\alpha/2} \sigma_p}{\sigma_D} \right\} + \Phi \left\{ \frac{-(p_2 - p_1) - c - z_{1-\alpha/2} \sigma_p}{\sigma_D} \right\} & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $\Phi(\cdot)$  is the cdf of the standard normal distribution, and  $c$  is the normal-approximation continuity correction. For equal sample sizes,  $n_1 = n_2 = n/2$ , the continuity correction is expressed as  $c = 2/n$  (Levin and Chen 1999).

For a one-sided test, given the allocation ratio  $R = n_2/n_1$ , the total sample size  $n$  is computed by inverting the corresponding power equation in (1),

$$n = \frac{\left\{ z_{1-\alpha} \sqrt{\bar{p}(1 - \bar{p})} - z_\beta \sqrt{w_2 p_1(1 - p_1) + w_1 p_2(1 - p_2)} \right\}^2}{w_1 w_2 (p_2 - p_1)^2} \quad (2)$$

where  $w_1 = 1/(1 + R)$  and  $w_2 = R/(1 + R)$ . Then  $n_1$  and  $n_2$  are computed as  $n_1 = n/(1 + R)$  and  $n_2 = R \times n_1$ , respectively. If the `continuity` option is specified, the sample size  $n_c$  for a one-sided test is computed as

$$n_c = \frac{n}{4} \left( 1 + \sqrt{1 + \frac{2}{n w_1 w_2 |p_2 - p_1|}} \right)^2$$

where  $n$  is the sample size computed without the correction. For unequal sample sizes, the continuity correction generalizes to  $c = 1/(2nw_1w_2)$  (Fleiss, Levin, and Paik 2003).

For a two-sided test, the sample size is computed by iteratively solving the two-sided power equation in (1). The initial values for the two-sided computations are obtained from (2) with the significance level  $\alpha/2$ .

If one of the group sizes is known, the other one is computed by iteratively solving the corresponding power equation in (1). The initial values are obtained from (2) by assuming that  $R = 1$ .

The experimental-group proportion  $p_2$  is computed by iteratively solving the corresponding power equations in (1). The default initial values are obtained by using a bisection method.

### Likelihood-ratio test

Let  $q_1 = 1 - p_1$ ,  $q_2 = 1 - p_2$ , and  $\bar{q} = 1 - \bar{p} = 1 - (n_1p_1 + n_2p_2)/(n_1 + n_2)$ . The likelihood-ratio test statistic is given by

$$G = \sqrt{2n \left\{ \frac{n_1p_1}{n} \ln \left( \frac{p_1}{\bar{p}} \right) + \frac{n_1q_1}{n} \ln \left( \frac{q_1}{\bar{q}} \right) + \frac{n_2p_2}{n} \ln \left( \frac{p_2}{\bar{p}} \right) + \frac{n_2q_2}{n} \ln \left( \frac{q_2}{\bar{q}} \right) \right\}}$$

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi(G - z_{1-\alpha}) & \text{for an upper one-sided test} \\ \Phi(-G - z_{1-\alpha}) & \text{for a lower one-sided test} \\ \Phi(G - z_{1-\alpha/2}) + \Phi(-G - z_{1-\alpha/2}) & \text{for a two-sided test} \end{cases} \quad (3)$$

For a one-sided test, given the allocation ratio  $R$ , the total sample size  $n$  is computed by inverting the corresponding power equation in (3),

$$n = \frac{(z_{1-\alpha} - z_\beta)^2}{2 \left\{ w_1p_1 \ln \left( \frac{p_1}{\bar{p}} \right) + w_1q_1 \ln \left( \frac{q_1}{\bar{q}} \right) + w_2p_2 \ln \left( \frac{p_2}{\bar{p}} \right) + w_2q_2 \ln \left( \frac{q_2}{\bar{q}} \right) \right\}} \quad (4)$$

where  $w_1 = 1/(1 + R)$  and  $w_2 = R/(1 + R)$ . Then  $n_1$  and  $n_2$  are computed as  $n_1 = n/(1 + R)$  and  $n_2 = R \times n_1$ , respectively.

For a two-sided test, the sample size is computed by iteratively solving the two-sided power equation in (3). The initial values for the two-sided computations are obtained from equation (4) with the significance level  $\alpha/2$ .

If one of the group sizes is known, the other one is computed by iteratively solving the corresponding power equation in (3). The initial values are obtained from (4) by assuming that  $R = 1$ .

The experimental-group proportion  $p_2$  is computed by iteratively solving the corresponding power equations in (3). The default initial values are obtained by using a bisection method.

## Fisher's exact conditional test

Power computation for Fisher's exact test is based on [Casagrande, Pike, and Smith \(1978\)](#). We present formulas from the original paper with a slight change in notation: we use  $\tilde{p}_1$  and  $\tilde{p}_2$  in place of  $p_1$  and  $p_2$  and  $\eta$  in place of  $\theta$ . The change in notation is to avoid confusion between our use of group proportions  $p_1$  and  $p_2$  and their use in the paper—compared with our definitions, the roles of  $p_1$  and  $p_2$  in the paper are reversed in the definitions of the hypotheses and other measures such as the odds ratio. In our definitions,  $\tilde{p}_1 = p_2$  is the proportion in the experimental group,  $\tilde{p}_2 = p_1$  is the proportion in the control group, and  $\eta = 1/\theta$ . Also we denote  $\tilde{n}_1 = n_2$  to be the sample size of the experimental group and  $\tilde{n}_2 = n_1$  to be the sample size of the control group.

Let  $k$  be the number of successes in the experimental group, and let  $m$  be the total number of successes in both groups. The conditional distribution of  $k$  is given by

$$p(k|m, \eta) = \frac{\binom{\tilde{n}_1}{k} \binom{\tilde{n}_2}{m-k} \eta^k}{\sum_i \binom{\tilde{n}_1}{i} \binom{\tilde{n}_2}{m-i} \eta^i}$$

where  $\eta = \{\tilde{p}_1(1 - \tilde{p}_2)\} / \{\tilde{p}_2(1 - \tilde{p}_1)\}$ , and the range of  $i$  is given by  $L = \max(0, m - \tilde{n}_2)$  to  $U = \min(\tilde{n}_1, m)$ .

Assume an upper one-sided test given by

$$H_0: \tilde{p}_1 = \tilde{p}_2 \quad \text{versus} \quad H_a: \tilde{p}_1 > \tilde{p}_2 \quad (5)$$

The hypothesis (5) in terms of  $\eta$  can be expressed as follows:

$$H_0: \eta = 1 \quad \text{versus} \quad H_a: \eta > 1$$

Let  $k_u$  be the critical value of  $k$  such that the following inequalities are satisfied:

$$\sum_{i=k_u}^U p(i|m, \eta = 1) \leq \alpha \quad \text{and} \quad \sum_{i=k_u-1}^U p(i|m, \eta = 1) > \alpha \quad (6)$$

The conditional power is

$$\beta(\eta|m) = \sum_{i=k_u}^U p(i|m, \eta)$$

For a lower one-sided hypothesis  $H_a: \tilde{p}_1 < \tilde{p}_2$ , the corresponding hypothesis in terms of  $\eta$  is given by

$$H_0: \eta = 1 \quad \text{versus} \quad H_a: \eta < 1$$

The conditional power in this case is

$$\beta(\eta|m) = \sum_{i=L}^{k_l} p(i|m, \eta)$$

where  $k_l$  is the critical value of  $k$  such that the following inequalities are satisfied:

$$\sum_{i=L}^{k_l} p(i|m, \eta = 1) \leq \alpha \quad \text{and} \quad \sum_{i=L}^{k_l+1} p(i|m, \eta = 1) > \alpha \quad (7)$$

For a two-sided test, the critical values  $k_l$  and  $k_u$  are calculated using the inequalities (6) and (7) with  $\alpha/2$ , respectively.

Finally, the unconditional power is calculated as

$$\beta(\eta) = \sum_j \beta(\eta|j)P(j)$$

where  $j$  takes the value from 0 to  $n$ , and

$$P(j) = \sum_{i=L}^U \binom{\widetilde{n}_1}{i} \widetilde{p}_1^i (1 - \widetilde{p}_1)^{\widetilde{n}_1 - i} \binom{\widetilde{n}_2}{j - i} \widetilde{p}_2^{j - i} (1 - \widetilde{p}_2)^{\widetilde{n}_2 - j + i}$$

where  $L = \max(0, j - \widetilde{n}_2)$  and  $U = \min(\widetilde{n}_1, j)$ .

## References

- Agresti, A. 2013. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- Casagrande, J. T., M. C. Pike, and P. G. Smith. 1978. The power function of the “exact” test for comparing two binomial distributions. *Journal of the Royal Statistical Society, Series C* 27: 176–180.
- Chernick, M. R., and C. Y. Liu. 2002. The saw-toothed behavior of power versus sample size and software solutions: Single binomial proportion using exact methods. *American Statistician* 56: 149–155.
- Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh: Oliver & Boyd.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley.
- Irwin, J. O. 1935. Tests of significance for differences between percentages based on small numbers. *Metron* 12: 83–94.
- Levin, B., and X. Chen. 1999. Is the one-half continuity correction used once or twice to derive a well-known approximate sample size formula to compare two independent binomial distributions? *American Statistician* 53: 62–66.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **bitest** — Binomial probability test
- [R] **prtest** — Tests of proportions

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power pairedproportions` computes sample size, power, or target discordant proportions for a two-sample paired-proportions test, also known as McNemar's test. By default, it computes sample size for given power and values of the discordant or marginal proportions. Alternatively, it can compute power for given sample size and values of the discordant or marginal proportions, or it can compute the target discordant proportions for given sample size and power. Also see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

*Using discordant proportions*

Sample size for two-sided McNemar's test of marginal homogeneity  $H_0: \pi_{12} = \pi_{21}$  when  $p_{12} = 0.1$  and  $p_{21} = 0.2$  with default power of 0.8 and significance level  $\alpha = 0.05$

```
power pairedproportions .1 .2
```

Same as above, specified as  $p_{12} = 0.1$  and difference between proportions  $p_{21} - p_{12} = 0.1$

```
power pairedproportions .1, diff(.1)
```

Same as above, specified as  $p_{12} = 0.1$  and ratio  $p_{21}/p_{12} = 2$

```
power pairedproportions .1, ratio(2)
```

As above, but for ratios of 1.8, 1.9, 2.0, 2.1, and 2.2

```
power pairedproportions .1, ratio(1.8(.1)2.2)
```

As above, but display results as a graph of sample size versus ratio

```
power pairedproportions .1, ratio(1.8(.1)2.2) graph
```

For a one-sided test with  $\alpha = 0.01$  and sample size of 300

```
power pairedproportions .1 .2, alpha(0.01) onesided n(300)
```

Effect size and target discordant proportions for sample size of 200, power of 0.9, and sum of discordant proportions  $p_{12} + p_{21} = 0.4$

```
power pairedproportions, n(200) power(.9) prdiscordant(.4)
```

*Using marginal proportions*

Sample size using marginal proportions  $p_{+1} = 0.6$  and  $p_{1+} = 0.7$  with a correlation of 0.35 between paired observations for default power of 0.8 and  $\alpha = 0.05$

```
power pairedproportions .6 .7, corr(.35)
```

Same as above, specified as  $p_{+1} = 0.6$  and difference  $p_{1+} - p_{+1} = 0.1$

```
power pairedproportions .6, corr(.35) diff(.1)
```

Same as above, specified as  $p_{+1} = 0.6$  and relative risk  $p_{1+}/p_{+1} = 1.167$

```
power pairedproportions .6, corr(.35) rrisk(1.167)
```

Same as above, specified as  $p_{+1} = 0.6$  and odds ratio  $\{p_{1+}/(1 - p_{1+})\}/\{p_{+1}/(1 - p_{+1})\} = 1.556$

```
power pairedproportions .6, corr(.35) oratio(1.556)
```

Power for a sample size of 250

```
power pairedproportions .6 .7, corr(.35) n(250)
```

## Menu

Statistics > Power and sample size



## Syntax

Compute sample size

Specify discordant proportions

```
power pairedproportions  $p_{12}$   $p_{21}$  [ , power(numlist) discordopts ]
```

Specify marginal proportions

```
power pairedproportions  $p_{1+}$   $p_{+1}$  , corr(numlist) [ power(numlist) margopts ]
```

Compute power

Specify discordant proportions

```
power pairedproportions  $p_{12}$   $p_{21}$  , n(numlist) [ discordopts ]
```

Specify marginal proportions

```
power pairedproportions  $p_{1+}$   $p_{+1}$  , corr(numlist) n(numlist) [ margopts ]
```

Compute effect size and target discordant proportions

```
power pairedproportions , n(numlist) power(numlist) prdiscordant(numlist)  
[ discordopts ]
```

where  $p_{12}$  is the probability of a success at occasion 1 and a failure at occasion 2, and  $p_{21}$  is the probability of a failure at occasion 1 and a success at occasion 2. Each represents the probability of a discordant pair.  $p_{1+}$  is the marginal probability of a success for occasion 1, and  $p_{+1}$  is the marginal probability of a success for occasion 2. Each may be specified either as one number or as a list of values in parentheses; see [U] 11.1.8 **numlist**.

<i>discordopts</i>	Description
Main	
* <u>a</u> lpha( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>p</u> ower( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>b</u> eta( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>n</u> fractional	allow fractional sample size
* <u>p</u> rdiscordant( <i>numlist</i> )	sum of the discordant proportions, $p_{12} + p_{21}$
* <u>s</u> um( <i>numlist</i> )	synonym for <code>prdiscordant()</code>
* <u>d</u> iff( <i>numlist</i> )	difference between the discordant proportions, $p_{21} - p_{12}$
* <u>r</u> atio( <i>numlist</i> )	ratio of the discordant proportions, $p_{21}/p_{12}$
<u>e</u> ffect( <i>effect</i> )	specify the type of effect to display; default is <code>effect(diff)</code>
<u>d</u> irection( <u>u</u> pper  <u>l</u> ower)	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>o</u> nesided	one-sided test; default is two sided
<u>p</u> arallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<u>[no]</u> <u>t</u> able [ <i>(tablespec)</i> ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>s</u> aving( <i>filename</i> [, <i>replace</i> ])	save the table data to <i>filename</i> ; use <i>replace</i> to overwrite existing <i>filename</i>
Graph	
<u>g</u> raph [ <i>(graphopts)</i> ]	graph results; see [PSS] <b>power, graph</b>
Iteration	
<u>i</u> nit(#)	initial value for the sample size or the difference between discordant proportions
<u>i</u> terate(#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>t</u> olerance(#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>f</u> tolerance(#)	function tolerance; default is <code>ftolerance(1e-12)</code>
<u>[no]</u> <u>l</u> og	suppress or display iteration log
<u>[no]</u> <u>d</u> ots	suppress or display iterations as dots
<u>n</u> otitle	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 **numlist**.

`notitle` does not appear in the dialog box.

<i>margopts</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample size
* <u>corr</u> ( <i>numlist</i> )	correlation between the paired observations
* <u>diff</u> ( <i>numlist</i> )	difference between the marginal proportions, $p_{+1} - p_{1+}$
* <u>ratio</u> ( <i>numlist</i> )	ratio of the marginal proportions, $p_{+1}/p_{1+}$
* <u>rrisk</u> ( <i>numlist</i> )	relative risk, $p_{+1}/p_{1+}$
* <u>oratio</u> ( <i>numlist</i> )	odds ratio, $\{p_{+1}(1 - p_{1+})\}/\{p_{1+}(1 - p_{+1})\}$
<u>effect</u> ( <i>effect</i> )	specify the type of effect to display; default is <code>effect(diff)</code>
<u>direction</u> ( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [ , <i>replace</i> ])	save the table data to <i>filename</i> ; use <i>replace</i> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
<u>notitle</u>	suppress the title
*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 <b>numlist</b> . <i>notitle</i> does not appear in the dialog box.	

<i>effect</i>	Description
<u>diff</u>	difference between the discordant proportions, $p_{21} - p_{12}$ , or marginal proportions, $p_{+1} - p_{1+}$ ; the default
<u>ratio</u>	ratio of the discordant proportions, $p_{21}/p_{12}$ , or of the marginal proportions, $p_{+1}/p_{1+}$
<u>rrisk</u>	relative risk, $p_{+1}/p_{1+}$ ; may only be specified with marginal proportions
<u>oratio</u>	odds ratio, $\{p_{+1}(1 - p_{1+})\}/\{p_{1+}(1 - p_{+1})\}$ ; may only be specified with marginal proportions

where *tablespec* is

```
column[:label] [column[:label] [...]] [, tableopts]
```

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	number of subjects	$N$
<code>delta</code>	effect size	$\delta$
<code>p12</code>	success–failure proportion	$p_{12}$
<code>p21</code>	failure–success proportion	$p_{21}$
<code>pmarg1</code>	success proportion in occasion 1	$p_{1+}$
<code>pmarg2</code>	success proportion in occasion 2	$p_{+1}$
<code>corr</code>	correlation between paired observations	$\rho$
<code>prdiscordant</code>	proportion of discordant pairs	$p_{12} + p_{21}$
<code>sum</code>	sum of discordant proportions	$p_{12} + p_{21}$
<code>diff</code>	difference between discordant proportions	$p_{21} - p_{12}$
	difference between marginal proportions	$p_{+1} - p_{1+}$
<code>ratio</code>	ratio of discordant proportions	$p_{21}/p_{12}$
	ratio of marginal proportions	$p_{+1}/p_{1+}$
<code>rrisk</code>	relative risk for marginal proportions	$p_{+1}/p_{1+}$
<code>oratio</code>	odds ratio for marginal proportions	$\theta$
<code>target</code>	target parameter; synonym for <code>p12</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `p12` and `p21` are shown in the default table if discordant proportions are specified.

Columns `pmarg1`, `pmarg2`, and `corr` are shown in the default table if marginal proportions are specified.

Columns `pmarg1`, `pmarg2`, `corr`, `rrisk`, and `oratio` are available only if marginal proportions are specified.

Columns `diff`, `ratio`, `prdiscordant`, `sum`, `rrisk`, and `oratio` are shown in the default table if specified.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**. The `nfractional` option is allowed only for sample-size determination.

`prdiscordant(numlist)` specifies the proportion of discordant pairs or the sum of the discordant proportions,  $p_{12} + p_{21}$ . See *Alternative ways of specifying effect* for details about the specification of this option.

`sum(numlist)` is a synonym for `prdiscordant()`. See *Alternative ways of specifying effect* for details about the specification of this option.

`corr(numlist)` specifies the correlation between paired observations. This option is required if marginal proportions are specified.

`diff(numlist)` specifies the difference between the discordant proportions,  $p_{21} - p_{12}$ , or the marginal proportions,  $p_{+1} - p_{1+}$ . See *Alternative ways of specifying effect* for details about the specification of this option.

`ratio(numlist)` specifies the ratio of the discordant proportions,  $p_{21}/p_{12}$ , or the marginal proportions,  $p_{+1}/p_{1+}$ . See *Alternative ways of specifying effect* for details about the specification of this option.

`rrisk(numlist)` specifies the relative risk of the marginal proportions,  $p_{+1}/p_{1+}$ . See *Alternative ways of specifying effect* for details about the specification of this option.

`oratio(numlist)` specifies the odds ratio of the marginal proportions,  $\{p_{+1}(1-p_{1+})\}/\{p_{1+}(1-p_{+1})\}$ . See *Alternative ways of specifying effect* for details about the specification of this option.

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. *effect* is one of `diff` or `ratio` for discordant proportions and one of `diff`, `ratio`, `rrisk`, or `oratio` for marginal proportions. By default, the effect size `delta` is the difference between proportions. If `diff()`, `ratio()`, `rrisk()`, or `oratio()` is specified, the effect size `delta` will contain the effect corresponding to the specified option. For example, if `ratio()` is specified, `delta` will contain the ratio of the proportions. See *Alternative ways of specifying effect* for details about the specification of this option.

`direction()`, `onesided`, `parallel`; see [PSS] **power**.

#### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

#### Graph

`graph`, `graph()`; see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

#### Iteration

`init(#)` specifies the initial value for the estimated parameter. The estimated parameter is sample size for sample-size determination or the difference between the discordant proportions for the effect-size determination.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power pairedproportions` but is not shown in the dialog box:

`notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using `power pairedproportions`*

*Alternative ways of specifying effect*

*Effect specifications for discordant proportions*

*Effect specifications for marginal proportions*

*Computing sample size*

*Computing power*

*Computing effect size and target discordant proportions*

*Testing a hypothesis about two correlated proportions*

This entry describes the `power pairedproportions` command and the methodology for power and sample-size analysis for a two-sample paired-proportions test. See [\[PSS\] intro](#) for a general introduction to power and sample-size analysis and [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

The analysis of paired proportions is used to compare two dependent binomial populations. Dependent binomial data arise from matched case–control studies, where the cases are matched to the controls on the basis of similar demographic characteristics, or from longitudinal studies, where the same cases serve as their own controls over time or for different treatments. In all cases, each observation represents a pair of correlated binary outcomes.

There are many examples of studies where a researcher would like to compare two dependent proportions. For example, a state highway department might be concerned that tollbooth workers may experience hearing loss because of chronic exposure to traffic noise. It wants to test whether the proportion of workers with moderate to severe hearing loss is the same between a sample of workers exposed to traffic noise and a sample of workers sheltered by the quieter interior of the booth. Or a pediatrician might conduct a study to compare the proportions of males and females with a particular food allergy in a study of male/female fraternal twins.

This entry describes power and sample-size analysis for correlated binary outcomes in a two-way contingency table. Consider a  $2 \times 2$  table from a study where the outcome of interest is a pair of results from “occasion 1” and “occasion 2”, each either a “success” or a “failure”.

	Occasion 2		
Occasion 1	Success	Failure	Total
Success	$n_{11}$	$n_{12}$	$n_{1+}$
Failure	$n_{21}$	$n_{22}$	$n_{2+}$
Total	$n_{+1}$	$n_{+2}$	$n$

$n$  is the total number of pairs;  $n_{11}$  is the number of pairs for which the response is a success for both occasions;  $n_{12}$  is the number of *success–failure pairs* for which the response is a success on occasion 1 and a failure on occasion 2;  $n_{21}$  is the number of *failure–success pairs* for which the response is a failure on occasion 1 and a success on occasion 2; and  $n_{22}$  is the number of pairs for which the response is a failure for both occasions. The success–failure and failure–success pairs form *discordant pairs*, and the remaining pairs form *concordant pairs*.

The above table can also be expressed in terms of the proportions.

	Occasion 2		
Occasion 1	Success	Failure	Total
Success	$p_{11}$	$p_{12}$	$p_{1+}$
Failure	$p_{21}$	$p_{22}$	$1 - p_{1+}$
Total	$p_{+1}$	$1 - p_{+1}$	1

$p_{1+}$  is the success probability for occasion 1, and  $p_{+1}$  is the success probability for occasion 2. The marginal probabilities,  $p_{1+}$  and  $p_{+1}$ , are used to compare the outcomes between occasion 1 and occasion 2.

The null hypothesis for the test of equality of marginal proportions, also known as the test of *marginal homogeneity*, is  $H_0: p_{+1} = p_{1+}$ . The null hypothesis can be formulated in terms of the discordant probabilities, the failure–success probability,  $p_{21}$ , and the success–failure probability,  $p_{12}$ , using the relationships  $p_{+1} = p_{11} + p_{21}$  and  $p_{1+} = p_{11} + p_{12}$ . The considered null hypothesis is then  $H_0: p_{21} = p_{12}$  versus the two-sided alternative hypothesis  $H_a: p_{21} \neq p_{12}$ , the upper one-sided alternative  $H_a: p_{21} > p_{12}$ , or the lower one-sided alternative  $H_a: p_{21} < p_{12}$ . For a  $2 \times 2$  table, the test of marginal homogeneity is also called a *test of symmetry*.

A large-sample McNemar’s test is commonly used for testing the above hypotheses. Under the null hypothesis, the test statistic is distributed as a  $\chi_1^2$  distribution with 1 degree of freedom.

`power pairedproportions` provides power and sample-size analysis for McNemar’s test of two correlated proportions.

## Using power pairedproportions

`power pairedproportions` computes sample size, power, or target discordant proportions for a two-sample paired-proportions test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

For sample-size and power determinations, `power pairedproportions` provides a number of ways of specifying the magnitude of an effect desired to be detected by the test. Below we describe the use of the command, assuming that the desired effect is expressed by the values of the two discordant proportions; see *Alternative ways of specifying effect* for other specifications.

To compute sample size, you must specify the discordant proportions,  $p_{12}$  and  $p_{21}$ , and, optionally, the power of the test in option `power()`. The default power is set to 0.8.

To compute power, you must specify the sample size in option `n()` and the discordant proportions,  $p_{12}$  and  $p_{21}$ .

The effect-size determination is available only for discordant proportions. To compute effect size and target discordant proportions, you must specify the sample size in option `n()`, the power in option `power()`, the sum of the discordant proportions in option `prdiscordant()`, and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the failure–success proportion,  $p_{21}$ , is assumed to be larger than the specified success–failure proportion,  $p_{12}$ . You can change the direction to lower, which means that  $p_{21}$  is assumed to be smaller than  $p_{12}$ , by specifying the `direction(lower)` option.

There are multiple definitions of effect size for a two-sample paired-proportions test. The `effect()` option specifies what definition `power pairedproportions` should use when reporting the effect size, which is labeled as `delta` in the output of the `power` command.

When you specify the discordant proportions, the available definitions are the difference  $p_{21} - p_{12}$  between the discordant proportions, `effect(diff)`, or the ratio  $p_{21}/p_{12}$  of the discordant proportions, `effect(ratio)`.

When you specify the marginal proportions, the available definitions are the difference  $p_{+1} - p_{1+}$  between the marginal proportions, `effect(diff)`; the relative risk or ratio  $p_{+1}/p_{1+}$  of the marginal proportions, `effect(rrisk)` or `effect(ratio)`; or the odds ratio  $\{p_{+1}(1-p_{1+})\}/\{p_{1+}(1-p_{+1})\}$  of the marginal proportions, `effect(oratio)`.

When `effect()` is specified, the effect size `delta` in the output of the `power` command contains the estimate of the corresponding effect and is labeled accordingly. By default, `delta` corresponds to the difference between proportions. If any one of the options `diff()`, `ratio()`, `rrisk()`, or `oratio()` is specified and `effect()` is not specified, `delta` will contain the effect size corresponding to the specified option.

Some of `power pairedproportions`'s computations require iteration. For example, a sample size for a two-sided test is obtained by iteratively solving a nonlinear power equation. The default initial value for the sample size for the iteration procedure is obtained using a closed-form one-sided formula. If you desire, you may change it by specifying the `init()` option. See [\[PSS\] power](#) for the descriptions of other options that control the iteration procedure.

## Alternative ways of specifying effect

To compute power or sample size, you must also specify the magnitude of the effect that is desired to be detected by the test. You can do this by specifying either the discordant proportions,  $p_{12}$  and  $p_{21}$ ,

```
power pairedproportions  $p_{12}$   $p_{21}$ , ...
```

or the marginal proportions,  $p_{1+}$  and  $p_{+1}$ :

```
power pairedproportions  $p_{1+}$   $p_{+1}$ , corr(numlist) ...
```

When you specify marginal proportions, you must also specify the correlation between paired observations in option `corr()`.

Below we describe other alternative specifications separately for discordant proportions and marginal proportions.

## Effect specifications for discordant proportions

Instead of specifying  $p_{21}$ , you may specify the discordant proportion  $p_{12}$  as the argument to the command and the sum of the discordant proportions,  $p_{12} + p_{21}$ , in option `prdiscordant()` or option `sum()`,

```
power pairedproportions  $p_{12}$ , prdiscordant(numlist) ...
```

```
power pairedproportions  $p_{12}$ , sum(numlist) ...
```

the difference between the discordant proportions,  $p_{21} - p_{12}$ , in option `diff()`,

```
power pairedproportions  $p_{12}$ , diff(numlist) ...
```

or the ratio of the discordant proportions,  $p_{21}/p_{12}$ , in option `ratio()`:

```
power pairedproportions  $p_{12}$ , ratio(numlist) ...
```

You may omit both command arguments  $p_{12}$  and  $p_{21}$  altogether and specify options `prdiscordant()` or `sum()`, `diff()`, and `ratio()` in pairs.



For example, you can specify the sum  $p_{12} + p_{21}$  and difference  $p_{21} - p_{12}$  of the discordant proportions:

```
power pairedproportions , prdiscordant(numlist) diff(numlist) ...
```

Or you can specify the sum  $p_{12} + p_{21}$  and ratio  $p_{21}/p_{12}$  of the discordant proportions:

```
power pairedproportions , sum(numlist) ratio(numlist) ...
```

Or you can specify the difference  $p_{21} - p_{12}$  and ratio  $p_{21}/p_{12}$  of the discordant proportions:

```
power pairedproportions , diff(numlist) ratio(numlist) ...
```

When discordant proportions are specified, the effect size may be expressed as the difference between discordant proportions,  $p_{21} - p_{12}$ , or the ratio of discordant proportions,  $p_{21}/p_{12}$ . You may choose what effect to compute by specifying the `effect()` option.

By default, effect size is the difference between the discordant proportions. For example, for the specification below, the effect size  $\delta$  is the difference between the discordant proportions.

```
power pairedproportions p12 p21 , ...
```

The above specification is equivalent to

```
power pairedproportions p12 p21 , effect(diff) ...
```

Alternatively, you may request the effect size to be the ratio instead of the difference.

```
power pairedproportions p12 p21 , effect(ratio) ...
```

Likewise, if you specify the `ratio()` option, the effect size is the ratio of the proportions.

## Effect specifications for marginal proportions

Instead of specifying  $p_{+1}$ , you may specify the marginal proportion,  $p_{1+}$ , as the argument to the command and the difference between the marginal proportions,  $p_{+1} - p_{1+}$ , in option `diff()`,

```
power pairedproportions p1+ , corr(numlist) diff(numlist) ...
```

the ratio of the marginal proportions or relative risk,  $p_{+1}/p_{1+}$ , in option `ratio()` or option `rrisk()`,

```
power pairedproportions p1+ , corr(numlist) ratio(numlist) ...
```

```
power pairedproportions p1+ , corr(numlist) rrisk(numlist) ...
```

or the odds ratio,  $\{p_{+1}(1 - p_{1+})\}/\{p_{1+}(1 - p_{+1})\}$ , in option `oratio()`:

```
power pairedproportions p1+ , corr(numlist) oratio(numlist) ...
```

Alternatively, you may omit both command arguments  $p_{1+}$  and  $p_{+1}$  and specify one of the combinations of `diff()` and `rrisk()` or `ratio()`, or `oratio()` and `rrisk()` or `ratio()`. You may not combine `diff()` and `oratio()`, because marginal proportions cannot be identified uniquely from this combination.

For example, you can specify the difference  $p_{+1} - p_{1+}$  and ratio  $p_{+1}/p_{1+}$  of the marginal proportions,

```
power pairedproportions , corr(numlist) diff(numlist) ratio(numlist) ...
```

or the odds ratio,  $\{p_{+1}(1 - p_{1+})\}/\{p_{1+}(1 - p_{+1})\}$ , and the relative risk,  $p_{+1}/p_{1+}$ :

```
power pairedproportions , corr(numlist) oratio(numlist) rrisk(numlist) ...
```

The effect size for marginal proportions may be defined as one of the difference between marginal proportions,  $p_{+1} - p_{1+}$ , the odds ratio,  $\{p_{+1}(1 - p_{1+})\}/\{p_{1+}(1 - p_{+1})\}$ , or the relative risk or, equivalently, the ratio  $p_{+1}/p_{1+}$ . The `effect()` option for marginal proportions may contain one of `diff`, `oratio`, `rrisk`, or `ratio`.

By default, effect size is defined as the difference between the marginal proportions. For example, the following specification,

```
power pairedproportions p1+ p+1 , corr(numlist) ...
```

is equivalent to

```
power pairedproportions p1+ p+1 , corr(numlist) effect(diff) ...
```

You may request other measures of effect size for marginal proportions such as the risk ratio,

```
power pairedproportions p1+ p+1 , corr(numlist) effect(rrisk) ...
```

or odds ratio:

```
power pairedproportions p1+ p+1 , corr(numlist) effect(oratio) ...
```

In the following sections, we describe the use of `power pairedproportions` accompanied by examples for computing sample size, power, and target discordant proportion.

## Computing sample size

To compute sample size, you must specify the discordant proportions,  $p_{12}$  and  $p_{21}$ , and, optionally, the power of the test in option `power()`. The default power is set to 0.8. Instead of the discordant proportions, you can specify an effect of interest as shown in [Alternative ways of specifying effect](#).

### ► Example 1: Sample size for a two-sample paired-proportions test

Consider a study from [Agresti \(2013, 413\)](#) where the same group of subjects was asked who they voted for in the 2004 and 2008 presidential elections. In 2008, we witnessed a shift from the Republican President George W. Bush, who was finishing his second term, to Democratic President Barack Obama, who was beginning his first term. Suppose that we would like to conduct another survey for the 2012 and 2016 elections to see whether a similar shift—this time from a Democrat to a Republican—would occur in 2016, when President Obama finishes his second term. We are interested in testing the hypothesis whether the proportions of votes for a Democratic president in 2012 will be the same as in 2016.

Consider the following  $2 \times 2$  table:

2012 Election	2016 Election		Total
	Democratic	Republican	
Democratic	$p_{11}$	$p_{12}$	$p_{1+}$
Republican	$p_{21}$	$p_{22}$	$1 - p_{1+}$
Total	$p_{+1}$	$1 - p_{+1}$	1

The test of marginal homogeneity that the proportion of Democratic votes in 2012 will be the same in 2016 is given by the null hypothesis  $H_0: p_{1+} = p_{+1}$ . This is equivalent to testing whether the proportion  $p_{12}$  of voters who changed parties from Democratic to Republican is the same as the proportion  $p_{21}$  of voters who changed parties from Republican to Democratic between 2012 and 2016. The corresponding null hypothesis that tests these discordant proportions is  $H_0: p_{12} = p_{21}$ .

Suppose that the previous survey reported that the proportion of respondents who voted for a Democratic president in 2012 is  $p_{1+} = 0.53$ . Using this empirical evidence, a political expert believes that the odds of a candidate being elected as the president when his or her party has already served two consecutive terms as the president are low. According to the expert's opinion, the odds of the population voting Democrat in 2016 to the odds of the population voting Democrat in 2012 are 2:3; that is, the corresponding odds ratio is  $\theta = 2/3 = 0.667$ . Using the relationship between the odds ratio and marginal probabilities, we compute the marginal probability  $p_{+1}$  to be  $= [0.667 \times \{0.53/(1 - 0.53)\}]/(1 + [(0.667 \times \{0.53/(1 - 0.53)\})]) = 0.4293$ .

It is generally believed that voting behavior is positively correlated, which means that a person voting for one party in an election year is very likely to vote for the same party in the next election year. Suppose the expert posits a correlation of 0.8.

We wish to compute the sample size required for our survey to detect the difference between the considered marginal proportions. To compute the minimum sample size, we specify the marginal proportions after the command name and the correlation in option `corr()`:

```
. power pairedproportions 0.53 0.4293, corr(0.8)
Performing iteration ...
Estimated sample size for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p+1 = p1+ versus Ha: p+1 != p1+
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.1007 (difference)
    p1+ =    0.5300
    p+1 =    0.4293
    corr =    0.8000
Estimated sample size:
    N =          82
```

We find that 82 subjects are required in our survey for a 5%-level two-sided McNemar's test to detect a change in the proportion voting Democrat from 0.53 in 2012 to 0.4293 in 2016, which corresponds to the difference of  $\delta = 0.4293 - 0.53 = -0.1007$ , with 80% power.

◀

## ► Example 2: Reporting odds ratio

By default, as in [example 1](#), the effect size  $\delta$  is the difference between the marginal proportions. Alternatively, we can request that the effect size be reported as the odds ratio by specifying option `effect(oratio)`.

```
. power pairedproportions 0.53 0.4293, corr(0.8) effect(oratio)
Performing iteration ...
Estimated sample size for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p+1 = p1+ versus Ha: p+1 != p1+
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.6671 (odds ratio)
      p1+ = 0.5300
      p+1 = 0.4293
      corr = 0.8000
Estimated sample size:
      N = 82
```

The effect size `delta` now contains the odds ratio estimated from the specified marginal proportions. Also see [Alternative ways of specifying effect](#) for other available measures of effect.

◀

### ▶ Example 3: Specifying odds ratio

In [example 1](#), we computed the second marginal proportion,  $p_{+1}$ , using the postulated values of the first marginal proportion,  $p_{1+}$ , and the odds ratio,  $\theta$ , and we specified the two marginal proportions with `power pairedproportions`. We can instead specify the first marginal proportion and the odds ratio directly:

```
. power pairedproportions 0.53, corr(0.8) oratio(0.667)
Performing iteration ...
Estimated sample size for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p+1 = p1+ versus Ha: p+1 != p1+
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.6670 (odds ratio)
      p1+ = 0.5300
      p+1 = 0.4293
      corr = 0.8000
      odds ratio = 0.6670
Estimated sample size:
      N = 82
```

When the `oratio()` option is specified, the reported effect size `delta` corresponds to the odds ratio, and the value of the odds ratio specified in `oratio()` is also reported in the output.

Also see [Alternative ways of specifying effect](#) for other ways of specifying an effect.

◀

### ▷ Example 4: Specifying discordant proportions

Instead of marginal proportions, as in [example 1](#), we can specify discordant proportions. We compute discordant proportions using (1) and the estimates of marginal proportions and correlation in this study:  $p_{12} = 0.53(1 - 0.4293) - 0.8\sqrt{0.53 \times (1 - 0.53) \times 0.4293 \times (1 - 0.4293)} = 0.105$  and  $p_{21} = 0.105 + 0.4293 - 0.53 = 0.004$ .

```
. power pairedproportions 0.105 0.004
Performing iteration ...
Estimated sample size for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p21 = p12 versus Ha: p21 != p12
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.1010 (difference)
    p12 =    0.1050
    p21 =    0.0040
Estimated sample size:
    N =          82
```

We obtain the same sample size of 82 as in [example 1](#). The reported effect size `delta` corresponds to the difference of  $-0.1010$  between the discordant proportions.

Also see [Alternative ways of specifying effect](#) for other ways of specifying the effect desired to be detected by the test. ◀

## Computing power

To compute power, you must specify the sample size in option `n()` and the discordant proportions,  $p_{12}$  and  $p_{21}$ . Instead of the discordant proportions, you can specify an effect of interest as shown in [Alternative ways of specifying effect](#).

### ▷ Example 5: Power of a two-sample paired-proportions test

Continuing with [example 4](#), we will suppose that we anticipate to obtain a sample of 100 subjects and want to compute the power corresponding to this sample size.

In addition to the discordant proportions, we specify the sample size of 100 in option `n()`:

```
. power pairedproportions 0.105 0.004, n(100)
Estimated power for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p21 = p12 versus Ha: p21 != p12
Study parameters:
    alpha =    0.0500
    N =        100
    delta =   -0.1010 (difference)
    p12 =    0.1050
    p21 =    0.0040
Estimated power:
    power =    0.8759
```

As expected, with a larger sample size, this example achieves a larger power, about 88%, compared with [example 4](#). ◀

### ▷ Example 6: Alternative specification of an effect for power determination

Continuing with [example 5](#), we can specify the first discordant proportion,  $p_{12}$ , and one of the options `prdiscordant()`, `diff()`, or `ratio()` instead of specifying both discordant proportions.

For example, let's specify the sum of the discordant proportions instead of the discordant proportion  $p_{21}$ :

```
. power pairedproportions 0.105, n(100) prdiscordant(0.109)
Estimated power for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p21 = p12 versus Ha: p21 != p12
Study parameters:
    alpha =    0.0500
      N =      100
    delta =  -0.1010 (difference)
     p12 =    0.1050
     p21 =    0.0040
    p12 + p21 =  0.1090
Estimated power:
    power =    0.8759
```

We obtain results identical to those in [example 5](#).

◀

### ▷ Example 7: Power determination with marginal proportions

We can compute the power corresponding to the sample of 100 subjects for the specification using marginal proportions from [example 1](#) by additionally specifying option `n(100)`:

```
. power pairedproportions 0.53 0.4293, corr(0.8) n(100)
Estimated power for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p+1 = p1+ versus Ha: p+1 != p1+
Study parameters:
    alpha =    0.0500
      N =      100
    delta =  -0.1007 (difference)
     p1+ =    0.5300
     p+1 =    0.4293
     corr =    0.8000
Estimated power:
    power =    0.8739
```

As expected, the estimated power of 0.8739 is very close to the estimated power of 0.8759 in [example 5](#). If we had used input values for discordant proportions with more precision in [example 5](#), we would have obtained nearly identical results.

◀

### ▷ Example 8: Multiple values of study parameters

Continuing with [example 7](#), we would like to assess the effect of varying correlation on the power of our study.

We specify a range of correlations between 0.2 and 0.8 with a step size of 0.1 in option `corr()`:

```
. power pairedproportions 0.53 0.4293, corr(0.2(0.1)0.8) n(100)
Estimated power for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p+1 = p1+ versus Ha: p+1 != p1+
```

alpha	power	N	delta	pmarg1	pmarg2	corr
.05	.3509	100	-.1007	.53	.4293	.2
.05	.3913	100	-.1007	.53	.4293	.3
.05	.4429	100	-.1007	.53	.4293	.4
.05	.5105	100	-.1007	.53	.4293	.5
.05	.6008	100	-.1007	.53	.4293	.6
.05	.7223	100	-.1007	.53	.4293	.7
.05	.8739	100	-.1007	.53	.4293	.8

For a given sample size, the power increases as the correlation increases, which means that for a given power, the required sample size decreases as the correlation increases. This demonstrates that a paired design can improve the precision of the statistical inference compared with an independent design, which corresponds to a correlation of zero.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## Computing effect size and target discordant proportions

The effect-size determination is available only for discordant proportions. As we describe in detail in [Alternative ways of specifying effect](#), there are multiple definitions of the effect size for a paired-proportions test. The default is the difference between the failure–success proportion,  $p_{21}$ , and the success–failure proportion,  $p_{12}$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding discordant proportions that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and the sum of the discordant proportions must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $p_{21} > p_{12}$ , or lower, meaning  $p_{21} < p_{12}$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ► Example 9: Compute effect size and target proportions

Suppose that we want to compute the corresponding discordant proportions for given sample size and power. To compute the discordant proportions, we must specify the proportion of discordant pairs (the sum of the discordant proportions) in addition to sample size and power.

Continuing with [example 4](#), we will compute the corresponding effect size and target proportions for the sample size of 82 and the power of 0.8 using 0.109 for the proportion of discordant pairs. We also specify the `direction(lower)` option because  $p_{21} < p_{12}$  in our example.

```

. power pairedproportions, prdiscordant(.109) n(82) power(0.8) direction(lower)
Performing iteration ...
Estimated discordant proportions for a two-sample paired-proportions test
Large-sample McNemar's test
Ho: p21 = p12 versus Ha: p21 != p12; p21 < p12
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =         82
    p12 + p21 =    0.1090
Estimated effect size and discordant proportions:
      delta =  -0.1007 (difference)
      p12 =    0.1048
      p21 =    0.0042

```

The estimated discordant proportions of 0.1048 and 0.0042 are very close to the respective original estimates of the discordant proportions of 0.105 and 0.004 from [example 4](#).

◀

## Testing a hypothesis about two correlated proportions

Suppose we collected data from two paired binomial samples and wish to test whether the two proportions of an outcome of interest are the same. We wish to use McNemar's test to test this hypothesis. We can use the `mcc` command to perform McNemar's test; see [\[R\] epitab](#) for details.

### ▶ Example 10: Testing for paired proportions

We use data provided in table 11.1 of [Agresti \(2013, 414\)](#) that present the results of a General Social Survey, which asked males who they voted for in the 2004 and 2008 presidential elections.

2004 Election	2008 Election		Total
	Democrat	Republican	
Democrat	175	16	191
Republican	54	188	42
Total	229	204	433

We wish to test whether there was a change in the voting behavior of males in 2008 compared with 2004 using McNemar's test. We use `mcci`, the immediate form of `mcc`, to perform this test.



```
. mcci 175 16 54 188
```

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	175	16	191
Unexposed	54	188	242
Total	229	204	433

```
McNemar's chi2(1) = 20.63 Prob > chi2 = 0.0000
```

```
Exact McNemar significance probability = 0.0000
```

```
Proportion with factor
```

```
Cases .4411085
```

```
Controls .5288684 [95% Conf. Interval]
```

```
difference -.0877598 -.1270274 -.0484923
```

```
ratio .8340611 .7711619 .9020907
```

```
rel. diff. -.1862745 -.2738252 -.0987238
```

```
odds ratio .2962963 .1582882 .5254949 (exact)
```

McNemar's test statistic is 20.63 with the corresponding two-sided  $p$ -value less than  $10^{-4}$ , which provides strong evidence of a shift in the Democratic direction among male voters in 2008.

We use the estimates of this study to perform a sample-size analysis we would have conducted before a new study. The discordant proportions are  $p_{12} = 16/433 = 0.037$  and  $p_{21} = 54/433 = 0.125$ .

```
. power pairedproportions 0.037 0.125
```

```
Performing iteration ...
```

```
Estimated sample size for a two-sample paired-proportions test
```

```
Large-sample McNemar's test
```

```
Ho: p21 = p12 versus Ha: p21 != p12
```

```
Study parameters:
```

```
alpha = 0.0500
```

```
power = 0.8000
```

```
delta = 0.0880 (difference)
```

```
p12 = 0.0370
```

```
p21 = 0.1250
```

```
Estimated sample size:
```

```
N = 162
```

We find that we need a sample of 162 respondents to detect a difference of 0.0880 between discordant proportions of 0.037 and 0.125 with 80% power using a 5%-level two-sided test.

## Stored results

`power pairedproportions` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(p12)</code>	success–failure proportion (first discordant proportion)
<code>r(p21)</code>	failure–success proportion (second discordant proportion)
<code>r(pmargin1)</code>	success proportion for occasion 1 (first marginal proportion)
<code>r(pmargin2)</code>	success proportion for occasion 2 (second marginal proportion)
<code>r(corr)</code>	correlation between paired observations
<code>r(diff)</code>	difference between proportions
<code>r(ratio)</code>	ratio of proportions
<code>r(prdiscordant)</code>	proportion of discordant pairs
<code>r(sum)</code>	sum of discordant proportions
<code>r(rrisk)</code>	relative risk
<code>r(oratio)</code>	odds ratio
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or the difference between discordant proportions
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(effect)</code>	<code>diff</code> , <code>ratio</code> , <code>oratio</code> , or <code>rrisk</code>
<code>r(type)</code>	test
<code>r(method)</code>	<code>pairedproportions</code>
<code>r(direction)</code>	<code>upper</code> or <code>lower</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Consider a  $2 \times 2$  contingency table formed with  $n$  pairs of observations. The first subscript  $i = 1, 2$  denotes the success of failure in occasion 1, and the second subscript  $j = 1, 2$  denotes the success of failure in occasion 2.

	Occasion 2		
Occasion 1	Success	Failure	Total
Success	$p_{11}$	$p_{12}$	$p_{1+}$
Failure	$p_{21}$	$p_{22}$	$1 - p_{1+}$
Total	$p_{+1}$	$1 - p_{+1}$	1

Each element in the above table denotes the probability of observing a specific pair. For example,  $p_{11}$  is the probability of jointly observing a success on occasion 1 and occasion 2, and  $p_{12}$  is the probability of observing a success on occasion 1 and a failure on occasion 2.  $p_{1+}$  is the marginal probability of a success on occasion 1, and  $p_{+1}$  is the marginal probability of a success on occasion 2. The off-diagonal proportions  $p_{12}$  and  $p_{21}$  are referred to as *discordant proportions*. The relationship between the discordant proportions and the marginal proportions is given by

$$p_{12} = p_{1+}(1 - p_{+1}) - \rho\sqrt{p_{1+}(1 - p_{1+}) p_{+1}(1 - p_{+1})}$$

$$p_{21} = p_{12} + p_{+1} - p_{1+} \tag{1}$$

where  $\rho$  is the correlation between the paired observations.

A two-sample paired proportions test involves testing the null hypothesis  $H_0: p_{+1} = p_{1+}$  versus the two-sided alternative hypothesis  $H_a: p_{+1} \neq p_{1+}$ , the upper one-sided alternative  $H_a: p_{+1} > p_{1+}$ , or the lower one-sided alternative  $H_a: p_{+1} < p_{1+}$ . Using the relationship  $p_{1+} = p_{11} + p_{12}$  and  $p_{+1} = p_{11} + p_{21}$ , test hypotheses may be stated in terms of the discordant proportions, for example,  $H_0: p_{21} = p_{12}$  versus  $H_a: p_{21} \neq p_{12}$ .

McNemar’s test statistic is:

$$\chi^2 = (n_{12} - n_{21})^2 / (n_{12} + n_{21})$$

where  $n_{ij}$  is the number of successes ( $i = 1$ ) or failures ( $i = 2$ ) on occasion 1 and the number of successes ( $j = 1$ ) or failures ( $j = 2$ ) on occasion 2; see Lachin (2011, chap. 5) for details. This test statistic has an approximately  $\chi^2$  distribution with 1 degree of freedom under the null hypothesis. The square root of the  $\chi^2$  test statistic is approximately normal with zero mean and variance of one.

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha/k}$  and  $z_\beta$  be the  $(1 - \alpha/k)$ th and the  $\beta$ th quantiles of the standard normal distribution.

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi \left\{ \frac{p_{\text{diff}} \sqrt{n} - z_{1-\alpha} \sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}} \right\} & \text{for an upper one-sided test} \\ \Phi \left\{ \frac{-p_{\text{diff}} \sqrt{n} - z_{1-\alpha} \sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}} \right\} & \text{for a lower one-sided test} \\ \Phi \left\{ \frac{p_{\text{diff}} \sqrt{n} - z_{1-\alpha/2} \sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}} \right\} + \Phi \left\{ \frac{-p_{\text{diff}} \sqrt{n} - z_{1-\alpha/2} \sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}} \right\} & \text{for a two-sided test} \end{cases} \tag{2}$$

where  $\Phi(\cdot)$  is the cdf of a standard normal distribution,  $p_{\text{diff}} = p_{21} - p_{12}$ , and  $p_{\text{disc}} = p_{12} + p_{21}$ .

The sample size  $n$  for a one-sided test is computed using

$$n = \left( \frac{z_\alpha \sqrt{p_{\text{disc}}} + z_{1-\beta} \sqrt{p_{\text{disc}} - p_{\text{diff}}^2}}{p_{\text{diff}}} \right)^2 \tag{3}$$

See Connor (1987) for details.

For a two-sided test, sample size is computed iteratively from the two-sided power equation in (2). The default initial value is obtained from the corresponding one-sided formula (3) with the significance level  $\alpha/2$ .

The effect size  $\delta = p_{21} - p_{12}$  is computed iteratively from the corresponding power equation in (2). The default initial value is  $p_{\text{disc}}/2$ .

## References

- Agresti, A. 2013. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- Connor, R. J. 1987. Sample size for testing differences in proportions for the paired-sample design. *Biometrics* 43: 207–211.
- Lachin, J. M. 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **epitab** — Tables for epidemiologists

# Title

**power onevariance** — Power analysis for a one-sample variance test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">Reference</a>	<a href="#">Also see</a>		

## Description

**power onevariance** computes sample size, power, or target variance for a one-sample variance test. By default, it computes sample size for given power and the values of the variance parameters under the null and alternative hypotheses. Alternatively, it can compute power for given sample size and values of the null and alternative variances or the target variance for given sample size, power, and the null variance. The results can also be obtained for an equivalent standard deviation test, in which case standard deviations are used instead of variances. Also see [\[PSS\] power](#) for a general introduction to the **power** command using hypothesis tests.

## Quick start

Sample size for test of  $H_0: \sigma^2 = 4$  vs  $H_a: \sigma^2 \neq 4$  with null variance  $v_0 = 4$  and alternative variance  $v_a = 9$  with default power of 0.8 and significance level  $\alpha = 0.05$

```
power onevariance 4 9
```

As above, but for  $H_0: \sigma = 2$  vs  $H_a: \sigma \neq 2$  with null standard deviation  $s_0 = 2$  and alternative standard deviation  $s_a = 3$

```
power onevariance 2 3, sd
```

As above, but for  $\alpha = 0.1$

```
power onevariance 2 3, sd alpha(0.1)
```

Sample sizes for  $v_a$  equal to 7, 8, 9, 10, and 11

```
power onevariance 4 (7(1)11)
```

As above, but display results in a graph showing sample size versus alternative variance

```
power onevariance 4 (7(1)11), graph
```

Specify  $v_0$  and the ratio of variances

```
power onevariance 4, ratio(2.25)
```

Power for a sample size of 30

```
power onevariance 4 9, n(30)
```

As above, but specify standard deviations rather than variances

```
power onevariance 2 3, n(30) sd
```

As above, but specify a one-sided test

```
power onevariance 2 3, sd n(30) onesided
```

Effect size and target variance for  $v_0 = 4$  with a sample size of 20 and power of 0.8

```
power onevariance 4, n(20) power(.8)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

*Variance scale*

```
power onevariance  $v_0$   $v_a$  [ , power(numlist) options ]
```

*Standard deviation scale*

```
power onevariance  $s_0$   $s_a$  , sd [ power(numlist) options ]
```

*Compute power*

*Variance scale*

```
power onevariance  $v_0$   $v_a$  , n(numlist) [options ]
```

*Standard deviation scale*

```
power onevariance  $s_0$   $s_a$  , sd n(numlist) [options ]
```

*Compute effect size and target parameter*

*Target variance*

```
power onevariance  $v_0$  , n(numlist) power(numlist) [options ]
```

*Target standard deviation*

```
power onevariance  $s_0$  , sd n(numlist) power(numlist) [options ]
```

where  $v_0$  and  $s_0$  are the null (hypothesized) variance and standard deviation or the value of the variance and standard deviation under the null hypothesis, and  $v_a$  and  $s_a$  are the alternative (target) variance and standard deviation or the value of the variance and standard deviation under the alternative hypothesis. Each argument may be specified either as one number or as a list of values in parentheses (see [U] [11.1.8 numlist](#)).

<i>options</i>	Description
<code>sd</code>	request computation using the standard deviation scale; default is the variance scale
Main	
* <code>alpha(numlist)</code>	significance level; default is <code>alpha(0.05)</code>
* <code>power(numlist)</code>	power; default is <code>power(0.8)</code>
* <code>beta(numlist)</code>	probability of type II error; default is <code>beta(0.2)</code>
* <code>n(numlist)</code>	sample size; required to compute power or effect size
<code>nfractional</code>	allow fractional sample size
* <code>ratio(numlist)</code>	ratio of variances, $v_a/v_0$ (or ratio of standard deviations, $s_a/s_0$ , if option <code>sd</code> is specified); specify instead of the alternative variance $v_a$ (or standard deviation $s_a$ )
<code>direction(upper lower)</code>	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<code>onesided</code>	one-sided test; default is two sided
<code>parallel</code>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<code>[no]table[ (tablespec) ]</code>	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<code>saving(filename [ , replace ])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph[ (graphopts) ]</code>	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<code>init(#)</code>	initial value for sample size or variance
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] [11.1.8 numlist](#).

`sd` does not appear in the dialog box; specification of `sd` is done automatically by the dialog box selected.

`notitle` does not appear in the dialog box.

where *tablespec* is

```
column[:label] [column[:label] [...]] [ , tableopts]
```

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	number of subjects	$N$
<code>delta</code>	effect size	$\delta$
<code>v0</code>	null variance	$\sigma_0^2$
<code>va</code>	alternative variance	$\sigma_a^2$
<code>s0</code>	null standard deviation	$\sigma_0$
<code>sa</code>	alternative standard deviation	$\sigma_a$
<code>ratio</code>	ratio of the alternative variance to the null variance or ratio of the alternative standard deviation to the null standard deviation (if <code>sd</code> is specified)	$\sigma_a^2/\sigma_0^2$ $\sigma_a/\sigma_0$
<code>target</code>	target parameter; synonym for <code>va</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `s0` and `sa` are displayed in the default table in place of the `v0` and `va` columns when the `sd` option is specified.

Column `ratio` is shown in the default table if specified. If the `sd` option is specified, this column contains the ratio of standard deviations. Otherwise, this column contains the ratio of variances.

## Options

`sd` specifies that the computation be performed using the standard deviation scale. The default is to use the variance scale.

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**. The `nfractional` option is allowed only for sample-size determination.

`ratio(numlist)` specifies the ratio of the alternative variance to the null variance,  $v_a/v_0$ , or the ratio of standard deviations,  $s_a/s_0$ , if the `sd` option is specified. You can specify either the alternative variance  $v_a$  as a command argument or the ratio of the variances in `ratio()`. If you specify `ratio(#)`, the alternative variance is computed as  $v_a = v_0 \times \#$ . This option is not allowed with the effect-size determination.

`direction()`, `onesided`, `parallel`; see [PSS] **power**.

### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

### Graph

`graph`, `graph()`; see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.



Iteration

`init(#)` specifies an initial value for the iteration procedure. Iteration is used to compute variance for a two-sided test and to compute sample size. The default initial value for the sample size is obtained from a closed-form normal approximation. The default initial value for the variance is obtained from a closed-form solution for a one-sided test with the significance level of  $\alpha/2$ .

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power onevariance` but is not shown in the dialog box: `notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

- [Introduction](#)
- [Using power onevariance](#)
- [Computing sample size](#)
- [Computing power](#)
- [Computing effect size and target variance](#)
- [Performing a hypothesis test on variance](#)

This entry describes the `power onevariance` command and the methodology for power and sample-size analysis for a one-sample variance test. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

The study of variance arises in cases where investigators are interested in making an inference on the variability of a process. For example, the precision of a thermometer in taking accurate measurements, the variation in the weights of potato chips from one bag to another, the variation in mileage across automobiles of the same model. Before undertaking the actual study, we may want to find the optimal sample size to detect variations that exceed the tolerable limits or industry-wide standards.

This entry describes power and sample-size analysis for the inference about the population variance performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: \sigma^2 = \sigma_0^2$  versus the two-sided alternative hypothesis  $H_a: \sigma^2 \neq \sigma_0^2$ , the upper one-sided alternative  $H_a: \sigma^2 > \sigma_0^2$ , or the lower one-sided alternative  $H_a: \sigma^2 < \sigma_0^2$ .

Hypothesis testing of variances relies on the assumption of normality of the data. For a random sample of size  $n$  from a normal distribution, the distribution of the sample variance  $s^2$  is scaled chi-squared. The  $\chi^2$  test statistic  $(n-1)s^2/\sigma_0^2$ , which has a chi-squared distribution  $\chi_{n-1}^2$  with  $n-1$  degrees of freedom, is used to test hypotheses on variance, and the corresponding test is known as a  $\chi^2$  test.

The test of a variance is equivalent to the test of a standard deviation with the null hypothesis  $H_0: \sigma = \sigma_0$ . The standard deviation test uses the same  $\chi^2$  test statistic. The only difference between the two tests is the scale or metric of the variability parameter: variance for the variance test and standard deviation for the standard deviation test. In some cases, standard deviations may provide a more meaningful interpretation than variances. For example, standard deviations of test scores or IQ have the same scale as the mean and provide information about the spread of the observations around the mean.

The `power onevariance` command provides power and sample-size analysis for the  $\chi^2$  test of a one-sample variance or a one-sample standard deviation.

## Using power onevariance

`power onevariance` computes sample size, power, or target variance for a one-sample variance test. If the `sd` option is specified, `power onevariance` computes sample size, power, or target standard deviation for an equivalent one-sample standard-deviation test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

In what follows, we describe the use of `power onevariance` in a variance metric. The corresponding use in a standard deviation metric, when the `sd` option is specified, is the same except variances  $v_0$  and  $v_a$  should be replaced with the respective standard deviations  $s_0$  and  $s_a$ . Note that computations using the variance and standard deviation scales yield the same results; the difference is only in the specification of the parameters.

To compute sample size, you must specify the variances under the null and alternative hypotheses,  $v_0$  and  $v_a$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

To compute power, you must specify the sample size in the `n()` option and the variances under the null and alternative hypotheses as arguments  $v_0$  and  $v_a$ , respectively.

Instead of the null and alternative variances  $v_0$  and  $v_a$ , you can specify the null variance  $v_0$  and the ratio of the alternative variance to the null variance in the `ratio()` option.

To compute effect size, the ratio of the alternative to the null variances, and target variance, you must specify the sample size in the `n()` option, the power in the `power()` option, the null variance  $v_0$ , and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the target variance is assumed to be larger than the specified null value. You can change the direction to lower, which means that the target variance is assumed to be smaller than the specified null value, by specifying the `direction(lower)` option.

By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS] **unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

The test statistic for a one-sample variance test follows a  $\chi^2$  distribution. Its degrees of freedom depends on the sample size; therefore, sample-size computations require iteration. The effect-size determination for a two-sided test also requires iteration. The default initial value of the sample size is obtained using a closed-form normal approximation. The default initial value of the variance for the effect-size determination is obtained by using the corresponding computation for a one-sided test with the significance level  $\alpha/2$ . The default initial values may be changed by specifying the `init()` option. See [PSS] **power** for the descriptions of other options that control the iteration procedure.

In the following sections, we describe the use of `power onevariance` accompanied by examples for computing sample size, power, and target variance.

## Computing sample size

To compute sample size, you must specify the variances under the null and alternative hypotheses,  $v_0$  and  $v_a$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a one-sample variance test

Consider a study where interest lies in testing whether the variability in mileage (measured in miles per gallon) of automobiles of a certain car manufacturer equals a specified value. Industry-wide standards maintain that a variation of at most two miles per gallon (mpg) from an average value is acceptable for commercial production.

The process engineer suspects that a faulty assembly line has been producing the variation higher than the acceptable standard. He or she wishes to test the null hypothesis of  $H_0: \sigma = 2$  versus a two-sided alternative  $H_a: \sigma \neq 2$  or, equivalently,  $H_0: \sigma^2 = 4$  versus  $H_0: \sigma^2 \neq 4$ . The engineer wants to find the minimum number of cars so that the 5%-level two-sided test achieves the power of 80% to detect the alternative variance of 9 (or standard deviation of 3 mpg) given the null variance of 4 (or standard deviation of 2 mpg). To obtain the sample size, we specify the null and alternative values of the variance in  $v_0$  and  $v_a$  after the command name:

```
. power onevariance 4 9
Performing iteration ...
Estimated sample size for a one-sample variance test
Chi-squared test
Ho: v = v0 versus Ha: v != v0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    2.2500
    v0 =     4.0000
    va =     9.0000
Estimated sample size:
    N =          24
```

We find that a sample of 24 subjects is required for this study.

As we mentioned in the previous section, sample-size computation requires iteration. By default, `power onevariance` suppresses the iteration log, which may be displayed by specifying the `log` option.

### ▷ Example 2: Specifying ratio of variances

Instead of the alternative variance as in [example 1](#), we can specify the ratio of the alternative and null variances of  $9/4 = 2.25$  in the `ratio()` option:

```
. power onevariance 4, ratio(2.25)
Performing iteration ...
Estimated sample size for a one-sample variance test
Chi-squared test
Ho: v = v0 versus Ha: v != v0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    2.2500
      v0 =     4.0000
      va =     9.0000
      ratio =    2.2500
Estimated sample size:
      N =         24
```

We obtain the same results as in [example 1](#). The ratio of the variances is now also displayed in the output.

◀

### ▷ Example 3: Standard deviation test

We can use the `sd` option to perform calculations in the standard deviation metric. We reproduce results from [example 1](#) using the corresponding null and standard deviations of 2 and 3.

```
. power onevariance 2 3, sd
Performing iteration ...
Estimated sample size for a one-sample standard-deviation test
Chi-squared test
Ho: s = s0 versus Ha: s != s0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    1.5000
      s0 =     2.0000
      sa =     3.0000
Estimated sample size:
      N =         24
```

The results are the same, except the output reports standard deviations instead of variances.

◀

## Computing power

To compute power, you must specify the sample size in the `n()` option and the variances under the null and alternative hypotheses,  $v_0$  and  $v_a$ , respectively.

### ▷ Example 4: Power of a one-sample variance test

Continuing with [example 1](#), we will suppose that we are designing a new study and anticipate obtaining a sample of 30 cars. To compute the power corresponding to this sample size given the study parameters from [example 1](#), we specify the sample size of 30 in the `n()` option:

```
. power onevariance 4 9, n(30)
Estimated power for a one-sample variance test
Chi-squared test
Ho: v = v0 versus Ha: v != v0
Study parameters:
    alpha =    0.0500
      N   =     30
    delta =    2.2500
     v0   =    4.0000
     va   =    9.0000
Estimated power:
    power =    0.8827
```

With a larger sample size, the power of the test increases to about 88%.

◀

### ▷ Example 5: Multiple values of study parameters

Suppose we would like to assess the effect of increasing the alternative variance on the power of the test. We do this by specifying a range of values in parentheses in the argument for the alternative variance:

```
. power onevariance 4 (4.5(0.5)10), n(30)
Estimated power for a one-sample variance test
Chi-squared test
Ho: v = v0 versus Ha: v != v0
```

alpha	power	N	delta	v0	va
.05	.08402	30	1.125	4	4.5
.05	.1615	30	1.25	4	5
.05	.2694	30	1.375	4	5.5
.05	.391	30	1.5	4	6
.05	.511	30	1.625	4	6.5
.05	.6189	30	1.75	4	7
.05	.7098	30	1.875	4	7.5
.05	.7829	30	2	4	8
.05	.8397	30	2.125	4	8.5
.05	.8827	30	2.25	4	9
.05	.9147	30	2.375	4	9.5
.05	.9382	30	2.5	4	10

The power is an increasing function of the effect size, which is measured by the ratio of the alternative variance to the null variance.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

◀

## Computing effect size and target variance

Effect size  $\delta$  for a one-sample variance test is defined as the ratio of the alternative variance to the null variance  $\delta = v_a/v_0$  or the ratio of the alternative standard deviation to the null standard deviation  $\delta = s_a/s_0$  when the `sd` option is specified.

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, power, sample size, and the null variance or the null standard deviation must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $v_a > v_0$  ( $s_a > s_0$ ), or lower, meaning  $v_a < v_0$  ( $s_a < s_0$ ). The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ► Example 6: Minimum detectable value of the variance

Continuing with [example 4](#), we may also be interested to find the minimum effect size that can be detected with a power of 80% given a sample of 30 subjects. To compute the smallest effect size and the corresponding target variance, after the command name, we specify the null variance of 4, sample size `n(30)`, and power `power(0.8)`:

```
. power onevariance 4, n(30) power(0.8)
Performing iteration ...
Estimated target variance for a one-sample variance test
Chi-squared test
Ho: v = v0 versus Ha: v != v0; va > v0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      N      =     30
      v0     =    4.0000
Estimated effect size and target variance:
      delta =    2.0343
      va    =    8.1371
```

The smallest detectable value of the effect size, the ratio of the variances, is 2.03, which corresponds to the alternative variance of 8.14. Compared with [example 1](#), for the same power of 80%, this example shows a smaller variance with a larger sample of 30 subjects.

Above we assumed the effect to be in the upper direction. The effect size and target variance in the lower direction can be obtained by specifying `direction(lower)`.

```

. power onevariance 4, n(30) power(0.8) direction(lower)
Performing iteration ...
Estimated target variance for a one-sample variance test
Chi-squared test
Ho: v = v0 versus Ha: v != v0; va < v0
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      30
        v0 =    4.0000
Estimated effect size and target variance:
      delta =    0.4567
        va =    1.8267

```

The smallest detectable value of the effect size is 0.46, which corresponds to the alternative variance of 1.83.

◀

## Performing a hypothesis test on variance

In this section, we demonstrate the use of the `sdtest` command for testing hypotheses about variances; see [R] [sdtest](#) for details. Suppose we wish to test the hypothesis that the variance or standard deviation is different from a specific null value on the collected data. We can use the `sdtest` command to do this.

### ► Example 7: Testing for variance

We use `auto.dta` to demonstrate the use of `sdtest`. We have data on mileage ratings of 74 automobiles and wish to test whether the overall standard deviation is different from 3 miles per gallon (mpg).

```

. use http://www.stata-press.com/data/r14/auto
(1978 Automobile Data)
. sdtest mpg == 3
One-sample test of variance

```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
mpg	74	21.2973	.6725511	5.785503	19.9569	22.63769

```

      sd = sd(mpg)                                c = chi2 = 271.4955
Ho: sd = 3                                       degrees of freedom = 73
      Ha: sd < 3                                Ha: sd != 3                                Ha: sd > 3
Pr(C < c) = 1.0000                            2*Pr(C > c) = 0.0000                            Pr(C > c) = 0.0000

```

We find statistical evidence to reject the null hypothesis of  $H_0: \sigma_{\text{mpg}} = 3$  versus a two-sided alternative  $H_a: \sigma_{\text{mpg}} \neq 3$  at the 5% significance level; the  $p$ -value  $< 0.0000$ .

We use the estimates of this study to perform a sample-size analysis we would have conducted before the study.

```

. power onevar 3 5.78, sd
Performing iteration ...
Estimated sample size for a one-sample standard-deviation test
Chi-squared test
Ho: s = s0 versus Ha: s != s0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    1.9267
      s0     =    3.0000
      sa     =    5.7800

Estimated sample size:
      N =      10

```

We find that the sample size required to detect a standard deviation of 5.78 mpg given the null value of 3 mpg with 80% power using a 5%-level two-sided test is only 10.

◀

## Stored results

`power onevariance` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(v0)</code>	variance under the null hypothesis (for variance scale, default)
<code>r(va)</code>	variance under the alternative hypothesis (for variance scale, default)
<code>r(s0)</code>	standard deviation under the null hypothesis (if option <code>sd</code> is specified)
<code>r(sa)</code>	standard deviation under the alternative hypothesis (if option <code>sd</code> is specified)
<code>r(ratio)</code>	ratio of the alternative variance to the null variance (or the ratio of standard deviations if option <code>sd</code> is specified)
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or the variance
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	onevariance
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats
<code>r(scale)</code>	variance or standard deviation

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------



## Methods and formulas

Consider a random sample of size  $n$  from a normal population with mean  $\mu$  and variance  $\sigma^2$ . Let  $\sigma_0^2$  and  $\sigma_a^2$  denote the null and alternative values of the variance parameter, respectively.

A one-sample variance test involves testing the null hypothesis  $H_0: \sigma^2 = \sigma_0^2$  versus the two-sided alternative hypothesis  $H_a: \sigma^2 \neq \sigma_0^2$ , the upper one-sided alternative  $H_a: \sigma^2 > \sigma_0^2$ , or the lower one-sided alternative  $H_a: \sigma^2 < \sigma_0^2$ .

The sampling distribution of the test statistic  $\chi^2 = (n-1)s^2/\sigma^2$  under the null hypothesis follows a  $\chi^2$  distribution with  $n-1$  degrees of freedom, where  $s^2$  is the sample variance. The corresponding test is known as a  $\chi^2$  test.

The following formulas are based on [Dixon and Massey \(1983, 110–112\)](#).

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $\chi_{n-1,1-\alpha}^2$  and  $\chi_{n-1,\beta}^2$  be the  $(1-\alpha)$ th and the  $\beta$ th quantiles of the  $\chi^2$  distribution with  $n-1$  degrees of freedom.

The following equality holds at the critical value of the accept/reject boundary for  $H_0$ :

$$\frac{\chi_{n-1,1-\alpha}^2}{n-1} \sigma_0^2 = \frac{\chi_{n-1,\beta}^2}{n-1} \sigma_a^2$$

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} 1 - \chi_{n-1}^2 \left( \frac{\sigma_0^2}{\sigma_a^2} \chi_{n-1,1-\alpha}^2 \right) & \text{for an upper one-sided test} \\ \chi_{n-1}^2 \left( \frac{\sigma_0^2}{\sigma_a^2} \chi_{n-1,\alpha}^2 \right) & \text{for a lower one-sided test} \\ 1 - \chi_{n-1}^2 \left( \frac{\sigma_0^2}{\sigma_a^2} \chi_{n-1,1-\alpha/2}^2 \right) + \chi_{n-1}^2 \left( \frac{\sigma_0^2}{\sigma_a^2} \chi_{n-1,\alpha/2}^2 \right) & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $\chi_{n-1}^2(\cdot)$  is the cdf of a  $\chi^2$  distribution with  $n-1$  degrees of freedom.

Sample size  $n$  is obtained by iteratively solving the corresponding power equation from (1) for  $n$ . The default initial value for the sample size is obtained by using a large-sample normal approximation.

For a large  $n$ , the log-transformed sample variance is approximately normal with mean  $2 \ln(\sigma)$  and standard deviation  $\sqrt{2/n}$ . The approximate sample size is then given by

$$n = \frac{1}{2} \left\{ \frac{z_{1-\alpha/k} - z_\beta}{\ln \left( \frac{\sigma_a}{\sigma_0} \right)} \right\}^2$$

where  $k = 1$  for a one-sided test and  $k = 2$  for a two-sided test.

For a one-sided test, the minimum detectable value of the variance is computed as follows:

$$\sigma_a^2 = \begin{cases} \sigma_0^2 \frac{\chi_{n-1,1-\alpha}^2}{\chi_{n-1,\beta}^2} & \text{for an upper one-sided test} \\ \sigma_0^2 \frac{\chi_{n-1,\alpha}^2}{\chi_{n-1,1-\beta}^2} & \text{for a lower one-sided test} \end{cases} \quad (2)$$

For a two-sided test, the minimum detectable value of the variance is computed by iteratively solving the two-sided power equation from (1) for  $\sigma_a^2$ . The default initial value is obtained from (2) with  $\alpha$  replaced by  $\alpha/2$ .

If the `nfractional` option is not specified, the computed sample size is rounded up.

## Reference

Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw-Hill.

## Also see

[PSS] [power](#) — Power and sample-size analysis for hypothesis tests

[PSS] [power, graph](#) — Graph results from the power command

[PSS] [power, table](#) — Produce table of results from the power command

[PSS] [Glossary](#)

[R] [sdtest](#) — Variance-comparison tests

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power twovariances` computes sample size, power, or the experimental-group variance (or standard deviation) for a two-sample variances test. By default, it computes sample size for given power and the values of the control-group and experimental-group variances. Alternatively, it can compute power for given sample size and values of the control-group and experimental-group variances or the experimental-group variance for given sample size, power, and the control-group variance. Also see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \sigma_1^2 = \sigma_2^2$  versus  $H_a: \sigma_1^2 \neq \sigma_2^2$  with control-group variance  $v_1 = 25$ , experimental-group variance  $v_2 = 36$ , default power of 0.8, and significance level  $\alpha = 0.05$

```
power twovariances 25 36
```

Same as above, but specified as standard deviations  $s_1 = 5$  and  $s_2 = 6$

```
power twovariances 5 6, sd
```

Sample size for  $v_1 = 25$  and  $v_2$  equals to 36, 38, 40, and 42

```
power twovariances 25 (36(2)42)
```

As above, but display results in a graph of sample size versus  $v_2$

```
power twovariances 25 (36(2)42), graph
```

Save results to the dataset `mydata.dta`

```
power twovariances 25 (36(2)42), saving(mydata)
```

Power for a total sample size of 300

```
power twovariances 25 36, n(300)
```

As above, but specify sample sizes of 200 and 100 for groups 1 and 2, respectively

```
power twovariances 25 36, n1(200) n2(100)
```

Effect size and experimental-group standard deviation given control-group standard deviation of 5, sample size of 200, and power of 0.8

```
power twovariances 5, sd n(200) power(0.8)
```

As above, but calculate experimental-group variance given control-group variance of 25

```
power twovariances 25, n(200) power(0.8)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

*Variance scale*

```
power twovariances  $v_1$   $v_2$  [ , power(numlist) options ]
```

*Standard deviation scale*

```
power twovariances  $s_1$   $s_2$  , sd [ power(numlist) options ]
```

*Compute power*

*Variance scale*

```
power twovariances  $v_1$   $v_2$  , n(numlist) [ options ]
```

*Standard deviation scale*

```
power twovariances  $s_1$   $s_2$  , sd n(numlist) [ options ]
```

*Compute effect size and target parameter*

*Experimental-group variance*

```
power twovariances  $v_1$  , n(numlist) power(numlist) [ options ]
```

*Experimental-group standard deviation*

```
power twovariances  $s_1$  , sd n(numlist) power(numlist) [ options ]
```

where  $v_1$  and  $s_1$  are the variance and standard deviation, respectively, of the control (reference) group, and  $v_2$  and  $s_2$  are the variance and standard deviation of the experimental (comparison) group. Each argument may be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 **numlist**).

<i>options</i>	Description
<code>sd</code>	request computation using the standard deviation scale; default is the variance scale
Main	
* <code>alpha(numlist)</code>	significance level; default is <code>alpha(0.05)</code>
* <code>power(numlist)</code>	power; default is <code>power(0.8)</code>
* <code>beta(numlist)</code>	probability of type II error; default is <code>beta(0.2)</code>
* <code>n(numlist)</code>	total sample size; required to compute power or effect size
* <code>n1(numlist)</code>	sample size of the control group
* <code>n2(numlist)</code>	sample size of the experimental group
* <code>nratio(numlist)</code>	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<code>compute(n1   n2)</code>	solve for $N1$ given $N2$ or for $N2$ given $N1$
<code>nfractional</code>	allow fractional sample sizes
* <code>ratio(numlist)</code>	ratio of variances, $v_2/v_1$ (or ratio of standard deviations, $s_2/s_1$ , if option <code>sd</code> is specified); specify instead of the experimental-group variance $v_2$ (or standard deviation $s_2$ )
<code>direction(upper lower)</code>	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<code>onesided</code>	one-sided test; default is two sided
<code>parallel</code>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<code>[no]table[ (tablespec) ]</code>	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<code>saving(filename [ , replace ])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph[ (graphopts) ]</code>	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<code>init(#)</code>	initial value for the sample sizes or experimental-group variance
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] [11.1.8 numlist](#).

`sd` does not appear in the dialog box; specification of `sd` is done automatically by the dialog box selected.

`notitle` does not appear in the dialog box.

where *tablespec* is

```
column[:label] [column[:label] [...]] [, tableopts]
```

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N1</code>	number of subjects in the control group	$N_1$
<code>N2</code>	number of subjects in the experimental group	$N_2$
<code>nratio</code>	ratio of sample sizes, experimental to control	$N_2/N_1$
<code>delta</code>	effect size	$\delta$
<code>v1</code>	control-group variance	$\sigma_1^2$
<code>v2</code>	experimental-group variance	$\sigma_2^2$
<code>s1</code>	control-group standard deviation	$\sigma_1$
<code>s2</code>	experimental-group standard deviation	$\sigma_2$
<code>ratio</code>	ratio of the experimental-group variance to the control-group variance or ratio of the experimental-group standard deviation to the control-group standard deviation (if <code>sd</code> is specified)	$\sigma_2^2/\sigma_1^2$ $\sigma_2/\sigma_1$
<code>target</code>	target parameter; synonym for <code>v2</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `s1` and `s2` are displayed in the default table in place of the `v1` and `v2` columns when the `sd` option is specified.

Column `ratio` is shown in the default table if specified. If the `sd` option is specified, this column contains the ratio of standard deviations. Otherwise, this column contains the ratio of variances.

## Options

`sd` specifies that the computation be performed using the standard deviation scale. The default is to use the variance scale.

Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `compute()`, `nfractional`; see [PSS] [power](#).

`ratio(numlist)` specifies the ratio of the experimental-group variance to the control-group variance,  $v_2/v_1$ , or the ratio of the standard deviations,  $s_2/s_1$ , if the `sd` option is specified. You can specify either the experimental-group variance  $v_2$  as a command argument or the ratio of the variances in `ratio()`. If you specify `ratio(#)`, the experimental-group variance is computed as  $v_2 = v_1 \times \#$ . This option is not allowed with the effect-size determination.

`direction()`, `onesided`, `parallel`; see [PSS] [power](#).

#### Table

`table`, `table()`, `notable`; see [PSS] [power](#), [table](#).

`saving()`; see [PSS] [power](#).

#### Graph

`graph`, `graph()`; see [PSS] [power](#), [graph](#). Also see the *column* table for a list of symbols used by the graphs.

#### Iteration

`init(#)` specifies the initial value for the estimated parameter. For sample-size determination, the estimated parameter is either the control-group size  $n_1$  or, if `compute(n2)` is specified, the experimental-group size  $n_2$ . For the effect-size determination, the estimated parameter is the experimental-group variance  $v_2$  or, if the `sd` option is specified, the experimental-group standard deviation  $s_2$ . The default initial values for the variance and standard deviation for a two-sided test are obtained as a closed-form solution for the corresponding one-sided test with the significance level  $\alpha/2$ . The default initial values for sample sizes for a  $\chi^2$  test are obtained from the corresponding closed-form normal approximation.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power twovariances` but is not shown in the dialog box:

`notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power twovariances*

*Computing sample size*

*Computing power*

*Computing effect size and experimental-group variance*

*Testing a hypothesis about two independent variances*

This entry describes the `power twovariances` command and the methodology for power and sample-size analysis for a two-sample variances test. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

Investigators are often interested in comparing the variances of two populations, such as comparing variances in yields of corn from two plots, comparing variances of stock returns from two companies, comparing variances of the alcohol concentrations from two different yeast strains, and so on. Before conducting the actual study, the investigators need to find the optimal sample size to detect variations that are beyond tolerable limits or industry-wide standards.

This entry describes power and sample-size analysis for the inference about two population variances performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: \sigma_2^2 = \sigma_1^2$  versus the two-sided alternative hypothesis  $H_a: \sigma_2^2 \neq \sigma_1^2$ , the upper one-sided alternative  $H_a: \sigma_2^2 > \sigma_1^2$ , or the lower one-sided alternative  $H_a: \sigma_2^2 < \sigma_1^2$ .

Hypothesis testing of variances relies on the assumption of normality. If two independent processes are assumed to follow a normal distribution, then the ratio of their sample variances follows an  $F$  distribution, and the corresponding test is known as an  $F$  test.

The test of variances is equivalent to the test of standard deviations with the null hypothesis  $H_0: \sigma_1 = \sigma_2$ . The standard deviation test uses the same  $F$  test statistic. The only difference between the two tests is the scale or metric of the variability parameters: variances for the variance test and standard deviations for the standard deviation test. In some cases, standard deviations may provide a more meaningful interpretation than variances. For example, standard deviations of test scores or IQ have the same scale as the mean and provide information about the spread of the observations around the mean.

The `power twovariances` command provides power and sample-size analysis for the  $F$  test of two-sample variances or standard deviations.

## Using power twovariances

`power twovariances` computes sample size, power, or experimental-group variance for a two-sample variances test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced- or equal-allocation design; see [PSS] **unbalanced designs** for a description of how to specify an unbalanced design.

In what follows, we describe the use of `power twovariances` in a variance metric. The corresponding use in a standard deviation metric, when the `sd` option is specified, is the same except variances  $v_1$  and  $v_2$  should be replaced with the respective standard deviations  $s_1$  and  $s_2$ . Note that computations using the variance and standard deviation scales yield the same results; the difference is only in the specification of the parameters.

To compute the total sample size, you must specify the control- and experimental-group variances,  $v_1$  and  $v_2$ , respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

Instead of the total sample size, you can compute one of the group sizes given the other one. To compute the control-group sample size, you must specify the `compute(n1)` option and the sample size of the experimental group in the `n2()` option. Likewise, to compute the experimental-group sample size, you must specify the `compute(n2)` option and the sample size of the control group in the `n1()` option.

To compute power, you must specify the total sample size in the `n()` option and the control and the experimental-group variances,  $v_1$  and  $v_2$ , respectively.

Instead of the experimental-group variance  $v_2$ , you may specify the ratio  $v_2/v_1$  of the experimental-group variance to the control-group variance in the `ratio()` option when computing sample size or power.

To compute effect size, the ratio of the experimental-group variance to the control-group variance, and the experimental-group variance, you must specify the total sample size in the `n()` option, the power in the `power()` option, the control-group variance  $v_1$ , and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the experimental-group



variance is assumed to be larger than the specified control-group value. You can change the direction to be lower, which means that the experimental-group variance is assumed to be smaller than the specified control-group value, by specifying the `direction(lower)` option.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see *Two samples* in [PSS] **unbalanced designs** for more details.

In the following sections, we describe the use of `power twovariances` accompanied by examples for computing sample size, power, and experimental-group variance.

## Computing sample size

To compute sample size, you must specify the control- and the experimental-group variances,  $v_1$  and  $v_2$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-sample variances test

Consider a study whose goal is to investigate whether the variability in weights (measured in ounces) of bags of potato chips produced by a machine at a plant A, the control group, differs from that produced by a similar machine at a new plant B, the experimental group. The considered null hypothesis is  $H_0: \sigma_A = \sigma_B$  versus a two-sided alternative hypothesis  $H_a: \sigma_A \neq \sigma_B$  or, equivalently,  $H_0: \sigma_A^2 = \sigma_B^2$  versus  $H_a: \sigma_A^2 \neq \sigma_B^2$ . The standard deviation of weights from plant A is 2 ounces. The standard deviation of weights from the new plant B is expected to be lower, 1.5 ounces. The respective variances of weights from plants A and B are 4 and 2.25. Investigators wish to obtain the minimum sample size required to detect the specified change in variability with 80% power using a 5%-level two-sided test assuming equal-group allocation. To compute sample size for this study, we specify the control- and experimental-group variances after the command name:

```
. power twovariances 4 2.25
Performing iteration ...
Estimated sample sizes for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.5625
      v1     =    4.0000
      v2     =    2.2500
Estimated sample sizes:
      N =      194
      N per group =    97
```

A total sample of 194 bags, 97 in each plant, must be obtained to detect the specified ratio of variances in the two plants with 80% power using a two-sided 5%-level test.

### ▷ Example 2: Standard deviation scale

We can also specify standard deviations instead of variances, in which case we must also specify the `sd` option:

```
. power twovariances 2 1.5, sd
Performing iteration ...
Estimated sample sizes for a two-sample standard-deviations test
F test
Ho: s2 = s1 versus Ha: s2 != s1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7500
     s1 =    2.0000
     s2 =    1.5000
Estimated sample sizes:
      N =      194
N per group =      97
```

We obtain the same sample sizes as in [example 1](#).

◀

### ▷ Example 3: Specifying ratio of variances or standard deviations

Instead of the experimental-group variance of 2.25 as in [example 1](#), we can specify the ratio of variances  $2.25/4 = 0.5625$  in the `ratio()` option.

```
. power twovariances 4, ratio(0.5625)
Performing iteration ...
Estimated sample sizes for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.5625
     v1 =    4.0000
     v2 =    2.2500
    ratio =    0.5625
Estimated sample sizes:
      N =      194
N per group =      97
```

The results are identical to those from [example 1](#).

Similarly, instead of the experimental-group standard deviation of 1.5 as in [example 2](#), we can specify the ratio of standard deviations  $1.5/2 = 0.75$  in the `ratio()` option and obtain the same results:

```

. power twovariances 2, sd ratio(0.75)
Performing iteration ...
Estimated sample sizes for a two-sample standard-deviations test
F test
Ho: s2 = s1 versus Ha: s2 != s1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.7500
    s1 = 2.0000
    s2 = 1.5000
    ratio = 0.7500
Estimated sample sizes:
    N = 194
    N per group = 97

```



#### ► Example 4: Computing one of the group sizes

Continuing with [example 1](#), we will suppose that investigators anticipate a sample of 100 bags from plant A and wish to compute the required number of bags from plant B. To compute sample size for plant B using the study parameters of [example 1](#), we use a combination of the `n1()` and `compute(n2)` options.

```

. power twovariances 4 2.25, n1(100) compute(n2)
Performing iteration ...
Estimated sample sizes for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.5625
    v1 = 4.0000
    v2 = 2.2500
    N1 = 100
Estimated sample sizes:
    N = 194
    N2 = 94

```

A slightly smaller sample of 94 bags is needed from plant B given a slightly larger sample of bags from plant A to achieve the same 80% power as in [example 1](#).

If the sample size for plant B is known a priori, you can compute the sample size for plant A by specifying the `n2()` and `compute(n1)` options.



#### ► Example 5: Unbalanced design

By default, `power twovariances` computes sample size for a balanced- or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size for an unbalanced design by specifying the `nratio()` option.

Continuing with [example 1](#), we will suppose that the new plant B is more efficient and more cost effective in producing bags of chips than plant A. Investigators anticipate twice as many bags from plant B than from plant A; that is,  $n_2/n_1 = 2$ . We compute the required sample size for this unbalanced design by specifying the `nratio()` option:

```
. power twovariances 4 2.25, nratio(2)
Performing iteration ...
Estimated sample sizes for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.5625
       v1 =    4.0000
       v2 =    2.2500
      N2/N1 =    2.0000
Estimated sample sizes:
      N =      225
      N1 =      75
      N2 =     150
```

The requirement on the total sample size increases for an unbalanced design compared with the balanced design from [example 1](#). Investigators must decide whether the decrease of 22 from 97 to 75 in the number of bags from plant A covers the cost of the additional 53 ( $150 - 97 = 53$ ) bags from plant B.

Also see [Two samples](#) in [\[PSS\] unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.

◀

## Computing power

To compute power, you must specify the total sample size in the `n()` option and the control- and experimental-group variances,  $v_1$  and  $v_2$ , respectively.

### ► Example 6: Power of a two-sample variances test

Continuing with [example 1](#), we will suppose that the investigators can afford a total sample of 250 bags, 125 from each plant, and want to find the power corresponding to this larger sample size.

To compute the power corresponding to this sample size, we specify the total sample size in the `n()` option:

```
. power twovariances 4 2.25, n(250)
Estimated power for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1
Study parameters:
      alpha =    0.0500
       N =      250
N per group =    125
      delta =    0.5625
       v1 =    4.0000
       v2 =    2.2500
Estimated power:
      power =    0.8908
```

With a total sample of 250 bags, 125 per plant, we obtain a power of roughly 89%.

◀

### ► Example 7: Multiple values of study parameters

In this example, we assess the effect of varying the variances of weights obtained from plant B on the power of our study. Continuing with [example 6](#), we vary the experimental-group variance from 1.5 to 3 in 0.25 increments. We specify the corresponding *numlist* in parentheses:

```
. power twovariances 4 (1.5(0.25)3), n(250)
Estimated power for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1
```

alpha	power	N	N1	N2	delta	v1	v2
.05	.9997	250	125	125	.375	4	1.5
.05	.9956	250	125	125	.4375	4	1.75
.05	.9701	250	125	125	.5	4	2
.05	.8908	250	125	125	.5625	4	2.25
.05	.741	250	125	125	.625	4	2.5
.05	.5466	250	125	125	.6875	4	2.75
.05	.3572	250	125	125	.75	4	3

The power decreases from 99.97% to 35.72% as the experimental-group variance gets closer to the control-group variance of 4.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

◀

## Computing effect size and experimental-group variance

Effect size  $\delta$  for a two-sample variances test is defined as the ratio of the experimental-group variance to the control-group variance,  $\delta = v_2/v_1$ . If the *sd* option is specified, effect size  $\delta$  is the ratio of the experimental-group standard deviation to the control-group standard deviation,  $\delta = s_2/s_1$ .

Sometimes, we may be interested in determining the smallest effect and the corresponding experimental-group variance that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and control-group variance must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $v_2 > v_1$ , or lower, meaning  $v_2 < v_1$ . The direction may be specified in the *direction()* option; *direction(upper)* is the default. If the *sd* option is specified, the estimated parameter is the experimental-group standard deviation instead of the variance.

### ► Example 8: Minimum detectable value of the experimental-group variance

Continuing with [example 6](#), we will compute the minimum plant B variance that can be detected given a total sample of 250 bags and 80% power. To find the variance, after the command name, we specify the plant A variance of 4, total sample size *n*(250), and power *power*(0.8):

```

. power twovariances 4, n(250) power(0.8)
Performing iteration ...
Estimated experimental-group variance for a two-sample variances test
F test
Ho: v2 = v1 versus Ha: v2 != v1; v2 > v1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      250
  N per group =   125
         v1 =    4.0000

Estimated effect size and experimental-group variance:
      delta =    1.6573
         v2 =    6.6291

```

We find that the minimum value of the experimental-group variance that would yield a statistically significant result in this study is 6.63, and the corresponding effect size is 1.66.

In this example, we computed the variance assuming an upper direction, or a ratio greater than 1,  $\delta > 1$ . To request a lower direction, or a ratio less than 1, we can specify the `direction(lower)` option.

◀

## Testing a hypothesis about two independent variances

In this section, we demonstrate the use of the `sdtest` command for testing a hypothesis about two independent variances; see [R] [sdtest](#) for details.

### ▶ Example 9: Comparing two variances

Consider the fuel dataset analyzed in [R] [sdtest](#). Suppose we want to investigate the effectiveness of a new fuel additive on the mileage of cars. We have a sample of 12 cars, where each car was run without the additive and later with the additive. The results of each run are stored in variables `mpg1` and `mpg2`.

```

. use http://www.stata-press.com/data/r14/fuel
. sdtest mpg1==mpg2
Variance ratio test

```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
mpg1	12	21	.7881701	2.730301	19.26525	22.73475
mpg2	12	22.75	.9384465	3.250874	20.68449	24.81551
combined	24	21.875	.6264476	3.068954	20.57909	23.17091

```

      ratio = sd(mpg1) / sd(mpg2)                                f =    0.7054
Ho: ratio = 1                                                    degrees of freedom = 11, 11
      Ha: ratio < 1                Ha: ratio != 1                Ha: ratio > 1
Pr(F < f) = 0.2862                2*Pr(F < f) = 0.5725                Pr(F > f) = 0.7138

```

`sdtest` uses the ratio of the control-group standard deviation to the experimental-group standard deviation as its test statistic. We do not have sufficient evidence to reject the null hypothesis of  $H_0: \sigma_1 = \sigma_2$  versus the two-sided alternative  $H_a: \sigma_1 \neq \sigma_2$ ; the  $p$ -value  $> 0.5725$ .

We use the estimates of this study to perform a sample-size analysis we would have conducted before the study. We assume an equal-group design and power of 80%.

```
. power twovariances 2.73 3.25, sd power(0.8)
Performing iteration ...
Estimated sample sizes for a two-sample standard-deviations test
F test
Ho: s2 = s1 versus Ha: s2 != s1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    1.1905
      s1 =     2.7300
      s2 =     3.2500
Estimated sample sizes:
      N =      522
      N per group = 261
```

The total sample size required by the test to detect the difference between the two standard deviations of 2.73 in the control group and of 3.25 in the experimental group is 522, 261 for each group, which is significantly larger than the sample of 12 cars in our fuel dataset.

◀

## Stored results

power twovariances stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(v1)</code>	control-group variance
<code>r(v2)</code>	experimental-group variance
<code>r(ratio)</code>	ratio of the experimental- to the control-group variances (or standard deviations if <code>sd</code> is specified)
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for sample sizes, experimental-group variance, or standard deviation
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

Macros	
r(type)	test
r(method)	twovariances
r(scale)	variance or standard deviation
r(direction)	upper or lower
r(columns)	displayed table columns
r(labels)	table column labels
r(widths)	table column widths
r(formats)	table column formats
Matrix	
r(pss_table)	table of results

## Methods and formulas

Consider two independent samples from a normal population with means  $\mu_1$  and  $\mu_2$  and variances  $\sigma_1^2$  and  $\sigma_2^2$ . The ratio  $(s_1/\sigma_1)^2/(s_2/\sigma_2)^2$  has an  $F$  distribution with  $n_1 - 1$  numerator and  $n_2 - 1$  denominator degrees of freedom.  $s_1^2$  and  $s_2^2$  are the sample variances, and  $n_1$  and  $n_2$  are the sample sizes.

Let  $\sigma_1^2$  and  $\sigma_2^2$  be the control-group and experimental-group variances, respectively.

A two-sample variances test involves testing the null hypothesis  $H_0: \sigma_2^2 = \sigma_1^2$  versus the two-sided alternative hypothesis  $H_a: \sigma_2^2 \neq \sigma_1^2$ , the upper one-sided alternative  $H_a: \sigma_2^2 > \sigma_1^2$ , or the lower one-sided alternative  $H_a: \sigma_2^2 < \sigma_1^2$ .

Equivalently, the hypotheses can be expressed in terms of the ratio of the two variances:  $H_0: \sigma_2^2/\sigma_1^2 = 1$  versus the two-sided alternative  $H_a: \sigma_2^2/\sigma_1^2 \neq 1$ , the upper one-sided alternative  $H_a: \sigma_2^2/\sigma_1^2 > 1$ , or the lower one-sided alternative  $H_a: \sigma_2^2/\sigma_1^2 < 1$ .

The following formulas are based on [Dixon and Massey \(1983, 116–119\)](#).

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $F_\alpha = F_{n_1-1, n_2-1, \alpha}$  and  $F_{n_1-1, n_2-1, \beta}$  be the  $\alpha$ th and the  $\beta$ th quantiles of an  $F$  distribution with  $n_1 - 1$  numerator and  $n_2 - 1$  denominator degrees of freedom.

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} 1 - F_{n_1-1, n_2-1} \left( \frac{\sigma_1^2}{\sigma_2^2} F_{1-\alpha} \right) & \text{for an upper one-sided test} \\ F_{n_1-1, n_2-1} \left( \frac{\sigma_1^2}{\sigma_2^2} F_\alpha \right) & \text{for a lower one-sided test} \\ 1 - F_{n_1-1, n_2-1} \left( \frac{\sigma_1^2}{\sigma_2^2} F_{1-\alpha/2} \right) + F_{n_1-1, n_2-1} \left( \frac{\sigma_1^2}{\sigma_2^2} F_{\alpha/2} \right) & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $F_{n_1-1, n_2-1}(\cdot)$  is the cdf of an  $F$  distribution with  $n_1 - 1$  numerator and  $n_2 - 1$  denominator degrees of freedom.

Let  $R = n_2/n_1$  denote the allocation ratio. Then  $n_2 = R \times n_1$  and power can be viewed as a function of  $n_1$ . Therefore, for sample-size determination, the control-group sample size  $n_1$  is computed first. The experimental-group size  $n_2$  is then computed as  $R \times n_1$ , and the total sample size is computed as  $n = n_1 + n_2$ . By default, sample sizes are rounded to integer values; see [Fractional sample sizes](#) in [\[PSS\] unbalanced designs](#) for details.

If either  $n_1$  or  $n_2$  is known, the other sample size is computed iteratively from the corresponding power equation in (1).

The initial values for the sample sizes are obtained from closed-form large-sample normal approximations; see, for example, [Mathews \(2010, 68\)](#).



For a one-sided test, the minimum detectable value of the experimental-group variance is computed as follows:

$$\sigma_2^2 = \begin{cases} \sigma_1^2 \frac{F_{n_1-1, n_2-1, 1-\alpha}}{F_{n_1-1, n_2-1, \beta}} & \text{for an upper one-sided test} \\ \sigma_1^2 \frac{F_{n_1-1, n_2-1, \alpha}}{F_{n_1-1, n_2-1, 1-\beta}} & \text{for a lower one-sided test} \end{cases} \quad (2)$$

For a two-sided test, the minimum detectable value of the experimental-group variance is computed iteratively using the two-sided power equation from (1). The default initial value is obtained from (2) with  $\alpha$  replaced by  $\alpha/2$ .

## References

- Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw-Hill.
- Mathews, P. 2010. *Sample Size Calculations: Practical Methods for Engineers and Scientists*. Fairport Harbor, OH: Mathews Malnar and Bailey.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **sdtest** — Variance-comparison tests

# Title

**power onecorrelation** — Power analysis for a one-sample correlation test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power onecorrelation` computes sample size, power, or target correlation for a one-sample correlation test. By default, it computes sample size for given power and the values of the correlation parameters under the null and alternative hypotheses. Alternatively, it can compute power for given sample size and values of the null and alternative correlations or the target correlation for given sample size, power, and the null correlation. Also see [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \rho = 0$  versus  $H_a: \rho \neq 0$  with null correlation  $r_0 = 0$ , alternative correlation  $r_a = 0.3$ , and default power of 0.8 and significance level  $\alpha = 0.05$

```
power onecorrelation 0 .3
```

As above, for alternative correlations of 0.2, 0.25, 0.3, 0.35, and 0.4

```
power onecorrelation 0 (.2(.05).4)
```

As above, but display a graph showing sample size versus alternative correlation

```
power onecorrelation 0 (.2(.05).4), graph
```

Sample size for  $r_0 = 0.1$ ,  $r_a = 0.3$  with power of 0.85 and  $\alpha = 0.01$

```
power onecorrelation .1 .3, power(.85) alpha(.01)
```

Power for a sample size of 30

```
power onecorrelation 0 .5, n(30)
```

Effect size and target correlation for a sample size of 20 and power of 0.8

```
power onecorrelation 0, n(20) power(.8)
```

As above, for sample sizes of 20, 30, 40, and 50

```
power onecorrelation 0, n(20(10)50) power(.8)
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power onecorrelation  $r_0$   $r_a$  [ , power(numlist) options ]
```

Compute power

```
power onecorrelation  $r_0$   $r_a$  , n(numlist) [options ]
```

Compute effect size and target correlation

```
power onecorrelation  $r_0$  , n(numlist) power(numlist) [options ]
```

where  $r_0$  is the null (hypothesized) correlation or the value of the correlation under the null hypothesis, and  $r_a$  is the alternative (target) correlation or the value of the correlation under the alternative hypothesis.  $r_0$  and  $r_a$  may each be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 **numlist**).

<i>options</i>	Description
Main	
* <u>a</u> lpha( <i>numlist</i> )	significance level; default is alpha(0.05)
* <u>p</u> ower( <i>numlist</i> )	power; default is power(0.8)
* <u>b</u> eta( <i>numlist</i> )	probability of type II error; default is beta(0.2)
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>n</u> fractional	allow fractional sample size
* <u>d</u> iff( <i>numlist</i> )	difference between the alternative correlation and the null correlation, $r_a - r_0$ ; specify instead of the alternative correlation $r_a$
<u>d</u> irection( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is direction(upper), which means that the postulated value of the parameter is larger than the hypothesized value
<u>o</u> nesided	one-sided test; default is two sided
<u>p</u> arallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>n</u> o ] <u>t</u> able [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>s</u> aving( <i>filename</i> [ , replace ])	save the table data to <i>filename</i> ; use replace to overwrite existing <i>filename</i>
Graph	
<u>g</u> raph [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
Iteration	
<u>i</u> nit(#)	initial value for the sample size or correlation
<u>i</u> terate(#)	maximum number of iterations; default is iterate(500)
<u>t</u> olerance(#)	parameter tolerance; default is tolerance(1e-12)
<u>f</u> tolerance(#)	function tolerance; default is ftolerance(1e-12)
[ <u>n</u> o ] <u>l</u> og	suppress or display iteration log
[ <u>n</u> o ] <u>d</u> ots	suppress or display iterations as dots
<u>n</u> otitle	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 **numlist**.

notitle does not appear in the dialog box.

where *tablespec* is

*column* [ :*label* ] [ *column* [ :*label* ] [ ... ] ] [ , *tableopts* ]

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
alpha	significance level	$\alpha$
power	power	$1 - \beta$
beta	type II error probability	$\beta$
N	number of subjects	$N$
delta	effect size	$\delta$
r0	null correlation	$\rho_0$
ra	alternative correlation	$\rho_a$
diff	difference between the alternative and null correlations	$\rho_a - \rho_0$
target	target parameter; synonym for ra	
_all	display all supported columns	

Column beta is shown in the default table in place of column power if specified.

## Options

### Main

alpha(), power(), beta(), n(), nfractional; see [PSS] power. The nfractional option is allowed only for sample-size determination.

diff(*numlist*) specifies the difference between the alternative correlation and the null correlation,  $r_a - r_0$ . You can specify either the alternative correlation  $r_a$  as a command argument or the difference between the two correlations in the diff(#) option. If you specify diff(#), the alternative correlation is computed as  $r_a = r_0 + \#$ . This option is not allowed with the effect-size determination.

direction(), onesided, parallel; see [PSS] power.

### Table

table, table(), notable; see [PSS] power, table.

saving(); see [PSS] power.

### Graph

graph, graph(); see [PSS] power, graph. Also see the *column* table for a list of symbols used by the graphs.

### Iteration

init(#) specifies the initial value for the iteration procedure. Iteration is used to compute sample size or target correlation for a two-sided test. The default initial value for the estimated parameter is obtained from the corresponding closed-form one-sided computation using the significance level  $\alpha/2$ .

iterate(), tolerance(), ftolerance(), log, nolog, dots, nodots; see [PSS] power.

The following option is available with power onecorrelation but is not shown in the dialog box: notitle; see [PSS] power.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*  
*Using power onecorrelation*  
*Computing sample size*  
*Computing power*  
*Computing effect size and target correlation*  
*Performing hypothesis tests on correlation*

This entry describes the **power onecorrelation** command and the methodology for power and sample-size analysis for a one-sample correlation test. See [PSS] **intro** for a general introduction to power and sample-size analysis and [PSS] **power** for a general introduction to the **power** command using hypothesis tests.

## Introduction

Correlation analysis emanates from studies quantifying dependence between random variables, such as dependence between height and weight of individuals, between blood pressure and cholesterol levels, between the SAT scores of students and their first-year grade point average, between the number of minutes per week customers spend using a new fitness program and the number of pounds lost, and many more.

The correlation coefficient  $\rho$  is commonly used to measure the strength of such dependence. We consider Pearson's correlation obtained by standardizing the covariance between two random variables. As a result, the correlation coefficient has no units and ranges between  $-1$  and  $1$ .

This entry describes power and sample-size analysis for the inference about the population correlation coefficient performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: \rho = \rho_0$  versus the two-sided alternative hypothesis  $H_a: \rho \neq \rho_0$ , the upper one-sided alternative  $H_a: \rho > \rho_0$ , or the lower one-sided alternative  $H_a: \rho < \rho_0$ . In most applications, the null value of the correlation,  $\rho_0$ , is set to zero.

Statistical inference on the correlation coefficient requires a distributional assumption between two random variables—bivariate normality with correlation  $\rho$ . The distribution of the sample correlation coefficient is rather complicated except under the null hypothesis of the true correlation being zero,  $H_0: \rho = 0$ , under which it is a Student's  $t$  distribution (for example, see Graybill [1961, 209]). The general inference of  $H_0: \rho = \rho_0$  is based on the asymptotic approximation.

One common approach in testing hypotheses concerning the correlation parameter is to use an inverse hyperbolic tangent transformation,  $\tanh^{-1}(x) = 0.5 \ln(1+x)/\ln(1-x)$ , also known as Fisher's  $z$  transformation when applied to the correlation coefficient (Fisher 1915). Specifically, if  $\hat{\rho}$  is the sample correlation coefficient and  $n$  is the sample size, Fisher (1915) showed that

$$\sqrt{n-3} \{ \tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho) \} \sim N(0, 1)$$

for  $n$  as small as 10, although the approximation tends to perform better for  $n > 25$ . The null hypothesis  $H_0: \rho = \rho_0$  is equivalent to  $H_0: \tanh^{-1}(\rho) = \tanh^{-1}(\rho_0)$ . The latter test is referred to as Fisher's  $z$  test.

**power onecorrelation** performs computations based on the asymptotic Fisher's  $z$  test.

## Using power onecorrelation

`power onecorrelation` computes sample size, power, or target correlation for a one-sample correlation test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

To compute sample size, you must specify the correlations under the null and alternative hypotheses,  $r_0$  and  $r_a$ , respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the sample size in the `n()` option and the correlations under the null and alternative hypotheses,  $r_0$  and  $r_a$ , respectively.

Instead of the alternative correlation,  $r_a$ , you may specify the difference  $r_a - r_0$  between the alternative correlation and the null correlation in the `diff()` option when computing sample size or power.

To compute effect size, the difference between the alternative and null correlations, and target correlation, you must specify the sample size in the `n()` option, the power in the `power()` option, the null correlation  $r_0$ , and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the target correlation is assumed to be larger than the specified null value. This is also equivalent to the assumption of a positive effect size. You can change the direction to lower, which means that the target correlation is assumed to be smaller than the specified null value, by specifying the `direction(lower)` option. This is equivalent to assuming a negative effect size.

By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS] **unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

The sample-size and effect-size determinations for a two-sided test require iteration. The default initial value for the estimated parameter is obtained from the corresponding closed-form one-sided computation using the significance level  $\alpha/2$ . The default initial value may be changed by specifying the `init()` option. See [PSS] **power** for the descriptions of other options that control the iteration procedure.

In the following sections, we describe the use of `power onecorrelation` accompanied by examples for computing sample size, power, and target correlation.

## Computing sample size

To compute sample size, you must specify the correlations under the null and alternative hypotheses,  $r_0$  and  $r_a$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a one-sample correlation test

Consider a study where the goal is to test the existence of a positive correlation between height and weight of individuals, that is,  $H_0: \rho = 0$  versus  $H_a: \rho > 0$ . Before conducting the study, we wish to determine the sample size required to detect a correlation of at least 0.5 with 80% power using a one-sided 5%-level test. We specify the values 0 and 0.5 of the null and alternative correlations after the command name. The only option we need to specify is `onesided`, which requests a one-sided test. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults.

```

. power onecorrelation 0 0.5, onesided
Estimated sample size for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r > r0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.5000
    r0 =      0.0000
    ra =      0.5000
Estimated sample size:
    N =        24

```

A sample of 24 individuals must be obtained to detect a correlation of at least 0.5 between height and weight with 80% power using an upper one-sided 5%-level test.

If we anticipate a stronger correlation of, say, 0.7, we will need a sample of only 12 subjects:

```

. power onecorrelation 0 0.7, onesided
Estimated sample size for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r > r0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7000
    r0 =      0.0000
    ra =      0.7000
Estimated sample size:
    N =        12

```

This is consistent with our expectation that in this example, as the alternative correlation increases, the required sample size decreases.

◀

## Computing power

To compute power, you must specify the sample size in the `n()` option and the correlations under the null and alternative hypotheses,  $r_0$  and  $r_a$ , respectively.

### ► Example 2: Power of a one-sample correlation test

Continuing with [example 1](#), we will suppose that we are designing a pilot study and would like to keep our initial sample small. We want to find out how much the power of the test would decrease if we sample only 15 individuals. To compute power, we specify the sample size in `n()`:



```

. power onecorrelation 0 0.5, n(15) onesided
Estimated power for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r > r0
Study parameters:
    alpha = 0.0500
      N = 15
    delta = 0.5000
      r0 = 0.0000
      ra = 0.5000
Estimated power:
    power = 0.6018

```

The power decreases to roughly 60%. The power to detect a correlation of 0.5 is low for a sample of 15 individuals. We should consider either a larger target correlation or a larger sample.

◀

### ▷ Example 3: Nonzero null hypothesis

To demonstrate power computation for a nonzero null hypothesis, we consider an example from Graybill (1961, 211) of a hypothetical study where the primary interest lies in the correlation between protein and fat for a particular breed of dairy cow. The sample contains 24 cows. Suppose we want to compute the power of a two-sided test with the null hypothesis  $H_0: \rho = 0.5$  versus the alternative hypothesis  $H_a: \rho \neq 0.5$ . Suppose that the correlation between protein and fat was estimated to be 0.3 in an earlier pilot study. We can compute power by typing

```

. power onecorrelation 0.5 0.3, n(24)
Estimated power for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r != r0
Study parameters:
    alpha = 0.0500
      N = 24
    delta = -0.2000
      r0 = 0.5000
      ra = 0.3000
Estimated power:
    power = 0.1957

```

The power of the test is estimated to be 0.1957 for the above study, which is unacceptably low for practical purposes. We should consider either increasing the sample size or choosing a different target correlation.

In fact, the author considers two values of the alternative correlation, 0.3 and 0.2. We can compute the corresponding powers for the multiple values of the alternative correlation by enclosing them in parentheses, as we demonstrate below.

```
. power onecorrelation 0.5 (0.3 0.2), n(24)
Estimated power for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r != r0
```

alpha	power	N	delta	r0	ra
.05	.1957	24	-.2	.5	.3
.05	.3552	24	-.3	.5	.2

The power increases from 20% to 36% as the alternative correlation decreases from 0.3 to 0.2. It may seem counterintuitive that the power increased as the value of the alternative correlation decreased. It is important to remember that the power of the correlation test is an increasing function of the magnitude of the effect size,  $\delta = \rho_a - \rho_0$ , the difference between the alternative and null values of the correlation. So the power of 20% corresponds to  $|\delta| = |0.3 - 0.5| = 0.2$ , and the power of 36% corresponds to  $|\delta| = 0.3$ .

◀

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

## Computing effect size and target correlation

In a one-sample correlation analysis, the effect size is commonly defined as the difference between the alternative correlation and the null correlation,  $\delta = \rho_a - \rho_0$ . For a zero-null hypothesis, the effect size is the alternative correlation.

The distribution of the sample correlation coefficient is symmetric when the population correlation  $\rho$  is zero and becomes negatively skewed as  $\rho$  approaches 1 and positively skewed as  $\rho$  approaches  $-1$ . Also the sampling variance approaches zero as  $|\rho|$  approaches 1. Clearly, the power of a correlation test depends on the magnitude of the true correlation; therefore,  $\delta$  alone is insufficient for determining the power. In other words, for the same difference  $\delta = 0.1$ , the power to detect the difference between  $\rho_a = 0.2$  and  $\rho_0 = 0.1$  is not the same as that between  $\rho_a = 0.3$  and  $\rho_0 = 0.2$ ; the former difference is associated with the larger sampling variability and thus has lower power to be detected.

Sometimes, we may be interested in determining the smallest effect (for a given null correlation) and the corresponding target correlation that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and null correlation must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $\rho_a > \rho_0$ , or lower, meaning  $\rho_a < \rho_0$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ► Example 4: Minimum detectable value of the correlation

Continuing [example 2](#), we will compute the minimum positive correlation that can be detected given a sample of 15 individuals and 80% power. To solve for target correlation, we specify the null correlation of 0 after the command, sample size `n(15)`, and power `power(0.8)`:

```

. power onecorrelation 0, n(15) power(0.8) onesided
Estimated target correlation for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r > r0; ra > r0
Study parameters:
      alpha =   0.0500
      power =   0.8000
         N =     15
        r0 =   0.0000
Estimated effect size and target correlation:
      delta =   0.6155
         ra =   0.6155

```

We find that the minimum value of the correlation as well as of the effect size that would yield a statistically significant result in this study is 0.6155.

In this example, we computed the correlation assuming an upper direction,  $\rho_a > \rho_0$ , or a positive effect,  $\delta > 0$ . To request a lower direction, or a negative effect, we can specify the `direction(lower)` option.

◀

## Performing hypothesis tests on correlation

After the data are collected, you can use the `pwcorr` command to test for correlation; see [R] [correlate](#).

### ▶ Example 5: Testing for correlation

For example, consider `auto.dta`, which contains various characteristics of 74 cars. Suppose that our study goal is to investigate the existence of a correlation between miles per gallon and weights of cars. The corresponding variables in the dataset are `mpg` and `weight`. We use `pwcorr` to perform the test.

```

. use http://www.stata-press.com/data/r14/auto
(1978 Automobile Data)
. pwcorr mpg weight, sig

```

	mpg	weight
mpg	1.0000	
weight	-0.8072	1.0000
	0.0000	

The test finds a statistically significant negative correlation of  $-0.8$  between `mpg` and `weight`; the  $p$ -value is less than 0.0001.

We use the parameters of this study to perform a sample-size analysis we would have conducted before the study.

```

. power onecorrelation 0 -0.8, power(0.9) alpha(0.01)
Performing iteration ...
Estimated sample size for a one-sample correlation test
Fisher's z test
Ho: r = r0 versus Ha: r != r0
Study parameters:
      alpha =    0.0100
      power =    0.9000
      delta =   -0.8000
      r0 =     0.0000
      ra =    -0.8000
Estimated sample size:
      N =        16

```

We find that the sample size required to detect a correlation of  $-0.8$  with 90% power using a 1%-level two-sided test is 16. The current sample contains many more cars (74), which would allow us to detect a potentially smaller (in absolute value) correlation.

We should point out that `pwcorr` performs an exact one-sample correlation test of  $H_0: \rho = 0$ , which uses a Student's  $t$  distribution. `power onecorrelation` performs computation for an approximate Fisher's  $z$  test, which uses the normal distribution.

◀

## Stored results

`power onecorrelation` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(r0)</code>	correlation under the null hypothesis
<code>r(ra)</code>	correlation under the alternative hypothesis
<code>r(diff)</code>	difference between the alternative and null correlations
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or the correlation
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	<code>onecorrelation</code>
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Let  $\rho$  denote Pearson's correlation and  $\hat{\rho}$  its estimator, the sample correlation coefficient.

A one-sample correlation test involves testing the null hypothesis  $H_0: \rho = \rho_0$  versus the two-sided alternative hypothesis  $H_a: \rho \neq \rho_0$ , the upper one-sided alternative  $H_a: \rho > \rho_0$ , or the lower one-sided alternative  $H_a: \rho < \rho_0$ .

The exact distribution of the sample correlation coefficient  $\hat{\rho}$  is complicated for testing a general hypothesis  $H_0: \rho = \rho_0 \neq 0$ . An alternative is to apply an inverse hyperbolic tangent,  $\tanh^{-1}(\cdot)$ , or Fisher's  $z$  transformation to  $\hat{\rho}$  to approximate its sampling distribution by a normal distribution (Fisher 1915). The hypothesis  $H_0: \rho = \rho_0$  is then equivalent to  $H_0: \tanh^{-1}(\rho) = \tanh^{-1}(\rho_0)$ .

Let  $Z = \tanh^{-1}(\hat{\rho}) = 0.5 \ln\{(1 + \hat{\rho})/(1 - \hat{\rho})\}$  denote Fisher's  $z$  transformation of  $\hat{\rho}$ . Then  $Z$  follows a normal distribution with mean  $\mu_z = \tanh^{-1}(\rho)$  and standard deviation  $\sigma_z = 1/\sqrt{n-3}$ , where  $n$  is the sample size (for example, see Graybill [1961, 211]; Anderson [2003, 134]).

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha/k}$  and  $z_\beta$  be the  $(1 - \alpha/k)$ th and the  $\beta$ th quantiles of the standard normal distribution. Denote  $\delta_z = \mu_a - \mu_0 = \tanh^{-1}(\rho_a) - \tanh^{-1}(\rho_0)$ .

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi\left(\frac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\ \Phi\left(-\frac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\ \Phi\left(\frac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $\Phi(\cdot)$  is the cdf of a standard normal distribution.

The sample size  $n$  for a one-sided test is computed using

$$n = 3 + \left(\frac{z_{1-\alpha} - z_\beta}{\delta_z}\right)^2$$

See, for example, Lachin (1981) for the case of  $\rho_0 = 0$ .

The minimum detectable value of the correlation is obtained by first finding the corresponding minimum value of  $\mu_a$  and then applying the inverse Fisher's  $z$  transformation to the  $\mu_a$ :

$$\rho_a = \frac{e^{2\mu_a} - 1}{e^{2\mu_a} + 1}$$

For a one-sided test,  $\mu_a = \mu_0 + \sigma_z(z_{1-\alpha} - z_\beta)$  when  $\mu_a > \mu_0$ , and  $\mu_a = \mu_0 - \sigma_z(z_{1-\alpha} - z_\beta)$  when  $\mu_a < \mu_0$ .

For a two-sided test, the sample size  $n$  and  $\mu_a$  are obtained by iteratively solving the two-sided power equation in (1) for  $n$  and  $\mu_a$ , respectively. The initial values are obtained from the respective formulas for the one-sided computation with the significance level  $\alpha/2$ .

If the `nfractional` option is not specified, the computed sample size is rounded up.

## References

- Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley.
- Fisher, R. A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10: 507–521.
- Graybill, F. A. 1961. *An Introduction to Linear Statistical Models, Vol. 1*. New York: McGraw–Hill.
- Lachin, J. M. 1981. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 2: 93–113.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **correlate** — Correlations (covariances) of variables or coefficients

# Title

**power twocorrelations** — Power analysis for a two-sample correlations test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power twocorrelations` computes sample size, power, or the experimental-group correlation for a two-sample correlations test. By default, it computes sample size for given power and the values of the control-group and experimental-group correlations. Alternatively, it can compute power for given sample size and values of the control-group and experimental-group correlations or the experimental-group correlation for given sample size, power, and the control-group correlation. Also see [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \rho_1 = \rho_2$  versus  $H_a: \rho_1 \neq \rho_2$  when control-group correlation  $r_1 = 0.4$  and experimental-group correlation  $r_2 = 0.1$  with default power of 0.8 and significance level  $\alpha = 0.05$

```
power twocorrelations .4 .1
```

As above, for  $r_2$  equal to  $-0.05$ ,  $0$ ,  $0.05$ , and  $0.1$

```
power twocorrelations .4 (-.05(.05).1)
```

As above, but for a one-sided test

```
power twocorrelations .4 (-.05(.05).1) onesided
```

Specify  $r_1$  and difference  $r_2 - r_1 = -0.3$

```
power twocorrelations .4, diff(-.3)
```

As above, but specify a sample size of 200 for group 1 and compute sample size for group 2

```
power twocorrelations 0.4, diff(-0.3) n1(200) compute(n2)
```

Power for  $r_1 = 0.4$  and  $r_2 = -0.15$  with a sample size of 100 and default  $\alpha = 0.05$

```
power twocorrelations .4 -.15, n(100)
```

As above, but specify sample size of 50 and 65 for groups 1 and 2, respectively

```
power twocorrelations .4 -.15 , n1(50) n2(65)
```

Power for sample sizes of 60, 70, 80, 90, and 100

```
power twocorrelations .4 -.15, n(60(10)100)
```

As above, but display results in a graph of power versus sample size

```
power twocorrelations .4 -.15, n(60(10)100) graph
```

Effect size and experimental-group correlation for  $r_1 = 0.4$  with sample size of 150 and power of 0.85

```
power twocorrelations .4, n(150) power(.85)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

```
power twocorrelations  $r_1$   $r_2$  [ , power(numlist) options ]
```

*Compute power*

```
power twocorrelations  $r_1$   $r_2$  , n(numlist) [options ]
```

*Compute effect size and experimental-group correlation*

```
power twocorrelations  $r_1$  , n(numlist) power(numlist) [options ]
```

where  $r_1$  is the correlation in the control (reference) group, and  $r_2$  is the correlation in the experimental (comparison) group.  $r_1$  and  $r_2$  may each be specified either as one number or as a list of values in parentheses (see [U] [11.1.8 numlist](#)).



<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
* <u>n1</u> ( <i>numlist</i> )	sample size of the control group
* <u>n2</u> ( <i>numlist</i> )	sample size of the experimental group
* <u>nratio</u> ( <i>numlist</i> )	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<u>compute</u> ( <i>n1</i>   <i>n2</i> )	solve for $N1$ given $N2$ or for $N2$ given $N1$
<u>nfractional</u>	allow fractional sample sizes
* <u>diff</u> ( <i>numlist</i> )	difference between the experimental-group and control-group correlations, $r_2 - r_1$ ; specify instead of the experimental-group correlation $r_2$
<u>direction</u> ( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<u>saving</u> ( <i>filename</i> [ , <u>replace</u> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<u>init</u> (#)	initial value for sample sizes or experimental-group correlation
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
[ <u>no</u> ] <u>log</u>	suppress or display iteration log
[ <u>no</u> ] <u>dots</u>	suppress or display iterations as dots
<u>notitle</u>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] [11.1.8 numlist](#).

`notitle` does not appear in the dialog box.

where *tablespec* is

```
column[:label] [column[:label] [...]] [, tableopts]
```

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N1</code>	number of subjects in the control group	$N_1$
<code>N2</code>	number of subjects in the experimental group	$N_2$
<code>nratio</code>	ratio of sample sizes, experimental to control	$N_2/N_1$
<code>delta</code>	effect size	$\delta$
<code>r1</code>	control-group correlation	$\rho_1$
<code>r2</code>	experimental-group correlation	$\rho_2$
<code>diff</code>	difference between the experimental-group correlation and the control-group correlation	$\rho_2 - \rho_1$
<code>target</code>	target parameter; synonym for <code>r2</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Columns `diff` and `nratio` are shown in the default table if specified.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `compute()`, `nfractional`; see [\[PSS\] power](#).

`diff(numlist)` specifies the difference between the experimental-group correlation and the control-group correlation,  $r_2 - r_1$ . You can specify either the experimental-group correlation  $r_2$  as a command argument or the difference between the two correlations in `diff()`. If you specify `diff(#)`, the experimental-group correlation is computed as  $r_2 = r_1 + \#$ . This option is not allowed with the effect-size determination.

`direction()`, `onesided`, `parallel`; see [\[PSS\] power](#).

### Table

`table`, `table()`, `notable`; see [\[PSS\] power, table](#).

`saving()`; see [\[PSS\] power](#).

### Graph

`graph`, `graph()`; see [\[PSS\] power, graph](#). Also see the *column* table for a list of symbols used by the graphs.

Iteration

`init(#)` specifies the initial value for the estimated parameter. For sample-size determination, the estimated parameter is either the control-group size  $n_1$  or, if `compute(n2)` is specified, the experimental-group size  $n_2$ . For the effect-size determination, the estimated parameter is the experimental-group correlation  $r_2$ . The default initial values for a two-sided test are obtained as a closed-form solution for the corresponding one-sided test with the significance level  $\alpha/2$ .

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power twocorrelations` but is not shown in the dialog box: `notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power twocorrelations*

*Computing sample size*

*Computing power*

*Computing effect size and experimental-group correlation*

*Testing a hypothesis about two independent correlations*

This entry describes the `power twocorrelations` command and the methodology for power and sample-size analysis for a two-sample correlations test. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

There are many examples of studies where a researcher may want to compare two correlations. A winemaker may want to test the null hypothesis that the correlation between fermentation time and alcohol level is the same for pinot noir grapes and merlot grapes. An education researcher may want to test the null hypothesis that the correlation of verbal and math SAT scores is the same for males and females. Or a genetics researcher may want to test the null hypothesis that the correlation of the cholesterol levels in identical twins raised together is equal to the correlation of the cholesterol levels in identical twins raised apart.

Correlation analysis emanates from studies quantifying dependence between random variables. The correlation coefficient  $\rho$  is commonly used to measure the strength of such dependence. We consider Pearson's correlation obtained by standardizing the covariance between two random variables. As a result, the correlation coefficient has no units and ranges between  $-1$  and  $1$ .

This entry describes power and sample-size analysis for the inference about two population correlation coefficients performed using hypothesis testing. Specifically, we consider the null hypothesis  $H_0: \rho_2 = \rho_1$  versus the two-sided alternative hypothesis  $H_a: \rho_2 \neq \rho_1$ , the upper one-sided alternative  $H_a: \rho_2 > \rho_1$ , or the lower one-sided alternative  $H_a: \rho_2 < \rho_1$ .

One common approach in testing hypotheses concerning the correlation parameter is to use an inverse hyperbolic tangent transformation,  $\tanh^{-1}(x) = 0.5 \ln(1+x)/\ln(1-x)$ , also known as Fisher's  $z$  transformation when applied to the correlation coefficient (Fisher 1915). Specifically, if  $\hat{\rho}$  is the sample correlation coefficient and  $n$  is the sample size, Fisher (1915) showed that

$$\sqrt{n-3} \{ \tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho) \} \sim N(0, 1)$$

for  $n$  as small as 10, although the approximation tends to perform better for  $n > 25$ . For a two-sample correlations test, the null hypothesis  $H_0: \rho_1 = \rho_2$  is equivalent to  $H_0: \tanh^{-1}(\rho_1) = \tanh^{-1}(\rho_2)$ . The latter test is referred to as the two-sample Fisher's  $z$  test.

`power twocorrelations` performs computations based on the asymptotic two-sample Fisher's  $z$  test.

## Using `power twocorrelations`

`power twocorrelations` computes sample size, power, or experimental-group correlation for a two-sample correlations test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced or equal-allocation design; see [PSS] **unbalanced designs** for a description of how to specify an unbalanced design.

To compute the total sample size, you must specify the control- and experimental-group correlations,  $r_1$  and  $r_2$ , respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

Instead of the total sample size, you can compute one of the group sizes given the other one. To compute the control-group sample size, you must specify the `compute(n1)` option and the sample size of the experimental group in the `n2()` option. Likewise, to compute the experimental-group sample size, you must specify the `compute(n2)` option and the sample size of the control group in the `n1()` option.

To compute power, you must specify the total sample size in the `n()` option and the control- and experimental-group correlations,  $r_1$  and  $r_2$ , respectively.

Instead of the experimental-group correlation  $r_2$ , you may specify the difference  $r_2 - r_1$  between the experimental-group correlation and the control-group correlation in the `diff()` option when computing sample size or power.

To compute effect size, the difference between the experimental-group and the control-group correlation, and the experimental-group correlation, you must specify the total sample size in the `n()` option, the power in the `power()` option, the control-group correlation  $r_1$ , and optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the experimental-group correlation is assumed to be larger than the specified control-group value. You can change the direction to be lower, which means that the experimental-group correlation is assumed to be smaller than the specified control-group value, by specifying the `direction(lower)` option.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see *Two samples* in [PSS] **unbalanced designs** for more details.

In the following sections, we describe the use of `power twocorrelations` accompanied with examples for computing sample size, power, and experimental-group correlation.

## Computing sample size

To compute sample size, you must specify the control- and experimental-group correlations,  $r_1$  and  $r_2$ , respectively, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ▷ Example 1: Sample size for a two-sample correlations test

Consider a study in which investigators are interested in testing whether the correlation between height and weight differs for males and females. The null hypothesis  $H_0: \rho_F = \rho_M$  is tested against the alternative hypothesis  $H_a: \rho_F \neq \rho_M$ , where  $\rho_F$  is the correlation between height and weight for females, and  $\rho_M$  is the correlation between height and weight for males.

Before conducting the study, investigators wish to determine the minimum sample size required to detect a difference between the correlation of 0.3 for females and a correlation of 0.5 for males with 80% power using a two-sided 5%-level test. We specify the values 0.3 and 0.5 as the control- and experimental-group correlations after the command name. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults. To compute the total sample size assuming equal-group allocation, we type

```
. power twocorrelations 0.3 0.5
Performing iteration ...
Estimated sample sizes for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.2000
    r1 = 0.3000
    r2 = 0.5000
Estimated sample sizes:
    N = 554
    N per group = 277
```

A total sample of 554 individuals, 277 in each group, must be obtained to detect the difference between correlations of females and males when the correlations are 0.3 and 0.5, respectively, with 80% power using a two-sided 5%-level test.

◀

### ▷ Example 2: Computing one of the group sizes

Continuing with [example 1](#), we will suppose that for some reason, we can enroll only 250 male subjects in our study. We want to know how many female subjects we need to recruit to maintain the 80% power for detecting the difference as described in [example 1](#). To do so, we use the combination of `compute(n1)` and `n2()`:

```
. power twocorrelations 0.3 0.5, n2(250) compute(n1)
Performing iteration ...
Estimated sample sizes for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.2000
      r1 = 0.3000
      r2 = 0.5000
      N2 = 250
Estimated sample sizes:
      N = 559
      N1 = 309
```

We need 309 females for a total sample size of 559 subjects, which is larger than the required total sample size for the corresponding balanced design from [example 1](#).

◀

### ▷ Example 3: Unbalanced design

By default, `power twocorrelations` computes sample size for a balanced or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size for an unbalanced design by specifying the `nratio()` option.

Continuing with [example 1](#), we will suppose that we anticipate to recruit twice as many males as females; that is,  $n_2/n_1 = 2$ . We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```
. power twocorrelations 0.3 0.5, nratio(2)
Performing iteration ...
Estimated sample sizes for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.2000
      r1 = 0.3000
      r2 = 0.5000
      N2/N1 = 2.0000
Estimated sample sizes:
      N = 624
      N1 = 208
      N2 = 416
```

Also see [Two samples in \[PSS\] unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.

◀

## Computing power

To compute power, you must specify the sample size in the `n()` option and the control- and the experimental-group correlations,  $r_1$  and  $r_2$ , respectively.

### ▷ Example 4: Power of a two-sample correlations test

Continuing with [example 1](#), we will suppose that we are designing a study and anticipate a total sample of 500 subjects. To compute the power corresponding to this sample size given the study parameters from [example 1](#), we specify the sample size in `n()`:

```
. power twocorrelations 0.3 0.5, n(500)
Estimated power for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1
Study parameters:
      alpha =    0.0500
         N =     500
  N per group =   250
      delta =    0.2000
         r1 =    0.3000
         r2 =    0.5000
Estimated power:
      power =    0.7595
```

With a smaller sample of 500 subjects compared with [example 1](#), we obtain a power of roughly 76%. ◀

### ▷ Example 5: Multiple values of study parameters

In this example, we assess the effect of varying the value of the correlation of the male group on the power of our study. We supply a list of correlations in parentheses for the second command argument:

```
. power twocorrelations 0.3 (0.4(0.1)0.9), n(500)
Estimated power for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1
```

alpha	power	N	N1	N2	delta	r1	r2
.05	.2452	500	250	250	.1	.3	.4
.05	.7595	500	250	250	.2	.3	.5
.05	.9894	500	250	250	.3	.3	.6
.05	1	500	250	250	.4	.3	.7
.05	1	500	250	250	.5	.3	.8
.05	1	500	250	250	.6	.3	.9

From the table, we see that the power increases from 25% to 100% as the correlation increases from 0.4 to 0.9.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

## Computing effect size and experimental-group correlation

Effect size  $\delta$  for a two-sample correlations test is defined as the difference between the experimental-group and control-group correlations,  $\delta = \rho_2 - \rho_1$ .

Sometimes, we may be interested in determining the smallest effect (for a given control-group correlation) that yields a statistically significant result for prespecified sample size and power. In this case, power, sample size, and control-group correlation must be specified. In addition, you must also decide on the direction of the effect: upper, meaning  $\rho_2 > \rho_1$ , or lower, meaning  $\rho_2 < \rho_1$ . The direction may be specified in the `direction()` option; `direction(upper)` is the default.

### ► Example 6: Minimum detectable value of the experimental-group correlation

Continuing with [example 4](#), we will compute the smallest positive correlation of the male group that can be detected given a total sample of 500 individuals and a power of 80%. To solve for the experimental-group correlation, after the command, we specify the control-group correlation of 0.3, total sample size `n(500)`, and power `power(0.8)`:

```
. power twocorrelations 0.3, n(500) power(0.8)
Performing iteration ...
Estimated experimental-group correlation for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1; r2 > r1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =     500
N per group =    250
         r1 =    0.3000
Estimated effect size and experimental-group correlation:
      delta =    0.2092
         r2 =    0.5092
```

We find that the minimum value of the experimental-group correlation that would yield a statistically significant result in this study is 0.5092, and the corresponding effect size is 0.2092.

In this example, we computed the correlation assuming an upper direction,  $\rho_2 > \rho_1$ , or a positive effect,  $\delta > 0$ . To request a lower direction, or a negative effect, we can specify the `direction(lower)` option.

◀

## Testing a hypothesis about two independent correlations

After data are collected, we can use the `mvtest` command to test the equality of two independent correlations using an asymptotic likelihood-ratio test; see [\[MV\] mvtest](#) for details. We can also manually perform a two-sample Fisher's  $z$  test on which `power twocorrelations` bases its computations. We demonstrate our examples using the `genderpsych` dataset from the [\[MV\] mvtest correlations](#) entry.

### ► Example 7: Comparing two correlations using mvtest

Consider a sample of 64 individuals with equal numbers of males and females. We would like to know whether the correlation between the pictorial inconsistencies (variable `y1`) and vocabulary (variable `y4`) is the same between males and females.



```
. use http://www.stata-press.com/data/r14/genderpsych
(Four Psychological Test Scores, Rencher and Christensen (2012))
. mvtest correlations y1 y4, by(gender)
Test of equality of correlation matrices across samples
      Jennrich chi2(1) =      2.16
      Prob > chi2 =      0.1415
```

The reported  $p$ -value is 0.1415, so we do not have sufficient evidence to reject the null hypothesis about the equality of the two correlations.

◀

### ▶ Example 8: Two-sample Fisher's $z$ test

To compute a two-sample Fisher's  $z$  test manually, we perform the following steps. We first compute the estimates of correlation coefficients for each group by using the `correlate` command; see [R] [correlate](#). We then compute Fisher's  $z$  test statistic and its corresponding  $p$ -value.

We compute and store correlation estimates for males in the `r1` scalar and the corresponding sample size in the `N1` scalar.

```
. /* compute and store correlation and sample size for males */
. correlate y1 y4 if gender==1
(obs=32)
```

	y1	y4
y1	1.0000	
y4	0.5647	1.0000

```
. scalar r1 = r(rho)
. scalar N1 = r(N)
```

We store the corresponding results for females in `r2` and `N2`.

```
. /* compute and store correlation and sample size for females */
. correlate y1 y4 if gender==2
(obs=32)
```

	y1	y4
y1	1.0000	
y4	0.2596	1.0000

```
. scalar r2 = r(rho)
. scalar N2 = r(N)
```

We now compute the  $z$  test statistic, stored in the `Z` scalar, by applying Fisher's  $z$  transformation to the obtained correlation estimates `r1` and `r2` and using the cumulative function of the standard normal distribution `normal()` to compute the  $p$ -value.

```
. /* compute Fisher's z statistic and p-value and display results */
. scalar mu_Z = atanh(r2) - atanh(r1)
. scalar sigma_Z = sqrt(1/(N1-3)+1/(N2-3))
. scalar Z = mu_Z/sigma_Z
. scalar pvalue = 2*normal(-abs(Z))
. display "Z statistic = " %8.4g Z _n "P-value = " %8.4g pvalue
Z statistic = -1.424
P-value = .1543
```

The  $p$ -value is 0.1543 and is close to the  $p$ -value reported by `mvtest` in [example 7](#).

The estimates of the correlations obtained from `correlate` are 0.5647 for males and 0.2596 for females. We may check how many subjects we need to detect the difference between these two correlation values.

```
. power twocorrelations 0.5647 0.2596
Performing iteration ...
Estimated sample sizes for a two-sample correlations test
Fisher's z test
Ho: r2 = r1 versus Ha: r2 != r1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.3051
     r1 =    0.5647
     r2 =    0.2596
Estimated sample sizes:
      N =      232
  N per group =   116
```

We need a total of 232 subjects, 116 per group in a balanced design, to have a power of 80% to detect the difference between the given values of correlations for males and females.

◀

## Stored results

`power twocorrelations` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(r1)</code>	control-group correlation
<code>r(r2)</code>	experimental-group correlation
<code>r(diff)</code>	difference between the experimental- and control-group correlations
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for sample sizes or experimental-group correlation
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

Macros	
r(type)	test
r(method)	twocorrelations
r(direction)	upper or lower
r(columns)	displayed table columns
r(labels)	table column labels
r(widths)	table column widths
r(formats)	table column formats
Matrix	
r(pss_table)	table of results

## Methods and formulas

Let  $\rho_1$  and  $\rho_2$  denote Pearson's correlation for the control and the experimental groups, respectively. Let  $\hat{\rho}_1$  and  $\hat{\rho}_2$  denote the corresponding estimators, the sample correlation coefficients.

A two-sample correlations test involves testing the null hypothesis  $H_0: \rho_2 = \rho_1$  versus the two-sided alternative hypothesis  $H_a: \rho_2 \neq \rho_1$ , the upper one-sided alternative  $H_a: \rho_2 > \rho_1$ , or the lower one-sided alternative  $H_a: \rho_2 < \rho_1$ .

The exact distribution of the difference between the sample correlation coefficients  $\hat{\rho}_1$  and  $\hat{\rho}_2$  is complicated for testing the hypothesis  $H_0: \rho_1 - \rho_2 \neq 0$ . An alternative is to apply an inverse hyperbolic tangent,  $\tanh^{-1}(\cdot)$ , or Fisher's  $z$  transformation to the estimators to approximate their sampling distribution by a normal distribution (Fisher 1915). The hypothesis  $H_0: \rho_1 = \rho_2$  is then equivalent to  $H_0: \tanh^{-1}(\rho_1) = \tanh^{-1}(\rho_2)$ .

Let  $Z_1 = \tanh^{-1}(\hat{\rho}_1) = 0.5 \ln\{(1 + \hat{\rho}_1)/(1 - \hat{\rho}_1)\}$  denote Fisher's  $z$  transformation of  $\hat{\rho}_1$ , and let  $Z_2$  denote the corresponding transformation of  $\hat{\rho}_2$ . Then the difference  $\delta_z = Z_2 - Z_1$  follows a normal distribution with mean  $\mu_z = \mu_2 - \mu_1 = \tanh^{-1}(\rho_2) - \tanh^{-1}(\rho_1)$  and standard deviation  $\sigma_z = \sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}$ , where  $n_1$  and  $n_2$  are the sample sizes of the control and the experimental groups, respectively (for example, see Graybill [1961, 211]; Anderson [2003, 134]).

Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $z_{1-\alpha}$  and  $z_\beta$  be the  $(1 - \alpha)$ th and the  $\beta$ th quantiles of the standard normal distribution.

The power  $\pi = 1 - \beta$  is computed using

$$\pi = \begin{cases} \Phi\left(\frac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\ \Phi\left(-\frac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\ \Phi\left(\frac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $\Phi(\cdot)$  is the cdf of a standard normal distribution.

Let  $R = n_2/n_1$  denote the allocation ratio. Then  $n_2 = R \times n_1$  and power can be viewed as a function of  $n_1$ . Therefore, for sample-size determination, the control-group sample size  $n_1$  is computed first. The experimental-group size  $n_2$  is then computed as  $R \times n_1$ , and the total sample size is computed as  $n = n_1 + n_2$ . By default, sample sizes are rounded to integer values; see *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

For a one-sided test, the control-group sample size  $n_1$  is computed as a positive root of the following quadratic equation:

$$\frac{(R + 1)n_1 - 6}{(n_1 - 3)(Rn_1 - 3)} = \left(\frac{\delta_z}{z_{1-\alpha} - z_\beta}\right)^2$$

For a one-sided test, if one of the group sizes is known, the other one is computed using the following formula. For example, to compute  $n_1$  given  $n_2$  we use

$$n_1 = 3 + \frac{1}{\left(\frac{\delta_z}{z_{1-\alpha} - z_\beta}\right)^2 - \frac{1}{n_2 - 3}}$$

The minimum detectable value of the experimental-group correlation is obtained by first finding the corresponding minimum value of  $\mu_2$  and then applying the inverse Fisher's  $z$  transformation to that  $\mu_2$ :

$$\rho_2 = \frac{e^{2\mu_2} - 1}{e^{2\mu_2} + 1}$$

For a one-sided test,  $\mu_2 = \mu_1 + \sigma_z(z_{1-\alpha} - z_\beta)$  when  $\mu_2 > \mu_1$ , and  $\mu_2 = \mu_1 - \sigma_z(z_{1-\alpha} - z_\beta)$  when  $\mu_2 < \mu_1$ .

For a two-sided test, the sample sizes and  $\mu_2$  are obtained by iteratively solving the two-sided power equation in (1) for the respective parameters. The initial values are obtained from the respective formulas for the one-sided computation with the significance level  $\alpha/2$ .

## References

- Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley.
- Fisher, R. A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10: 507–521.
- Graybill, F. A. 1961. *An Introduction to Linear Statistical Models, Vol. 1*. New York: McGraw–Hill.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [MV] **mvtest** — Multivariate tests
- [R] **correlate** — Correlations (covariances) of variables or coefficients

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power oneway` computes sample size, power, or effect size for one-way analysis of variance (ANOVA). By default, it computes sample size for given power and effect size. Alternatively, it can compute power for given sample size and effect size or compute effect size for given sample size, power, and number of groups. Also see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a test of  $H_0: \mu_1 = \mu_2 = \mu_3$  for cell means of 21, 19, and 18 and within-group variance of 16 with default power of 0.8 and significance level  $\alpha = 0.05$

```
power oneway 21 19 18, varerror(16)
```

Same as above, but specify a between-group variance of 1.55 for three groups

```
power oneway, varerror(16) varmeans(1.55) ngroups(3)
```

Same as above, but specify that delta (Cohen's  $f$ ), given by  $\sqrt{\sigma_m^2/\sigma_e^2}$ , equals 0.31

```
power oneway, ngroups(3) delta(.31)
```

As above, but specify delta equal to 0.2, 0.25, 0.3, 0.35, and 0.4

```
power oneway, ngroups(3) delta(.2(.05).4)
```

As above, but show results in a graph of sample size versus effect size (delta)

```
power oneway, ngroups(3) delta(.2(.05).4) graph
```

Sample size for contrast of  $H_0: \mu_1 = (\mu_2 + \mu_3)/2$  or, equivalently,  $H_0: -\mu_1 + 0.5\mu_2 + 0.5\mu_3 = 0$

```
power oneway 21 19 18, varerror(16) contrast(-1 0.5 0.5)
```

For an unbalanced design where the sample size in group 1 is 1.5 times the sample sizes for groups 2 and 3

```
power oneway 21 19 18, varerror(16) grweights(3 2 2)
```

Specify group means for 5 groups using matrix `means` and within-group variance of 34

```
matrix means = (21, 19, 19, 22, 27)
```

```
power oneway means, varerror(34)
```

Power for sample sizes of 99, 108, 117, and 126 with balanced group sizes

```
power oneway 21 19 18, varerror(16) n(99(9)126)
```

As above, but with sample sizes of 25, 25, and 50 for groups 1, 2 and 3, respectively

```
power oneway 21 19 18, varerror(16) n1(25) n2(25) n3(50)
```

As above, sample size allocations treated as parallel sets, each with total sample size 100

```
power oneway 21 19 18, varerror(16) n1(50 25 25) n2(25 50 25) ///  
n3(25 25 50) parallel
```

Effect size and target between-group variance for a sample size of 150 for three groups, power of 0.9, and  $\alpha = 0.01$

```
power oneway, n(150) ngroups(3) power(.9) alpha(0.1)
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power oneway meanspec [, power(numlist) options]
```

Compute power

```
power oneway meanspec, n(numlist) [options]
```

Compute effect size and target between-group variance

```
power oneway, n(numlist) power(numlist) ngroups(#) [varerror(numlist) options]
```

where *meanspec* is either a matrix *matname* containing group means or individual group means specified in a matrix form:

$$m_1 \ m_2 \ [m_3 \ \dots \ m_J]$$

$m_j$ , where  $j = 1, 2, \dots, J$ , is the alternative group mean or the group mean under the alternative hypothesis for the  $j$ th group. Each  $m_j$  may be specified either as one number or as a list of values in parentheses; see [U] 11.1.8 **numlist**.

*matname* is the name of a Stata matrix with  $J$  columns containing values of alternative group means. Multiple rows are allowed, in which case each row corresponds to a different set of  $J$  group means or, equivalently, column  $j$  corresponds to a *numlist* for the  $j$ th group mean.

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample sizes
* <u>npergroup</u> ( <i>numlist</i> )	number of subjects per group; implies balanced design
* <u>n#</u> ( <i>numlist</i> )	number of subjects in group #
<u>grweights</u> ( <i>wgtspec</i> )	group weights; default is one for each group, meaning equal group sizes
<u>ngroups</u> (#)	number of groups
* <u>varmeans</u> ( <i>numlist</i> )	variance of the group means or between-group variance
* <u>varerror</u> ( <i>numlist</i> )	error (within-group) variance; default is <code>varerror(1)</code>
<u>contrast</u> ( <i>contrastspec</i> )	contrast specification for group means
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [ , <i>replace</i> ])	save the table data to <i>filename</i> ; use <i>replace</i> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
Iteration	
<u>init</u> (#)	initial value for the sample size or the effect size; default is to use a bisection algorithm to bound the solution
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
[ <u>no</u> ] <u>log</u>	suppress or display iteration log
[ <u>no</u> ] <u>dots</u>	suppress or display iterations as dots
<u>notitle</u>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 **numlist**.

`notitle` does not appear in the dialog box.

<i>wgtspec</i>	Description
<i>#<sub>1</sub> #<sub>2</sub> ... #<sub>J</sub></i>	<i>J</i> group weights. Weights must be positive and must be integers unless option <code>nfractional</code> is specified. Multiple values for each group weight <i>#<sub>j</sub></i> can be specified as a <i>numlist</i> enclosed in parentheses.
<i>matname</i>	matrix with <i>J</i> columns containing <i>J</i> group weights. Multiple rows are allowed, in which case each row corresponds to a different set of <i>J</i> weights or, equivalently, column <i>j</i> corresponds to a <i>numlist</i> for the <i>j</i> th weight.

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N_per_group</code>	number of subjects per group	$N/N_g$
<code>N_avg</code>	average number of subjects per group	$N_{avg}$
<code>N#</code>	number of subjects in group #	$N_{\#}$
<code>delta</code>	effect size	$\delta$
<code>N_g</code>	number of groups	$N_g$
<code>m#</code>	group mean #	$\mu_{\#}$
<code>Cm</code>	mean contrast	$C \cdot \mu$
<code>c0</code>	null mean contrast	$c_0$
<code>Var_m</code>	group means (between-group) variance	$\sigma_m^2$
<code>Var_Cm</code>	contrast variance	$\sigma_{C\mu}^2$
<code>Var_e</code>	error (within-group) variance	$\sigma_e^2$
<code>grwtg#</code>	group weight #	$w_{\#}$
<code>target</code>	target parameter; synonym for <code>Var_m</code> or <code>Var_Cm</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Column `N_per_group` is available and is shown in the default table only for balanced designs.

Columns `N_avg` and `N#` are shown in the default table only for unbalanced designs.

Columns `m#` are shown in the default table only if group means are specified.

Column `Var_m` is not shown in the default table if the `contrast()` option is specified.

Columns `Cm`, `c0`, and `Var_Cm` are shown in the default table only if the `contrast()` option is specified.

Columns `grwtg#` are not shown in the default table.



## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**.

`npergroup(numlist)` specifies the group size. Only positive integers are allowed. This option implies a balanced design. `npergroup()` cannot be specified with `n()`, `n#()`, or `grweights()`.

`n#(numlist)` specifies the size of the *#*th group. Only positive integers are allowed. All group sizes must be specified. For example, all three options `n1()`, `n2()`, and `n3()` must be specified for a design with three groups. `n#()` cannot be specified with `n()`, `npergroup()`, or `grweights()`.

`grweights(wgtspec)` specifies *J* group weights for an unbalanced design. The weights may be specified either as a list of values or as a matrix, and multiple sets of weights are allowed; see *wgtspec* for details. The weights must be positive and must also be integers unless the `nfractional` option is specified. `grweights()` cannot be specified with `npergroup()` or `n#()`.

`ngroups(#)` specifies the number of groups. At least two groups must be specified. This option is required if *meanspec* is not specified. This option is also required for effect-size determination unless `grweights()` is specified.

`varmeans(numlist)` specifies the variance of the group means or the between-group variance. `varmeans()` cannot be specified with *meanspec* or `contrast()`, nor is it allowed with effect-size determination.

`varerror(numlist)` specifies the error (within-group) variance. The default is `varerror(1)`.

`contrast(contrastspec)` specifies a contrast for group means containing *J* contrast coefficients that must sum to zero. *contrastspec* is

```
#1 #2 [#3 ...#J] [ , null(numlist) onesided ]
```

`null(numlist)` specifies the null or hypothesized value of the mean contrast. The default is `null(0)`.

`onesided` requests a one-sided *t* test. The default is *F* test.

`parallel`; see [PSS] **power**.

### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

### Graph

`graph`, `graph()`; see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

### Iteration

`init(#)` specifies the initial value of the sample size for the sample-size determination or the initial value of the effect size  $\delta$  for the effect-size determination. The default uses a bisection algorithm to bracket the solution.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power oneway` but is not shown in the dialog box:

`notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power oneway*

*Alternative ways of specifying effect*

*Computing sample size*

*Computing power*

*Computing effect size and between-group variance*

*Testing hypotheses about multiple group means*

*Video examples*

This entry describes the **power oneway** command and the methodology for power and sample-size analysis for one-way ANOVA. See [\[PSS\] intro](#) for a general introduction to power and sample-size analysis and [\[PSS\] power](#) for a general introduction to the **power** command using hypothesis tests.

## Introduction

The comparison of multiple group means using one-way ANOVA models is a commonly used approach in a wide variety of statistical studies. The term “one way” refers to a single factor containing an arbitrary number of groups or levels. In what follows, we will assume that the factor levels are fixed. For two groups, the ANOVA model is equivalent to an unpaired two-sample  $t$  test; see [\[PSS\] power twomeans](#) for the respective power and sample-size analysis. One-way ANOVA uses an  $F$  test based on the ratio of the between-group variance to the within-group variance to compare means of multiple groups.

For example, consider a type of drug with three levels of dosage in treating a medical condition. An investigator may wish to test whether the mean response of the drug is the same across all levels of dosage. This is equivalent to testing the null hypothesis  $H_0: \mu_1 = \mu_2 = \mu_3$  versus the alternative hypothesis  $H_a: \mu_1 \neq \mu_2$  or  $\mu_1 \neq \mu_3$  or  $\mu_2 \neq \mu_3$ ; that is, at least one of the three group means is different from all the others. Rejection of the null hypothesis, however, does not provide any specific information about the individual group means. Therefore, in some cases, investigators may want to form a hypothesis for a mean contrast,  $c = \sum_{j=1}^k c_j \mu_j$ , a linear combination of group means where weights  $c_j$  sum to zero, and compare individual means by testing a hypothesis  $H_0: c = c_0$  versus  $H_a: c \neq c_0$ .

This entry describes power and sample-size analysis for the inference about multiple population means using hypothesis testing based on one-way ANOVA. Specifically, we consider the null hypothesis  $H_0: \mu_1 = \dots = \mu_J$ , which tests the equality of  $J$  group means against the alternative hypothesis  $H_a: \mu_i \neq \mu_j$  for some  $i, j$ . The test statistic for this hypothesis is based on the ratio of the between-group variance to the within-group variance and has an  $F$  distribution under the null hypothesis. The corresponding test is known as an overall  $F$  test, which tests the equality of multiple group means. This test is nondirectional.

For testing a single mean contrast,  $H_0: c = c_0$  versus  $H_a: c \neq c_0$ , a test statistic is a function of the ratio of the contrast variance to the error (or within-group) variance, and either an  $F$  test or a  $t$  test can be used for a two-sided alternative. For a one-sided alternative,  $H_a: c > c_0$  or  $H_a: c < c_0$ , only a  $t$  test can be used.

Power and sample-size computations use the distribution of the test statistic under the alternative hypothesis, which is a noncentral  $F$  distribution for the considered tests. Power is a function of the noncentrality parameter, and the noncentrality parameter is a function of the ratio of the standard deviation of the tested effect to the standard deviation of the errors. As such, the effect size for the overall  $F$  test is defined as the square root of the ratio of the between-group variance to the

within-group variance. For testing a mean contrast, the effect size is defined as the square root of the contrast variance to the error or within-group variance.

## Using power oneway

`power oneway` computes sample size, power, or effect size and target variance of the effect for a one-way fixed-effects analysis of variance. All computations are performed assuming a significance level of 0.05. You may change the significance level by specifying the `alpha()` option.

By default, the computations are performed for an overall  $F$  test testing the equality of all group means. The within-group or error variance for this test is assumed to be 1 but may be changed by specifying the `varerror()` option.

To compute the total sample size, you must specify the alternative *meanspec* and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the total sample size in the `n()` option and the alternative *meanspec*.

Instead of the alternative group means, you can specify the number of groups in the `ngroups()` option and the variance of the group means (or the between-group variance) in the `varmeans()` option when computing sample size or power.

To compute effect size, the square root of the ratio of the between-group variance to the error variance, and the target between-group variance, you must specify the total sample size in the `n()` option, the power in the `power()` option, and the number of groups in the `ngroups()` option.

To compute sample size or power for a test of a mean contrast, in addition to the respective options `power()` or `n()` as described above, you must specify the alternative *meanspec* and the corresponding contrast coefficients in the `contrast()` option. A contrast coefficient must be specified for each of the group means, and the specified coefficients must sum to zero. The null value for the specified contrast is assumed to be zero but may be changed by specifying the `null()` suboption within `contrast()`. The default test is an  $F$  test. You can instead request a one-sided  $t$  test by specifying the `onesided` suboption within `contrast()`. Effect-size determination is not available when testing a mean contrast.

For all the above computations, the error or within-group variance is assumed to be 1. You can change this value by specifying the `varerror()` option.

By default, all computations assume a balanced- or equal-allocation design. You can use the `grweights()` option to specify an unbalanced design for power, sample-size, or effect-size computations. For power and effect-size computations, you can specify individual group sizes in options `n1()`, `n2()`, and so on instead of a combination of `n()` and `grweights()` to accommodate an unbalanced design. For a balanced design, you can also specify the `npergroup()` option to specify a group size instead of a total sample size in `n()`.

In a one-way analysis of variance, sample size and effect size depend on the noncentrality parameter of the  $F$  distribution, and their estimation requires iteration. The default initial values are obtained from a bisection search that brackets the solution. If you desire, you may change this by specifying your own value in the `init()` option. See [PSS] **power** for the descriptions of other options that control the iteration procedure.

**Alternative ways of specifying effect**

To compute power or sample size, you must specify the magnitude of the effect desired to be detected by the test. With `power oneway`, you can do this by specifying either the individual alternative *meanspec*, for example,

```
power oneway  $m_1$   $m_2$  ...  $m_J$ , ...
```

or the variance of  $J$  group means (between-group variance) and the number of groups  $J$ :

```
power oneway, varmeans(#) ngroups(#) [...]
```

You can also specify multiple values for variances in `varmeans()`.

There are multiple ways in which you can supply the group means to `power oneway`.

As we showed above, you may specify group means following the command line as

```
power oneway  $m_1$   $m_2$  ...  $m_J$ , ...
```

At least two means must be specified.

Alternatively, you can define a Stata matrix as a row or a column vector, and use it with `power oneway`. The dimension of the Stata matrix must be at least 2. For example,

```
matrix define meanmat = ( $m_1$ ,  $m_2$ , ...,  $m_J$ )
power oneway meanmat, ...
```

You can also specify multiple values or a *numlist* for each of the group means in parentheses:

```
power oneway ( $m_{1,1}$   $m_{1,2}$  ...  $m_{1,K_1}$ ) ( $m_{2,1}$   $m_{2,2}$  ...  $m_{2,K_2}$ ) ..., ...
```

Each of the *numlists* may contain different numbers of values,  $K_1 \neq K_2 \neq \dots \neq K_J$ . `power oneway` will produce results for all possible combinations of values across *numlists*. If instead you would like to treat each specification as a separate scenario, you can specify the `parallel` option.

Similarly, you can accommodate multiple sets of group means in a matrix form by adding a row for each specification. The columns of a matrix with multiple rows correspond to  $J$  group means, and values within each column  $j$  correspond to multiple specifications of the  $j$ th group mean or a *numlist* for the  $j$ th group mean.

For example, the following two specifications are the same:

```
power oneway ( $m_{1,1}$   $m_{1,2}$   $m_{1,3}$ ) ( $m_{2,1}$   $m_{2,2}$   $m_{2,3}$ ), ...
```

and

```
matrix define meanmat = ( $m_{1,1}$ ,  $m_{2,1}$  \  $m_{1,2}$ ,  $m_{2,2}$  \  $m_{1,3}$ ,  $m_{2,3}$ )
power oneway meanmat, ...
```

In the above specification, if you wish to specify a *numlist* only for the first group, you may define your matrix as

```
matrix define meanmat = ( $m_{1,1}$ ,  $m_{2,1}$  \  $m_{1,2}$ , . \  $m_{1,3}$ , .)
```

and the results of

```
power oneway meanmat, ...
```

will be the same as the results of

```
power oneway ( $m_{1,1}$   $m_{1,2}$   $m_{1,3}$ )  $m_{2,1}$ , ...
```

If you wish to treat the rows of *meanmat* as separate scenarios, you must specify the `parallel` option.

In the following sections, we describe the use of `power oneway` accompanied by examples for computing sample size, power, and effect size.

## Computing sample size

To compute sample size, you must specify the alternative group means or their variance and, optionally, the power of the test in the `power()` option. A power of 0.8 is assumed if `power()` is not specified.

### ▷ Example 1: Sample size for a one-way ANOVA

Consider an example from [van Belle, Fisher, Heagerty, and Lumley \(2004, 367\)](#), where the authors report the results of a study of the association between cholesterol level and the number of diseased blood vessels, which indicate the presence of coronary artery disease, in patients undergoing coronary bypass surgery. Suppose we wish to plan a similar new study in which a cholesterol level is considered a risk factor that may be associated with the number of diseased blood vessels, our grouping variable. We consider three groups of subjects with 1, 2, or 3 numbers of diseased blood vessels. We would like to know how many subjects we need to observe in each group to detect differences between cholesterol levels across groups. Our projected cholesterol levels in three groups are 260, 289, and 295 mg/dL, respectively. Suppose that we anticipate the within-group or error variance is 4,900. To compute the total sample size for `power oneway`'s default setting of a balanced design with 5% significance level and 80% power, we type

```
. power oneway 260 289 295, varerror(4900)
Performing iteration ...
Estimated sample size for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.2183
      N_g =      3
      m1 =   260.0000
      m2 =   289.0000
      m3 =   295.0000
      Var_m =   233.5556
      Var_e =  4900.0000

Estimated sample sizes:
      N =      207
      N per group =    69
```

We find that a total sample of 207 subjects, 69 subjects per group, is required to detect a change in average cholesterol levels in at least 1 of the 3 groups for this study.

In addition to the specified and implied study parameters, `power oneway` reports the value of the effect size,  $\delta = \sqrt{233.5556/4900} = 0.2183$ , computed as a square root of the ratio between the variance of the group means `Var_m` and the error variance `Var_e`. The two variances are also often referred to as the between-group variance and the within-group variance, respectively. The effect size  $\delta$  provides a unit-less measure of the magnitude of an effect with a lower bound of zero meaning

no effect. It corresponds to Cohen's effect-size measure  $f$  (Cohen 1988). Cohen's convention is that  $f = 0.1$  means small effect size,  $f = 0.25$  means medium effect size, and  $f = 0.4$  means large effect size. According to this convention, the effect size considered in our example is medium.



### ► Example 2: Alternative ways of specifying effect

Instead of specifying the alternative means as in [example 1](#), we can specify the variance between them and the number of groups. From [example 1](#), the variance between the group means was computed to be 233.5556. We specify this value in `varmeans()` as well as the number of groups in `ngroups()`:

```
. power oneway, varmeans(233.5556) ngroups(3) varerror(4900)
Estimated sample size for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.2183
      N_g =      3
      Var_m = 233.5556
      Var_e = 4900.0000
Estimated sample sizes:
      N =      207
      N per group =    69
```

We obtain the exact same results as in [example 1](#).

Instead of specifying alternative means directly following the command line, we can define a matrix, say, `means`, containing these means and use it with `power oneway`:

```
. matrix define means = (260,289,295)
. power oneway means, varerror(4900)
(output omitted)
```

You can verify that the results are identical to the previous results.



### ► Example 3: Computing sample size for a mean contrast

Continuing [example 1](#), suppose we would like to test whether the average of the first two cholesterol levels is different from the cholesterol level of the third group. We construct the following contrast  $(0.5, 0.5, -1)$  to test this hypothesis. To compute sample size, we specify the contrast coefficients in the `contrast()` option:

```
. power oneway 260 289 295, varerror(4900) contrast(.5 .5 -1)
Performing iteration ...
Estimated sample size for one-way ANOVA
F test for contrast of means
Ho: Cm = 0 versus Ha: Cm != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.1381
    N_g =      3
    m1 =    260.0000
    m2 =    289.0000
    m3 =    295.0000
    C*m =   -20.5000
    c0 =      0.0000
    Var_Cm =   93.3889
    Var_e =  4900.0000
Estimated sample sizes:
    N =      414
    N per group =   138
```

The required sample size to achieve this study objective is 414 with 138 subjects per group.

For a test of a mean contrast, we can also test a directional hypothesis by specifying the `onesided` option within `contrast()`. In this case, the computation is based on the  $t$  test instead of the  $F$  test.

For example, to test whether the average of the first two cholesterol levels is less than the cholesterol level of the third group, we type

```
. power oneway 260 289 295, varerror(4900) contrast(.5 .5 -1, onesided)
Performing iteration ...
Estimated sample size for one-way ANOVA
t test for contrast of means
Ho: Cm = 0 versus Ha: Cm < 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.1381
    N_g =      3
    m1 =    260.0000
    m2 =    289.0000
    m3 =    295.0000
    C*m =   -20.5000
    c0 =      0.0000
    Var_Cm =   93.3889
    Var_e =  4900.0000
Estimated sample sizes:
    N =      327
    N per group =   109
```

The results show that the required sample size reduces to a total of 327 subjects for this lower one-sided hypothesis.

The default null value for the contrast is zero, but you can change this by specifying `contrast()`'s suboption `null()`.

### ▶ Example 4: Unbalanced design

Continuing [example 1](#), suppose we anticipate that the first group will have twice as many subjects as the second and the third groups. We can accommodate this unbalanced design by specifying the corresponding group weights in the `grweights()` option:

```
. power oneway 260 289 295, varerror(4900) grweights(2 1 1)
Performing iteration ...
Estimated sample size for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.2306
      N_g    =         3
      m1     = 260.0000
      m2     = 289.0000
      m3     = 295.0000
      Var_m  = 260.5000
      Var_e  = 4900.0000
Estimated sample sizes:
      N      =    188
Average N   =   62.6667
      N1     =     94
      N2     =     47
      N3     =     47
```

The required total sample size for this unbalanced design is 188 with 94 subjects in the first group and 47 subjects in the second and third groups. The average number of subjects per group is 62.67.

We can compute results for multiple sets of group weights. The specification of group weights within `grweights()` is exactly the same as the specification of group means described in [Alternative ways of specifying effect](#). Suppose that we would like to compute sample sizes for two unbalanced designs. The first design has twice as many subjects in the first group as the other two groups. The second design has the first two groups with twice as many subjects as the third group. We specify multiple group weights for the first and second groups in parentheses. We also specify the `parallel` option to treat multiple weight values in parallel instead of computing results for all possible combinations of these values, which would have been done by default.

```
. local tabcols alpha power N N1 N2 N3 grwgt1 grwgt2 grwgt3 Var_m Var_e
. power oneway 260 289 295, varerror(4900) grweights((2 2) (1 2) 1) parallel
> table('tabcols', formats("%6.0g"))
Performing iteration ...
Estimated sample size for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
```

alpha	power	N	N1	N2	N3	grwgt1	grwgt2	grwgt3	Var_m	Var_e
.05	.8	188	94	47	47	2	1	1	260.5	4900
.05	.8	205	82	82	41	2	2	1	235.4	4900

The default table does not include group weights, so we request a table with custom columns containing group weights via `table()`. We also request a smaller format to make the table more compact.



## Computing power

To compute power, you must specify the total sample size in the `n()` option and the alternative group means or their variance.

### ▷ Example 5: Power for a one-way ANOVA

Continuing [example 1](#), suppose that we anticipate obtaining a total sample of 300 subjects. To compute the corresponding power, we specify the sample size of 300 in `n()`:

```
. power oneway 260 289 295, n(300) varerror(4900)
Estimated power for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
        N =     300
N per group =     100
      delta =    0.2183
        N_g =      3
        m1 = 260.0000
        m2 = 289.0000
        m3 = 295.0000
      Var_m = 233.5556
      Var_e = 4900.0000

Estimated power:
      power =    0.9308
```

The power increases to 93.08% for a larger sample of 300 subjects.

◀

### ▷ Example 6: Multiple values of study parameters

Continuing [example 5](#), we may want to check powers for several sample sizes. We simply list multiple sample-size values in `n()`:

```
. power oneway 260 289 295, n(100 200 300) varerror(4900)
> table(, labels(N_per_group "N/N_g") formats("%6.2g"))
Estimated power for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
```

alpha	power	N	N/N_g	delta	N_g	m1	m2	m3	Var_m	Var_e
.05	.47	100	33	.22	3	260	289	295	234	4900
.05	.78	200	66	.22	3	260	289	295	234	4900
.05	.93	300	100	.22	3	260	289	295	234	4900

To shorten our default table, we specified a shorter label for the `N_per_group` column and reduced the default display format for all table columns by specifying the corresponding options within the `table()` option.

We can compute results for multiple values of group means. For example, to see how power changes when the first group mean takes values of 245, 260, and 280, we specify these values in parentheses:

```
. power oneway (245 260 280) 289 295, n(300) varerror(4900)
> table(, labels(N_per_group "N/N_g") formats("%6.2g"))

Estimated power for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
```

alpha	power	N	N/N_g	delta	N_g	m1	m2	m3	Var_m	Var_e
.05	1	300	100	.32	3	245	289	295	497	4900
.05	.93	300	100	.22	3	260	289	295	234	4900
.05	.25	300	100	.088	3	280	289	295	38	4900

We can compute results for a combination of multiple sample sizes and multiple mean values or a combination of multiple values of other study parameters.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## Computing effect size and between-group variance

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, power, sample size, and the number of groups must be specified.

The effect size is defined as a square root of the ratio of the variance of the tested effect, for example, the between-group variance, to the error variance. Both the effect size and the target between-group variance are computed.

The effect-size determination is not available for testing a mean contrast.

### ► Example 7: Effect size for a one-way analysis of variance

Continuing [example 5](#), we now want to compute the effect size that can be detected for a sample of 300 subjects and a power of 80%. We specify both parameters in the respective options. For the effect-size determination, we must also specify the number of groups in `ngroups()`:

```
. power oneway, varerror(4900) n(300) power(0.80) ngroups(3)
Performing iteration ...
Estimated between-group variance for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0

Study parameters:
      alpha =    0.0500
      power =    0.8000
         N  =    300
N per group =    100
       N_g  =     3
      Var_e = 4900.0000

Estimated effect size and between-group variance:
      delta =    0.1801
      Var_m = 158.9648
```

For a larger sample size, given the same power, we can detect a smaller effect size, 0.18, compared with the effect size of 0.22 from [example 1](#). The corresponding estimate of the between-group variance is 158.96, given the error variance of 4,900.

◀

## Testing hypotheses about multiple group means

There are several ways in which you can compare group means of a single factor on the collected data. Two commonly used commands to do this are `oneway` (or `anova`) and `contrast`. We demonstrate a quick use of these commands here using the systolic blood pressure example; see [\[R\] oneway](#) and [\[R\] contrast](#) for more examples. Also see [\[R\] anova](#) for general ANOVA models.

### ▶ Example 8: One-way ANOVA

Consider `systolic.dta` containing 58 patients undergoing 4 different drug treatments for reducing systolic blood pressure. The `systolic` variable records the change in systolic blood pressure and the `drug` variable records four treatment levels. We would like to test whether the average change in systolic blood pressure is the same for all treatments. We use `oneway` to do this:

```
. use http://www.stata-press.com/data/r14/systolic
(Systolic Blood Pressure Data)
. oneway systolic drug
```

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	3133.23851	3	1044.41284	9.09	0.0001
Within groups	6206.91667	54	114.942901		
Total	9340.15517	57	163.862371		

Bartlett's test for equal variances:  $\chi^2(3) = 1.0063$  Prob> $\chi^2 = 0.800$

We reject the null hypothesis that all treatment means are equal at the 5% significance level; the  $p$ -value is less than 0.0001.

Suppose we wish to design a new similar study. We use the estimates from this study to perform a sample-size analysis for our new study. First, we estimate the means of systolic blood pressure for different treatment levels:

```
. mean systolic, over(drug)
```

Mean estimation	Number of obs	=	58
1: drug = 1			
2: drug = 2			
3: drug = 3			
4: drug = 4			

Over	Mean	Std. Err.	[95% Conf. Interval]	
systolic				
1	26.06667	3.014989	20.02926	32.10408
2	25.53333	2.999788	19.52636	31.54031
3	8.75	2.892323	2.958224	14.54178
4	13.5	2.330951	8.832351	18.16765

From the `oneway` output, the estimate of the error variance is roughly 115. Second, we specify the means and the error variance with `power oneway` and compute the required sample size for a balanced design assuming 5% significance level and 90% power.

```

. power oneway 26.07 25.53 8.75 13.5, vareerror(115) power(.9)
Performing iteration ...
Estimated sample size for one-way ANOVA
F test for group effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha = 0.0500
    power = 0.9000
    delta = 0.7021
    N_g = 4
    m1 = 26.0700
    m2 = 25.5300
    m3 = 8.7500
    m4 = 13.5000
    Var_m = 56.6957
    Var_e = 115.0000
Estimated sample sizes:
    N = 36
    N per group = 9

```

The effect size for this ANOVA model specification is rather large, 0.7021. So we need only 36 subjects, 9 per group, to detect the effect of this magnitude with 90% power.

Suppose that in addition to testing the overall equality of treatment means, we are interested in testing a specific hypothesis of whether the average of the first two treatment means is equal to the average of the last two treatment means.

To perform this test on the collected data, we can use the `contrast` command. The `contrast` command is not available after `oneway`, so we repeat our one-way ANOVA analysis using the `anova` command before using `contrast`; see [R] [contrast](#) for details.

```

. anova systolic i.drug

```

Source	Partial SS	df	MS	F	Prob>F
Model	3133.2385	3	1044.4128	9.09	0.0001
drug	3133.2385	3	1044.4128	9.09	0.0001
Residual	6206.9167	54	114.9429		
Total	9340.1552	57	163.86237		

```

. contrast {drug .5 .5 -.5 -.5}
Contrasts of marginal linear predictions
Margins      : asbalanced

```

	df	F	P>F
drug	1	26.85	0.0000
Denominator	54		

```


```

	Contrast	Std. Err.	[95% Conf. Interval]	
drug (1)	14.675	2.832324	8.996533	20.35347

As with the overall equality test, we find statistical evidence to reject this hypothesis as well.

To compute the required sample size for this hypothesis, we specify the contrast coefficients in the `contrast()` option of `power oneway`:

```
. power oneway 26.07 25.53 8.75 13.5, varerror(115) power(.9)
> contrast(.5 .5 -.5 -.5)
Performing iteration ...
Estimated sample size for one-way ANOVA
F test for contrast of means
Ho: Cm = 0 versus Ha: Cm != 0
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =    0.6842
    N_g =      4
    m1 =    26.0700
    m2 =    25.5300
    m3 =     8.7500
    m4 =    13.5000
    C*m =    14.6750
    c0 =     0.0000
    Var_Cm =   53.8389
    Var_e =  115.0000
Estimated sample sizes:
    N =      28
    N per group =    7
```

The required sample size is 28 subjects with 7 subjects per group, which is smaller than the required sample size computed earlier for the overall test of the equality of means.

◀

## Video examples

[Sample-size calculation for one-way analysis of variance](#)

[Power calculation for one-way analysis of variance](#)

[Minimum detectable effect size for one-way analysis of variance](#)

## Stored results

`power oneway` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N_avg)</code>	average sample size
<code>r(N#)</code>	number of subjects in group #
<code>r(N_per_group)</code>	number of subjects per group
<code>r(N_g)</code>	number of groups
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(balanced)</code>	1 for a balanced design; 0 otherwise
<code>r(grwgt#)</code>	group weight #
<code>r(onesided)</code>	1 for a one-sided test of a mean contrast; 0 otherwise
<code>r(m#)</code>	group mean #
<code>r(Cm)</code>	mean contrast
<code>r(c0)</code>	null mean contrast
<code>r(Var_m)</code>	group-means (between-group) variance
<code>r(Var_Cm)</code>	contrast variance
<code>r(Var_e)</code>	error (within-group) variance
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or effect size
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>oneway</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrices

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Consider a single factor  $A$  with  $J$  groups or levels, where each level comprises  $n_j$  observations for  $j = 1, \dots, J$ . The total number of observations is  $n = \sum_{j=1}^J n_j$ . Let  $Y_{ij}$  denote the response for the  $j$ th level of the  $i$ th individual and  $\mu_j$  denote the factor-level or group means. Individual responses from  $J$  populations are assumed to be normally distributed with mean  $\mu_j$  and a constant variance  $\sigma_e^2$ .

The hypothesis of an overall  $F$  test of the equality of group means is

$$H_0: \mu_1 = \dots = \mu_J$$

versus

$$H_a: \mu_j\text{'s are not all equal} \quad (1)$$

The hypothesis for a test of a mean contrast is

$$H_0: \sum_{j=1}^J c_j \mu_j = c_0 \quad (2)$$

versus a two-sided  $H_a: \sum_{j=1}^J c_j \mu_j \neq c_0$ , upper one-sided  $H_a: \sum_{j=1}^J c_j \mu_j > c_0$ , and lower one-sided  $H_a: \sum_{j=1}^J c_j \mu_j < c_0$ , where  $c_j$ 's are contrast coefficients such that  $\sum_{i=1}^J c_j = 0$  and  $c_0$  is a null value of a mean constant. The two-sided hypothesis is tested using an  $F$  test, and one-sided hypotheses are tested using a  $t$  test.

## Computing power

Hypotheses (1) and (2) can be tested in a general linear model framework. Consider a linear model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\epsilon}$$

where  $\mathbf{y}$  is an  $n \times 1$  vector of observations,  $\mathbf{X}$  is an  $n \times p$  matrix of predictors,  $\mathbf{b}$  is a  $p \times 1$  vector of unknown and fixed coefficients, and  $\boldsymbol{\epsilon}$  is an  $n \times 1$  vector of error terms that are independent and identically distributed as  $N(0, \sigma_e^2)$ .

For the one-way model,  $p = J$  and the contents of  $\mathbf{b} = (b_1, \dots, b_J)'$  are the  $\mu_j$ ,  $j = 1, \dots, J$ .

A general linear hypothesis in this framework is given by

$$\mathbf{C}\mathbf{b} = \mathbf{c}_0$$

where  $\mathbf{C}$  is a  $\nu \times p$  matrix with  $\text{rank}(\mathbf{C}) = \nu \leq p$ , and  $\mathbf{c}_0$  is a vector of constants. For an overall test of the means in (1),  $\mathbf{c}_0 = 0$ . The estimates of  $\mathbf{b}$  and  $\sigma_e^2$ , respectively, are given by

$$\begin{aligned} \hat{\mathbf{b}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \hat{\sigma}_e^2 &= (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})/(n - p) \end{aligned}$$

A general test statistic for testing hypotheses (1) and (2) is given by

$$F_C = \frac{SS_C}{(p - 1)\hat{\sigma}_e^2} \quad (3)$$

where  $SS_C = (\mathbf{C}\hat{\mathbf{b}} - \mathbf{c}_0)' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \}^{-1} (\mathbf{C}\hat{\mathbf{b}} - \mathbf{c}_0)$ . Let  $\alpha$  be the significance level,  $\beta$  be the probability of a type II error, and  $F_{p-1, n-p, 1-\alpha}$  be the  $(1 - \alpha)$ th quantile of an  $F$  distribution with  $p - 1$  numerator and  $n - p$  denominator degrees of freedom. We reject the null hypothesis if we observe a statistic  $F_C > F_{p-1, n-p, 1-\alpha}$ .

The test statistic in (3) under the alternative hypothesis is distributed as a noncentral  $F$  distribution with  $p - 1$  numerator and  $n - p$  denominator degrees of freedom with a noncentrality parameter  $\lambda$  given by

$$\begin{aligned} \lambda &= (\mathbf{C}\mathbf{b} - \mathbf{c}_0)' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \}^{-1} (\mathbf{C}\mathbf{b} - \mathbf{c}_0) / \sigma_e^2 \\ &= n(\mathbf{C}\mathbf{b} - \mathbf{c}_0)' \{ \mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}' \}^{-1} (\mathbf{C}\mathbf{b} - \mathbf{c}_0) / \sigma_e^2 \\ &= n\delta^2 \end{aligned}$$

where the matrix  $\ddot{\mathbf{X}}$  contains the unique rows of  $\mathbf{X}$  and  $\mathbf{W} = \text{diag}(w_1, \dots, w_p)$ . We define  $\delta$  as the effect size.

For the one-way design, the dimension of  $\ddot{\mathbf{X}}$  is  $J \times J$ . The weights are formally  $w_j = n_j/n$  but can also be expressed in terms of the group weights (specified in `grweights()`), normalized by the sum of the group weights, making  $w_j$  independent of  $n$ . Specifically, let the group weights be denoted  $\tilde{w}_j$ , then define a cell sample size multiplier as  $n_c = n/\sum_j \tilde{w}_j$  so that  $n_j = n_c \tilde{w}_j$ . The cell-means parameterization simplifies  $\ddot{\mathbf{X}}$  to the identity matrix,  $\mathbf{I}_J$ .

See O'Brien and Muller (1993) for details.

The power of the overall  $F$  test in (1) is given by

$$1 - \beta = F_{\nu, n-p, \lambda} (F_{\nu, n-p, 1-\alpha}) \quad (4)$$

where  $F_{\cdot, \cdot, \lambda}(\cdot)$  is the cdf of a noncentral  $F$  distribution.

Total sample size and effect size are obtained by iteratively solving the nonlinear equation (4). When the `grweights()` option is specified, a constant multiplier  $n_c$  is computed and rounded to an integer unless the `nfractional` option is specified. The group sizes are then computed as  $\tilde{w}_j n_c$ . The actual sample size, `N_a`, is the sum of the group sizes.

See Kutner et al. (2005) for details.

The power of the test (2) for a mean contrast is given by

$$1 - \beta = \begin{cases} 1 - T_{n-1, \tilde{\lambda}}(t_{n-1, 1-\alpha}) & \text{for an upper one-sided test} \\ T_{n-1, \tilde{\lambda}}(-t_{n-1, 1-\alpha}) & \text{for a lower one-sided test} \\ F_{1, n-p, \lambda}(F_{1, n-p, 1-\alpha}) & \text{for a two-sided test} \end{cases} \quad (5)$$

where  $T_{\cdot, \tilde{\lambda}}(\cdot)$  is the cumulative of a noncentral Student's  $t$  distribution with the noncentrality parameter  $\tilde{\lambda}$  given by

$$\begin{aligned} \tilde{\lambda} &= \sqrt{n}(\mathbf{C}\mathbf{b} - \mathbf{c}_0)' \sqrt{\{\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}'\}^{-1} / \sigma_e^2} \\ &= \sqrt{n}\tilde{\delta} \end{aligned}$$

Sample size is obtained by iteratively solving the nonlinear equation (5).

## References

- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied Linear Statistical Models*. 5th ed. New York: McGraw-Hill/Irwin.
- O'Brien, R. G., and K. E. Muller. 1993. Unified power analysis for  $t$ -tests through multivariate hypotheses. In *Applied Analysis of Variance in Behavioral Science*, ed. L. K. Edwards, 297–344. New York: Dekker.
- van Belle, G., L. D. Fisher, P. J. Heagerty, and T. S. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences*. 2nd ed. New York: Wiley.



## Also see

[PSS] [power](#) — Power and sample-size analysis for hypothesis tests

[PSS] [power repeated](#) — Power analysis for repeated-measures analysis of variance

[PSS] [power twomeans](#) — Power analysis for a two-sample means test

[PSS] [power twoway](#) — Power analysis for two-way analysis of variance

[PSS] [power, graph](#) — Graph results from the power command

[PSS] [power, table](#) — Produce table of results from the power command

[PSS] [Glossary](#)

[R] [anova](#) — Analysis of variance and covariance

[R] [contrast](#) — Contrasts and linear hypothesis tests after estimation

[R] [oneway](#) — One-way analysis of variance

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power twoway` computes sample size, power, or effect size for two-way analysis of variance (ANOVA). By default, it computes sample size for given power and effect size. Alternatively, it can compute power for given sample size and effect size or compute effect size for given sample size, power, and number of cells. You can choose between testing for main row or column effect or their interaction. Also see [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for the main effect of the row factor for a  $2 \times 3$  design specified using cell means with a within-cell variance of 27 and default power of 0.8 and significance level  $\alpha = 0.05$

```
power twoway 19 18 32 \ 23 25 26, varerror(27)
```

Same as above, but specify cell means in the matrix `cm`

```
matrix cm = (19, 18, 32 \ 23, 25, 26)
power twoway cm, varerror(27)
```

Same as above

```
power twoway cm, varerror(27) factor(row)
```

As above, but calculate sample size for the main effect of the column factor

```
power twoway cm, varerror(27) factor(column)
```

As above, but calculate sample size for the interaction of the row and column factors

```
power twoway cm, varerror(27) factor(rowcol)
```

As above, but for within-cell variances of 20, 25, 30, and 35

```
power twoway cm, varerror(20(5)35) factor(rowcol)
```

Sample size for the row factor with power of 0.85 and  $\alpha = 0.01$

```
power twoway cm, varerror(27) power(.85) alpha(.01)
```

Specify that the groups in the second row have twice the sample size as those in the first row

```
power twoway cm, varerror(27) cellweights(1 1 1\2 2 2) showcellsizes
```

Sample size when variance of the main effect of the row factor equals 1.2 in a  $2 \times 3$  design

```
power twoway, varerror(27) factor(row) vareffect(1.2) ///
nrows(2) ncols(3)
```

Power for the main effect of the row factor with a sample size of 180 for cell means stored in matrix `cm`

```
power twoway cm, varerror(27) factor(row) n(180)
```

Same as above, but specify that the sample size per cell is 40

```
power twoway cm, varerror(27) factor(row) npercell(40)
```

As above, but specify cell sample sizes of 40, 45, 50, and 55

```
power twoway cm, varerror(27) factor(row) npercell(40(5)55)
```

As above, but display results as a graph of power versus sample size

```
power twoway cm, varerror(27) factor(row) npercell(40(5)55) graph
```

Effect size and target between-group variance for three groups, sample size of 150, and power of 0.8

```
power twoway, varerror(27) nrows(2) ncols(3) n(150) power(.8)
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power twoway meanspec [ , power(numlist) options ]
```

Compute power

```
power twoway meanspec, n(numlist) [options]
```

Compute effect size and target effect variance

```
power twoway, n(numlist) power(numlist) nrows(#) ncols(#) [options]
```

where *meanspec* is either a matrix *matname* containing cell means or individual cell means in a matrix form:

$$m_{1,1} \ m_{1,2} \ [\dots] \ \backslash \ m_{2,1} \ m_{2,2} \ [\dots] \ [\ \backslash \dots \backslash \ m_{J,1} \ \dots \ m_{J,K} ]$$

$m_{jk}$ , where  $j = 1, 2, \dots, J$  and  $k = 1, 2, \dots, K$ , is the alternative cell mean or the cell mean of the  $j$ th row and  $k$ th column under the alternative hypothesis.

*matname* is the name of a Stata matrix with  $J$  rows and  $K$  columns containing values of alternative cell means.

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample sizes
* <u>npercell</u> ( <i>numlist</i> )	number of subjects per cell; implies balanced design
<u>cellweights</u> ( <i>wgtspec</i> )	cell weights; default is one for each cell, meaning equal cell sizes
<u>nrows</u> (#)	number of rows
<u>ncols</u> (#)	number of columns
<u>factor</u> (row   column   rowcol)	tested effect
* <u>vareffect</u> ( <i>numlist</i> )	variance explained by the tested effect in <code>factor()</code>
* <u>varrow</u> ( <i>numlist</i> )	variance explained by the row effect; synonym for <code>factor(row)</code> and <code>vareffect(numlist)</code>
* <u>varcolumn</u> ( <i>numlist</i> )	variance explained by the column effect; synonym for <code>factor(column)</code> and <code>vareffect(numlist)</code>
* <u>varrowcolumn</u> ( <i>numlist</i> )	variance explained by the row–column interaction effect; synonym for <code>factor(rowcol)</code> and <code>vareffect(numlist)</code>
* <u>varerror</u> ( <i>numlist</i> )	error variance; default is <code>varerror(1)</code>
<u>showmatrices</u>	display cell means and sample sizes as matrices
<u>showmeans</u>	display cell means
<u>showcellsizes</u>	display cell sizes
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<u>[no]</u> <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<u>saving</u> ( <i>filename</i> [ , replace ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<u>init</u> (#)	initial value for the sample size or the effect size; default is to use a bisection algorithm to bound the solution
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
<u>[no]</u> <u>log</u>	suppress or display iteration log
<u>[no]</u> <u>dots</u>	suppress or display iterations as dots
<u>notitle</u>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 [numlist](#).

notitle does not appear in the dialog box.

<i>wgtspec</i>	Description
$\#_{1,1} \dots \#_{1,K} \setminus \dots \setminus \#_{J,1} \dots \#_{J,K}$	$J \times K$ cell weights; weights must be positive and must be integers unless option <code>nfractional</code> is specified
<i>matname</i>	$J \times K$ matrix containing cell weights

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N_per_cell</code>	number of subjects per cell	$N/N_{rc}$
<code>N_avg</code>	average number of subjects per cell	$N_{avg}$
<code>N#1_#2</code>	number of subjects in cell ( $\#_1, \#_2$ )	$N_{\#_1, \#_2}$
<code>delta</code>	effect size	$\delta$
<code>N_rc</code>	number of cells	$N_{rc}$
<code>N_r</code>	number of rows	$N_r$
<code>N_c</code>	number of columns	$N_c$
<code>m#1_#2</code>	cell mean ( $\#_1, \#_2$ )	$\mu_{\#_1, \#_2}$
<code>Var_r</code>	variance explained by the row effect	$\sigma_r^2$
<code>Var_c</code>	variance explained by the column effect	$\sigma_c^2$
<code>Var_rc</code>	variance explained by the row–column interaction	$\sigma_{rc}^2$
<code>Var_e</code>	error variance	$\sigma_e^2$
<code>cwgt#1_#2</code>	cell weight ( $\#_1, \#_2$ )	$w_{\#_1, \#_2}$
<code>target</code>	target parameter; synonym for target effect variance	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Column `N_per_cell` is available and is shown in the default table only for balanced designs.

Column `N_avg` is shown in the default table only for unbalanced designs.

Columns `N#1_#2`, `N_rc`, `m#1_#2`, and `cwgt#1_#2` are not shown in the default table.

## Options

Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] [power](#).

`npercell` (*numlist*) specifies the cell size. Only positive integers are allowed. This option implies a balanced design. `npercell()` cannot be specified with `n()` or `cellweights()`.

`cellweights(wgtspec)` specifies  $J \times K$  cell weights for an unbalanced design. The weights must be positive and must also be integers unless the `nfractional` option is specified. `cellweights()` cannot be specified with `npercell()`.

`nrows(#)` specifies the number of rows or the number of levels of the row factor in a two-way ANOVA. At least two rows must be specified. This option is required if `meanspec` is not specified. This option is also required for effect-size determination unless `cellweights()` is specified.

`ncols(#)` specifies the number of columns or the number of levels of the column factor in a two-way ANOVA. At least two columns must be specified. This option is required if `meanspec` is not specified. This option is also required for effect-size determination unless `cellweights()` is specified.

`factor(row|column|rowcol)` specifies the effect of interest for which power and sample-size analysis is to be performed. In a two-way ANOVA, the tested effects include the main effects of a row factor (row effect), the main effects of a column factor (column effect), or the interaction effects between the row and column factors (row–column effect). The default is `factor(row)`.

`vareffect(numlist)` specifies the variance explained by the tested effect specified in `factor()`. For example, if `factor(row)` is specified, `vareffect()` specifies the variance explained by the row factor. This option is required if the `factor()` option is specified and cell means are not specified. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varrow()`, `varcolumn()`, or `varrowcolumn()` may be specified.

`varrow(numlist)` specifies the variance explained by the row factor. This option is equivalent to specifying `factor(row)` and `vareffect(numlist)` and thus cannot be combined with `factor()`. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varrow()`, `varcolumn()`, or `varrowcolumn()` may be specified.

`varcolumn(numlist)` specifies the variance explained by the column factor. This option is equivalent to specifying `factor(column)` and `vareffect(numlist)` and thus cannot be combined with `factor()`. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varrow()`, `varcolumn()`, or `varrowcolumn()` may be specified.

`varrowcolumn(numlist)` specifies the variance explained by the interaction between row and column factors. This option is equivalent to specifying `factor(rowcol)` and `vareffect(numlist)` and thus cannot be combined with `factor()`. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varrow()`, `varcolumn()`, or `varrowcolumn()` may be specified.

`varerror(numlist)` specifies the error (within-cell) variance. The default is `varerror(1)`.

`showmatrices` specifies that the matrices of cell means and cell sizes be displayed, when applicable. The cell means will be displayed only if specified. The cell sizes will be displayed only for an unbalanced design.

`showmeans` specifies that the cell means be reported. For a text or graphical output, this option is equivalent to `showmatrices` except only the cell-mean matrix will be reported. For a tabular output, the columns containing cell means will be included in the default table.

`showcellsizes` specifies that the cell sizes be reported. For a text or graphical output, this option is equivalent to `showmatrices` except only the cell-sizes matrix will be reported. For a tabular output, the columns containing cell sizes will be included in the default table.

`parallel`; see [PSS] **power**.

Table

`table`, `table()`, `notable`; see [PSS] **power**, **table**.

`saving()`; see [PSS] **power**.

## Graph

`graph`, `graph()`; see [PSS] [power](#), [graph](#). Also see the [column](#) table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value of the sample size for the sample-size determination or the initial value of the effect size  $\delta$  for the effect-size determination. The default uses a bisection algorithm to bracket the solution.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power twoway` but is not shown in the dialog box:

`notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power twoway*

*Alternative ways of specifying effect*

*Computing sample size*

*Computing power*

*Computing effect size and target variance explained by the tested effect*

*Testing hypotheses about means from multiple populations*

This entry describes the `power twoway` command and the methodology for power and sample-size analysis for two-way ANOVA. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

ANOVA has been one of the most widely used statistical tools in many scientific applications. Two-way ANOVA models allow analysts to study the effects of two factors simultaneously. The term “two way” refers to two factors each containing an arbitrary number of groups or levels.

For example, consider a type of drug with three levels of dosage in reducing blood pressure for males and females. In this case, three interesting hypotheses arise: an investigator may wish to test whether the average change in blood pressure is the same for both genders, whether the average change in blood pressure is the same across all levels of dosage regardless of gender, or whether there is any interaction between dosage levels and gender.

This entry describes power and sample-size analysis for the inference about main and interaction effects of two factors based on hypothesis testing. Let  $\mu_{jk}$  be the cell mean of the  $j$ th row and the  $k$ th column in a two-way cell-means ANOVA model,  $\mu_{j\cdot}$  be the marginal mean of the  $j$ th row,  $\mu_{\cdot k}$  be the marginal mean of the  $k$ th column, and  $\mu_{\cdot\cdot}$  be the grand mean. The  $j$ th-row-by- $k$ th-column interaction effect is then  $(ab)_{jk} = \mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu_{\cdot\cdot}$ . We consider the null hypotheses 1)  $H_0: \mu_{\cdot 1} = \dots = \mu_{\cdot J}$ , for testing the main row effect; 2)  $H_0: \mu_{\cdot 1} = \dots = \mu_{\cdot K}$ , for testing the main column effect; and 3)  $H_0: \text{all } (ab)_{jk} = 0$ , for testing the row-by-column interaction effect.

The test statistic for each of the three hypotheses is based on the ratio of the variance explained by the tested effect to the error variance. Under the null hypothesis, the test statistics used for items 1, 2, and 3 above have an  $F$  distribution. We will refer to the corresponding tests as  $F$  tests for row, column, and row-by-column effects. For power analysis, we consider the distribution of the test statistic under the alternative hypothesis. This distribution is a noncentral  $F$  distribution for all the considered tests. Power is a function of the noncentrality parameter, and the noncentrality parameter is a function of the ratio of the standard deviation of the tested effect to the standard deviation of the errors. As such, the effect size for each of the  $F$  tests is defined as the square root of the ratio of the variance explained by the tested effect to the error variance.

`power twoway` performs power and sample-size computation for a two-way fixed-effects ANOVA model based on an  $F$  test of the effect of interest.

## Using power twoway

`power twoway` computes sample size, power, or effect size and target variance of the effect for a two-way fixed-effects ANOVA. All computations are performed assuming a significance level of 0.05. You may change the significance level by specifying the `alpha()` option.

By default, the computations are performed for an  $F$  test of the main row effects; `factor(row)` is assumed. You can instead request a test of the main column effects by specifying `factor(column)` or a test of the row-by-column interaction effects by specifying `factor(rowcol)`. The error variance for all tests is assumed to be 1 but may be changed by specifying the `varerror()` option.

To compute the total sample size, you must specify the alternative `meanspec` and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the total sample size in the `n()` option and the alternative `meanspec`.

Instead of the alternative cell means, you can specify the number of rows in the `nrows()` option, the number of columns in the `ncols()` option, and the variance explained by the tested effect in the `vareffect()` option when computing sample size or power. See [Alternative ways of specifying effect](#).

To compute effect size, the square root of the ratio of the variance explained by the tested factor to the error variance, and the target variance explained by the tested factor, you must specify the total sample size in the `n()` option, the power in the `power()` option, and the number of rows and columns in `nrows()` and `ncols()`, respectively.

By default, all computations assume a balanced- or an equal-allocation design. You can use the `cellweights()` option to specify an unbalanced design for power, sample-size, or effect-size computations. For power and effect-size computations of a balanced design, you can also specify the `npercell()` option to specify a cell size instead of a total sample size in `n()`.

In a two-way ANOVA, sample size and effect size depend on the noncentrality parameter of the  $F$  distribution, and their estimation requires iteration. The default initial values are obtained from a bisection search that brackets the solution. If you desire, you may change this by specifying your own value in the `init()` option. See [\[PSS\] power](#) for the descriptions of other options that control the iteration procedure.

In a two-way ANOVA, all computations depend on the noncentrality parameter of the  $F$  distribution. The sample-size and effect-size computation requires a nonlinear search algorithm where the default initial value is obtained using a bisection search algorithm that brackets the solution.

`power twoway` provides several ways of specifying the study parameters that we discuss next.



## Alternative ways of specifying effect

To compute power or sample size, you must specify the magnitude of the effect desired to be detected by the test. With `power twoway`, you can do this in several ways. For example, consider a two-way model with  $J = 2 \geq 2$  row-factor levels and  $K = 3 \geq 2$  column-factor levels. You can specify either the individual alternative *meanspec*,

```
power twoway m1,1 m1,2 m1,3 \ m2,1 m2,2 m2,3 [, factor() ...]
```

or the variance of the tested effect and the number of rows  $J$  and columns  $K$ :

```
power twoway, factor() vareffect(#) nrows(2) ncols(3) [...]
```

You can also replace `vareffect()` and `factor()` in the above with the `varrow()`, `varcolumn()`, or `varrowcolumn()` option. And you can specify multiple values of the variances in these options.

As an alternative to directly specifying alternative cell means following the command name, you can define a Stata matrix containing these means and use it with `power twoway`. For example,

```
matrix define meanmat = (m1,1, m1,2, m1,3 \ m2,1, m2,2, m2,3)
```

The matrix must have at least two rows and two columns.

```
power twoway meanmat, ...
```

In the following sections, we describe the use of `power twoway` accompanied by examples for computing sample size, power, and effect size.

## Computing sample size

To compute sample size, you must specify the alternative cell means or the variance of the tested effect and, optionally, the power of the test in the `power()` option. A power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-way ANOVA—row effect

van Belle et al. (2004, 376) provide an example of an experimental study that investigates the effect of an automobile emission pollutant, nitrogen dioxide ( $NO_2$ ). The experiment considers the effect of  $NO_2$  exposure on protein leakage in the lungs of mice. In the experimental group, mice were exposed to 0.5 part per million (ppm)  $NO_2$  for 10, 12, and 14 days. Measurements on the response variable, serum fluorescence, were taken on mice in the experimental (exposed) and control (unexposed) groups. The analysis of these data used a two-way ANOVA model with the exposure status as a row factor and the number of days of exposure to  $NO_2$  as a column factor. The row factor has two levels, exposed or unexposed, and the column factor has three levels: 10, 12, and 14 days.

Suppose that investigators are planning to conduct another similar study. They would like to know how many subjects, mice, they need for the experiment. We will use the estimates of parameters from the above study to answer this question.

The estimated cell means from this study over the number of days are 134.4, 143, and 91.3 in the control group and 106.4, 173.2, and 145.5 in the experimental group. From the ANOVA table on page 379 (van Belle et al. 2004), the estimated residual variance is 1417.35. For convenience, we round these numbers down to the nearest integers in our computations.

We begin by computing sample size for testing the main treatment (exposure) effects using `power twoway`'s defaults for other aspects of the study: a balanced design, a 5% significance level, and 80% power. (Option `factor(row)` is assumed by default.)

```

. power twoway 134 143 91 \ 106 173 145, varerror(1417)
Performing iteration ...
Estimated sample size for two-way ANOVA
F test for row effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.2479
    N_r =      2
    N_c =      3
    means = <matrix>
    Var_r =    87.1111
    Var_e =   1417.0000
Estimated sample sizes:
    N =      132
    N per cell =    22

```

Assuming a balanced design, we need a total of 132 mice with 22 mice per cell to detect the effect of exposure to  $NO_2$  on the protein leakage of mice.

Like all other power methods, `power twoway` reports study parameters first and the estimated parameters next. The reported study parameters include the specified and implied parameters such as significance level, power, number of rows, number of columns, and so on. `power twoway` does not display the specified cell means by default but indicates in the output that the means are specified. You can specify the `showmeans` option to display cell means as a matrix.

In addition to the specified and implied study parameters, `power twoway` reports the value of the effect size,  $\delta = \sqrt{87.1111/1417} = 0.2479$ , computed as a square root of the ratio between the variance of the row effect `Var_r` and the error variance `Var_e`. As for the one-way ANOVA models, the effect size  $\delta$  provides a unit-less measure of the magnitude of an effect with a lower bound of zero, meaning no effect. It corresponds to Cohen's effect-size measure  $f$  (Cohen 1988). Cohen's convention is that  $f = 0.1$  means small effect size,  $f = 0.25$  means medium effect size, and  $f = 0.4$  means large effect size. According to this convention, the effect size considered in our example is medium.

◀

### ▷ Example 2: Sample size for a two-way ANOVA—column effect

Continuing with [example 1](#), we can compute the required sample size for the main column effects by specifying the `factor(column)` option:

```
. power twoway 134 143 91 \ 106 173 145, varerror(1417) factor(column)
Performing iteration ...
Estimated sample size for two-way ANOVA
F test for column effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.4889
      N_r = 2
      N_c = 3
      means = <matrix>
      Var_c = 338.6667
      Var_e = 1417.0000
Estimated sample sizes:
      N = 48
      N per cell = 8
```

Assuming a balanced design, we need a total of 48 mice with 8 mice per cell to detect the effect of the length of exposure to  $NO_2$  on the protein leakage of mice.

Similarly to the row effect, the effect size for the column effect,  $\delta = \sqrt{338.6667/1417} = 0.4889$ , is computed as a square root of the ratio between the variance of the column effect `Var_c` and the error variance `Var_e`. The interpretation remains the same but with respect to the main column effects. According to Cohen's scale, the effect size corresponding to the test of the main column effects is large, so we need fewer subjects to detect the column effect than we need to detect the previous row effect.

◀

### ▶ Example 3: Sample size for a two-way ANOVA—row-by-column effect

Continuing with [example 2](#), we can also compute the required sample size for the row-by-column effects interaction by specifying the `factor(rowcol)` option:

```

. power twoway 134 143 91 \ 106 173 145, varerror(1417) factor(rowcol)
Performing iteration ...
Estimated sample size for two-way ANOVA
F test for row-by-column effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.4572
      N_r    =         2
      N_c    =         3
      means  = <matrix>
      Var_rc = 296.2222
      Var_e  = 1417.0000
Estimated sample sizes:
      N      =        54
      N per cell =         9

```

For a balanced design, we need a total of 54 mice with 9 mice per cell to detect the joint effects of exposure and the length of exposure to  $NO_2$  on the protein leakage of mice.

Similarly to the row-by-column effects, the effect size for the row-by-column effect,  $\delta = \sqrt{296.2222/1417} = 0.4572$ , is computed as a square root of the ratio between the variance of the row-by-column effect `Var_rc` and the error variance `Var_e`. The interpretation is again the same but with respect to the interaction of row-by-column effects. According to Cohen's scale, the effect size corresponding to the test of the row-by-column effects is also large, so we need fewer subjects to detect this effect than we need to detect the row effect. The effect size is similar to the column effect size, so the required numbers of subjects are comparable for the two tests. As a final sample size, we would choose the largest of the three sizes to ensure that we have enough subjects to detect any of the considered effects.

◀

#### ▷ Example 4: Alternative ways of specifying effect

Instead of specifying the alternative cell means as in previous examples, we can specify the variance explained by the corresponding tested effect and the numbers of rows and columns.

For instance, from [example 2](#), the variance explained by the column effect was computed to be 338.6667. We specify this value in `vareffect()` as well as the number of rows in `nrows()` and the number of columns in `ncols()`:

```

. power twoway, varerror(1417) factor(column) vareffect(338.6667)
> nrows(2) ncols(3)
Performing iteration ...
Estimated sample size for two-way ANOVA
F test for column effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.4889
      N_r =         2
      N_c =         3
      Var_c =   338.6667
      Var_e =  1417.0000
Estimated sample sizes:
      N =         48
      N per cell =         8

```

We obtain the exact same results as in example 2.

A shorthand for the specification of `factor(column)` and `vareffect()` is the `varcolumn()` option. You can verify that the specification

```

. power twoway, varerror(1417) varcolumn(338.6667) nrows(2) ncols(3)
  (output omitted)

```

produces results identical to the results above.

You can also use similar alternative specifications for the tests of row and row-by-column effects with intuitive modifications to the syntax.

`power twoway` also provides another alternative specification of the cell means. Instead of specifying alternative cell means directly following the command line, as in [example 2](#), we can define a matrix, say, `means`, containing these means and use it with `power twoway`:

```

. matrix define means = (134, 143, 91 \ 106, 173, 145)
. power twoway means, varerror(1417) factor(column)
  (output omitted)

```

You can again verify that the results are identical to the previous results.

◀

## ▶ Example 5: Unbalanced design

Continuing with [example 1](#), let's compute the required sample size for an unbalanced design. For instance, consider a design in which the control group (the first row) contains twice as many subjects as the experimental group (the second row) for each level of the other factor. We use the `cellweights()` option to specify weights for each cell.

```

. power twoway 134 143 91 \ 106 173 145, varerror(1417) cellweights(2 2 2 \ 1 1 1)
> showcellsizes
Performing iteration ...
Estimated sample size for two-way ANOVA
F test for row effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.2338
      N_r = 2
      N_c = 3
      means = <matrix>
      Var_r = 77.4321
      Var_e = 1417.0000
Estimated sample sizes:
      N = 153
      Average N = 25.5000
Cell sample sizes

```

		columns		
		1	2	3
rows	1	34	34	34
	2	17	17	17

The required total sample size for this unbalanced design is 153 with the average number of subjects in a cell of 25.5. We specified the `showcellsizes` option to display the number of subjects for each cell along with the total and average sample sizes that are displayed by default for an unbalanced design.

You can alternatively specify cell weights as a matrix.



## Computing power

To compute power, you must specify the total sample size in the `n()` option and the desired effect size, expressed using alternative cell means, for example. See [Alternative ways of specifying effect](#).

### ▷ Example 6: Power for a two-way ANOVA

Continuing with [example 1](#), suppose that we anticipate a sample of 90 mice. To compute the corresponding power, we specify the sample size of 90 in `n()`.

```
. power twoway 134 143 91 \ 106 173 145, varerror(1417) n(90)
Estimated power for two-way ANOVA
F test for row effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
  alpha = 0.0500
    N = 90
  N per cell = 15
  delta = 0.2479
    N_r = 2
    N_c = 3
  means = <matrix>
  Var_r = 87.1111
  Var_e = 1417.0000
Estimated power:
  power = 0.6426
```

For this smaller sample size, the power for detecting the effect size of 0.25 is only 64%.

◀

### ▷ Example 7: Multiple values of study parameters

Continuing with [example 6](#), we may want to check powers for several sample sizes. We simply list multiple sample-size values in `n()`:

```
. power twoway 134 143 91 \ 106 173 145, varerror(1417) n(90 114 126)
Estimated power for two-way ANOVA
F test for row effect
Ho: delta = 0 versus Ha: delta != 0
  means = <matrix>
```

alpha	power	N	N_per_cell	delta	N_r	N_c	Var_r	Var_e
.05	.6426	90	15	.2479	2	3	87.11	1417
.05	.7466	114	19	.2479	2	3	87.11	1417
.05	.7884	126	21	.2479	2	3	87.11	1417

The larger the sample size, the larger the power.

We can even compute results for multiple sample sizes and, for example, multiple values of error variances.

```
. power twoway 134 143 91 \ 106 173 145, varerror(1000 1800) n(90 114 126)
> table(, sep(3))
Estimated power for two-way ANOVA
F test for row effect
Ho: delta = 0 versus Ha: delta != 0
means = <matrix>
```

alpha	power	N	N_per_cell	delta	N_r	N_c	Var_r	Var_e
.05	.7904	90	15	.2951	2	3	87.11	1000
.05	.8776	114	19	.2951	2	3	87.11	1000
.05	.9076	126	21	.2951	2	3	87.11	1000
.05	.5411	90	15	.22	2	3	87.11	1800
.05	.6436	114	19	.22	2	3	87.11	1800
.05	.6878	126	21	.22	2	3	87.11	1800

We specified `table()`'s suboption `separator()`, abbreviated to `sep()`, to improve readability of the table.

We can also compute results for combinations of multiple values of other study parameters.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## Computing effect size and target variance explained by the tested effect

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, power, sample size, and the numbers of rows and columns must be specified.

The effect size is defined as a square root of the ratio of the variance explained by the tested effect to the error variance. The effect size and the target variance explained by the tested effect are computed.

### ► Example 8: Effect size for a two-way ANOVA—row effect

Continuing with [example 6](#), we now want to compute the effect size that can be detected for a sample of 90 subjects and a power of 80%. We specify both parameters in the respective options. For the effect-size determination, we must also specify the number of rows in `nrows()` and the number of columns in `ncols()`:



```

. power twoway, varerror(1417) n(90) power(0.8) nrows(2) ncols(3)
Performing iteration ...
Estimated row variance for two-way ANOVA
F test for row effect
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
        N =      90
N per cell =     15
      N_r =      2
      N_c =      3
      Var_e = 1417.0000

Estimated effect size and row variance:
      delta =    0.2987
      Var_r = 126.4634

```

With a smaller sample size, given the same power, we can only detect a larger effect size of 0.2987, compared with the effect size of 0.2479 from [example 1](#). The corresponding estimate of the variance explained by the row effect is 126.46, given the error variance of 1417.

◀

## Testing hypotheses about means from multiple populations

### ▶ Example 9: Two-way ANOVA

After the initial power and sample-size planning, we can use Stata's `anova` command to perform inference for two-way ANOVA based on the collected sample. We show a quick example of how to do this here; see [\[R\] anova](#) for more examples and details.

We use data on systolic blood pressure. Consider a sample of 58 patients, each suffering from 1 of 3 different diseases, who were randomly assigned to 1 of 4 different drug treatments and whose change in systolic blood pressure was recorded. To test for the effects of the drug and the disease and their interaction, we type the following:

```

. use http://www.stata-press.com/data/r14/systolic
(Systolic Blood Pressure Data)
. anova systolic drug disease drug#disease

```

Source	Partial SS	df	MS	F	Prob>F
Model	4259.3385	11	387.21259	3.51	0.0013
drug	2997.4719	3	999.15729	9.05	0.0001
disease	415.87305	2	207.93652	1.88	0.1637
drug#disease	707.26626	6	117.87771	1.07	0.3958
Residual	5080.8167	46	110.45254		
Total	9340.1552	57	163.86237		

We find that only the main effect of the drug is significant.

Suppose that we would like to conduct a similar study. We use the estimates from the study above to compute the required sample size for our new study. We are particularly interested in testing the interaction between the drug and disease, so we would like to compute the sample size for this test.

First, we estimate the cell means of systolic blood pressure for different treatment and disease levels:

```
. table drug disease, contents(mean systolic) format(%9.0f)
```

Drug Used	Patient's Disease		
	1	2	3
1	29	28	20
2	28	34	18
3	16	4	8
4	14	13	14

From the `twoway` output, the estimate of the error variance is roughly 110. Second, we specify the means and the error variance with `power twoway` and compute the required sample size for a balanced design assuming 5% significance level and 80% power for the test of interaction effects.

```
. power twoway 29 28 20 \ 28 34 18 \ 16 4 8 \ 14 13 14, varerror(110) f(rowcol)
```

```
Performing iteration ...
```

```
Estimated sample size for two-way ANOVA
```

```
F test for row-by-column effect
```

```
Ho: delta = 0 versus Ha: delta != 0
```

```
Study parameters:
```

```
alpha = 0.0500
```

```
power = 0.8000
```

```
delta = 0.3465
```

```
N_r = 4
```

```
N_c = 3
```

```
means = <matrix>
```

```
Var_rc = 13.2083
```

```
Var_e = 110.0000
```

```
Estimated sample sizes:
```

```
N = 132
```

```
N per cell = 11
```

We need a total of 132 subjects with 11 subjects per cell to detect the drug-by-disease effect size of 0.3465 for this design.

To determine the final sample size, you may want to repeat the same computations for the tests of the main effects of drug and the main effects of disease and select the sample size based on the three tests.

## Stored results

`power twoway` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N_avg)</code>	average sample size
<code>r(N#_1_#_2)</code>	number of subjects in cell ( $\#_1$ , $\#_2$ )
<code>r(N_per_cell)</code>	number of subjects per cell
<code>r(N_rc)</code>	number of cells
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(balanced)</code>	1 for a balanced design; 0 otherwise
<code>r(cwgt#_1_#_2)</code>	cell weight ( $\#_1$ , $\#_2$ )
<code>r(N_r)</code>	number of rows
<code>r(N_c)</code>	number of columns
<code>r(m#_1_#_2)</code>	cell mean ( $\#_1$ , $\#_2$ )
<code>r(Var_r)</code>	row variance
<code>r(Var_c)</code>	column variance
<code>r(Var_rc)</code>	row-by-column variance
<code>r(Var_e)</code>	error variance
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for sample size or effect size
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>twoway</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrices

<code>r(pss_table)</code>	table of results
<code>r(Nij)</code>	cell-sizes matrix
<code>r(means)</code>	cell-means matrix
<code>r(cwgt)</code>	cell-weights matrix

## Methods and formulas

Consider factor  $A$  with  $J$  levels and factor  $B$  with  $K$  levels. Let  $\mu_{jk}$  be the mean of cell  $(j, k)$  in a table formed by the levels of factors  $A$  and  $B$ . For example, let  $J = 3$  and  $K = 3$ ; then the following cell-means table summarizes the experiment.

Factor A	Factor B			Total
	$k = 1$	$k = 2$	$k = 3$	
$j = 1$	$\mu_{11}$	$\mu_{12}$	$\mu_{13}$	$\mu_{1\cdot}$
$j = 2$	$\mu_{21}$	$\mu_{22}$	$\mu_{23}$	$\mu_{2\cdot}$
$j = 3$	$\mu_{31}$	$\mu_{32}$	$\mu_{33}$	$\mu_{3\cdot}$
Total	$\mu_{\cdot 1}$	$\mu_{\cdot 2}$	$\mu_{\cdot 3}$	$\mu_{\cdot\cdot}$

Methods and formulas are presented under the following headings:

*Main effects*  
*Interaction effects*  
*Hypothesis testing*

## Main effects

Main effects measure the deviation of the factor-level means from the overall or grand mean. The larger the main effect, the more likely you can detect the effect. From the above table, the main effect of factor  $A$  at the  $j$ th level is  $a_j = \mu_{j\cdot} - \mu_{\cdot\cdot}$ . Similarly, the main effect of factor  $B$  at the  $k$ th level is  $b_k = \mu_{\cdot k} - \mu_{\cdot\cdot}$ . The overall mean can be expressed as

$$\mu_{\cdot\cdot} = \frac{\sum_{j=1}^J \sum_{k=1}^K \mu_{jk}}{JK} = \frac{\sum_{j=1}^J \mu_{j\cdot}}{J} = \frac{\sum_{k=1}^K \mu_{\cdot k}}{K}$$

This implies that

$$\sum_{j=1}^J a_j = 0 \quad \text{and} \quad \sum_{k=1}^K b_k = 0$$

## Interaction effects

Unlike main effects that measure the effect of individual factors on the dependent variable, interaction effects measure the effect of the two factors jointly on the dependent variable. For example, the interaction effect of factor  $A$  at the  $j$ th level and factor  $B$  at the  $k$ th level is  $(ab)_{jk} = \mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu_{\cdot\cdot}$ .

The sum of interaction effects is zero:

$$\sum_{j=1}^J (ab)_{jk} = 0 \text{ at each level of } k = 1, \dots, K$$

$$\sum_{k=1}^K (ab)_{jk} = 0 \text{ at each level of } j = 1, \dots, J$$

This implies

$$\sum_{j=1}^J \sum_{k=1}^K (ab)_{jk} = 0$$

## Hypothesis testing

Let  $n$  denote the total sample size and  $y_{ijk}$  denote the response of  $i$ th individual at the  $j$ th level of factor  $A$  and  $k$ th level of factor  $B$  for  $i = 1, \dots, n_{jk}$  such that  $n = \sum_{j,k} n_{jk}$ . ANOVA models assume that responses  $y_{ijk}$  within each cell are independent and identically distributed random normal with constant variance  $\sigma_e^2$ . Using the definitions of  $a_j$ ,  $b_k$ , and  $(ab)_{jk}$  from the previous sections, our linear model is expressed as

$$\begin{aligned} y_{ijk} &= \mu_{..} + a_j + b_k + (ab)_{jk} + e_{ijk} \\ &= \mu_{jk} + e_{ijk} \end{aligned}$$

where  $\mu_{..}$  is the overall mean, and  $e_{ijk}$ 's are the independent error terms that have the standard normal distribution. The first equation corresponds to the formulation of an ANOVA model using effects, and the second formulation corresponds to the cell-means formulation.

The variance explained by the row effects is  $\sigma_a^2 = \sum_j a_j^2/J$ , by the column effects is  $\sigma_b^2 = \sum_k b_k^2/K$ , and by the row-by-column effects is  $\sigma_{(ab)}^2 = \sum_{j,k} (ab)_{jk}^2/JK$ .

The following sets of hypotheses are of interest in a two-way ANOVA:

$$H_0: \text{all } a_j = 0 \quad \text{versus} \quad H_a: \text{at least one } a_j \neq 0 \quad (1)$$

$$H_0: \text{all } b_k = 0 \quad \text{versus} \quad H_a: \text{at least one } b_k \neq 0 \quad (2)$$

$$H_0: \text{all } (ab)_{jk} = 0 \quad \text{versus} \quad H_a: \text{at least one } (ab)_{jk} \neq 0 \quad (3)$$

Hypotheses (1) and (2) test the main effects of factors  $A$  and  $B$ , respectively, and hypothesis (3) tests the interaction effects between  $A$  and  $B$ .

To test the above hypotheses, we can use the general linear model framework discussed in [Methods and formulas](#) of [PSS] **power oneway**. We recapitulate it here with application to the two-way model.

A general test statistic for testing hypotheses like (1), (2), and (3) is given by

$$F_C = \frac{SS_C}{\nu \hat{\sigma}_e^2} \quad (4)$$

where  $SS_C = (\mathbf{Cb})' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}' \}^{-1} (\mathbf{Cb})$ . The  $n \times JK$  matrix  $\mathbf{X}$  specifies the coding for the two-way design. Matrix  $\mathbf{C}$  is  $\nu \times JK$  and contains the contrasts for the means that are used to test each of the three hypotheses. For the two-way design,  $\nu$  is  $J - 1$ ,  $K - 1$ , or  $(J - 1)(K - 1)$  for hypotheses (1), (2), and (3), respectively.

Let  $\alpha$  be the significance level and  $F_{\nu, n-JK, 1-\alpha}$  be the  $(1 - \alpha)$ th quantile of an  $F$  distribution with  $\nu$  numerator and  $n - JK$  denominator degrees of freedom. We reject the null hypothesis if we observe a statistic  $F_C > F_{\nu, n-JK, 1-\alpha}$ .

Under the alternative hypothesis, the test statistic (4) is distributed as a noncentral  $F$  distribution with  $\nu$  numerator and  $n - JK$  denominator degrees of freedom and a noncentrality parameter  $\lambda$  given by

$$\begin{aligned} \lambda &= n(\mathbf{Cb})' \{ \mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1} \mathbf{C}' \}^{-1} (\mathbf{Cb}) / \sigma_e^2 \\ &= n\delta^2 \end{aligned}$$

where the matrix  $\ddot{\mathbf{X}}$  contains the unique rows of  $\mathbf{X}$  such that  $\boldsymbol{\mu} = \ddot{\mathbf{X}}\mathbf{b}$ ,  $\mathbf{W} = \text{diag}(w_1, \dots, w_{JK})$ , and  $\delta$  is the effect size. For a two-way design, the dimension of  $\ddot{\mathbf{X}}$  is  $JK \times JK$ , and the weights are  $w_i = n_{jk}/n$ ,  $i = (k - 1)J + j$ . The cell-means parameterization simplifies  $\ddot{\mathbf{X}}$  to the identity matrix,  $\mathbf{I}_{JK}$ .

The rows of  $\ddot{X}$  and weights  $w_i$  are associated with the column-major, `vec()`, order of the two-way table with factor  $A$  indexed on the rows and factor  $B$  indexed on the columns. (See the  $3 \times 3$  table in the beginning of this section, and scan each column  $k = 1, 2, 3$ .) The weight  $w_i$  can be reexpressed as a cell weight  $\tilde{w}_i$ , which is independent of the sample size  $n$ ; see *Methods and formulas* of **[PSS] power oneway** for details.

When the `cellweights()` option is specified, a constant cell-size multiplier  $n_c$  is computed and rounded to an integer unless the `nfractional` option is specified. The cell sizes are then computed as  $\tilde{w}_j n_c$ . The actual sample size, `N_a`, is the sum of the cell sizes.

## References

- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- van Belle, G., L. D. Fisher, P. J. Heagerty, and T. S. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences*. 2nd ed. New York: Wiley.

## Also see

- [PSS] power** — Power and sample-size analysis for hypothesis tests
- [PSS] power oneway** — Power analysis for one-way analysis of variance
- [PSS] power repeated** — Power analysis for repeated-measures analysis of variance
- [PSS] power, graph** — Graph results from the power command
- [PSS] power, table** — Produce table of results from the power command
- [PSS] Glossary**
- [R] anova** — Analysis of variance and covariance

# Title

**power repeated** — Power analysis for repeated-measures analysis of variance

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power repeated` computes sample size, power, or effect size for one-way or two-way repeated-measures analysis of variance (ANOVA). By default, it computes sample size for given power and effect size. Alternatively, it can compute power for given sample size and effect size or compute effect size for given sample size, power, and number of groups. Also see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Quick start

Sample size for a repeated measures design with one 3-level within-subject factor, a correlation of 0.3 between measurements, and an error variance of 42 with default power of 0.8 and significance level  $\alpha = 0.05$

```
power repeated 25 27 22, varerror(42) corr(.3)
```

Same as above, specified as 3 measurements on 1 group with within-subject variance of 4.22

```
power repeated, varerror(42) corr(.3) nrepeated(3) ngroups(1) ///  
varwithin(4.22)
```

Same as above, specified as cell means in matrix `cm`

```
matrix cm = (25,27,22)  
power repeated cm, corr(.3) varerror(42)
```

As above, and show the mean and covariance matrices in the output

```
power repeated cm, corr(.3) varerror(42) showmatrices
```

Sample size for the between effect in a design with a 3-level within-subject factor and a 2-level between-subject factor

```
power repeated 18 14 12\14 13 10, covmatrix(24 9 9\9 24 9\9 9 24)
```

Same as above, specified as cell means in matrix `cm2` and covariances in matrix `cov`

```
matrix cm2 = (18,14,12\14,13,10)  
matrix cov = (24,9,9\9,24,9\9,9,24)  
power repeated cm2, covmatrix(cov)
```

As above, but for the within effect

```
power repeated cm2, covmatrix(cov) factor(within)
```

As above, but for the between–within effect

```
power repeated cm2, covmatrix(cov) factor(bwithin)
```

Power for a design with one within-subject factor, a sample size of 25, and  $\alpha = 0.01$

```
power repeated 25 27 22, varerror(42) corr(.3) n(25) alpha(.01)
```

As above, but for sample sizes of 20, 24, 28, and 32

```
power repeated 25 27 22, varerror(42) corr(.3) n(20(4)32)
```

As above, but show results in a graph of sample size versus power

```
power repeated 25 27 22, varerror(42) corr(.3) n(20(4)32) graph
```

Power for the between effect of a design with a 3-level within-subject factor and a 2-level between-subject factor with a sample size of 160

```
power repeated cm2, covmatrix(cov) n(160)
```

As above, but specify sample sizes of 100 and 140 for groups 1 and 2, respectively

```
power repeated cm2, covmatrix(cov) n1(100) n2(140)
```

Effect size for a one-group repeated-measures design

```
power repeated, varerror(42) corr(.3) n(24) ngroups(1) ///  
nrepeated(3) power(.8)
```

Effect size for the within-subject effect of a design with a 3-level within-subject factor, a 2-level between-subject factor, and a sample size of 160

```
matrix cov = (24,9,9\9,24,9\9,9,24)  
power repeated, covmatrix(cov) n(160) power(.8) ngroups(2) ///  
factor(within)
```

## Menu

Statistics > Power and sample size



## Syntax

Compute sample size

```
power repeated meanspec, corrspec [power(numlist) options]
```

Compute power

```
power repeated meanspec, n(numlist) corrspec [options]
```

Compute effect size

```
power repeated, n(numlist) power(numlist) ngroups(#) corrspec [options]
```

where *meanspec* is either a matrix *matname* containing cell means or individual cell means in a matrix form:

$$m_{1,1} \ m_{1,2} \ [\dots \ m_{1,K}] \ [\backslash \dots \ [\backslash \ m_{J,1} \ m_{J,2} \ [\dots \ m_{J,K}]]]$$

$m_{jk}$ , where  $j = 1, 2, \dots, J$  and  $k = 1, 2, \dots, K$ , is the alternative cell mean or the cell mean of the  $j$ th row (group) and  $k$ th column (repeated measure) under the alternative hypothesis.

*matname* is the name of a Stata matrix with  $J$  rows and  $K$  columns containing values of alternative cell means.

At least one group,  $J = 1$ , and two repeated measures,  $K = 2$ , must be specified.

where *corrspec* for computing power and sample size is {corr(*numlist*) | covmatrix(*matname*)}, and *corrspec* for computing effect size is {nrepeated(#) corr(*numlist*) | covmatrix(*matname*)}.

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample sizes
* <u>npergroup</u> ( <i>numlist</i> )	number of subjects per group; implies balanced design
* <u>n#</u> ( <i>numlist</i> )	number of subjects in group #
<u>grweights</u> ( <i>wgtspec</i> )	group weights; default is one for each group, meaning equal group sizes
<u>ngroups</u> (#)	number of groups
<u>nrepeated</u> (#)	number of repeated measures
* <u>corr</u> ( <i>numlist</i> )	correlation between repeated measures; one of <code>corr()</code> or <code>covmatrix()</code> is required
<u>covmatrix</u> ( <i>matname</i> )	covariance between repeated measures; one of <code>corr()</code> or <code>covmatrix()</code> is required
<u>factor</u> ( <u>between</u>   <u>within</u>   <u>bwithin</u> )	tested effect: between, within, or between–within; default is <code>factor(between)</code>
* <u>vareffect</u> ( <i>numlist</i> )	variance explained by the tested effect specified in <code>factor()</code>
* <u>varbetween</u> ( <i>numlist</i> )	variance explained by the between-subjects effect; synonym for <code>factor(between)</code> and <code>vareffect(numlist)</code>
* <u>varwithin</u> ( <i>numlist</i> )	variance explained by the within-subject effect; synonym for <code>factor(within)</code> and <code>vareffect(numlist)</code>
* <u>varbwithin</u> ( <i>numlist</i> )	variance explained by the between–within effect; synonym for <code>factor(bwithin)</code> and <code>vareffect(numlist)</code>
* <u>varerror</u> ( <i>numlist</i> )	error variance; default is <code>varerror(1)</code> when <code>corr()</code> is specified; not allowed with <code>covmatrix()</code>
<u>showmatrices</u>	display cell-means matrix and covariance matrix
<u>showmeans</u>	display cell means
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [, <code>replace</code> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>

Iteration	
<code>init(#)</code>	initial value for the sample size or the effect size; default is to use a bisection algorithm to bound the solution
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 [numlist](#).

`notitle` does not appear in the dialog box.

<i>wgtspec</i>	Description
<code>#<sub>1</sub> #<sub>2</sub> ... #<sub>J</sub></code>	<i>J</i> group weights. Weights must be positive and must be integers unless option <code>nfractional</code> is specified. Multiple values for each group weight # <sub><i>j</i></sub> can be specified as a <a href="#">numlist</a> enclosed in parentheses.
<i>matname</i>	matrix with <i>J</i> columns containing <i>J</i> group weights. Multiple rows are allowed, in which case each row corresponds to a different set of <i>J</i> weights or, equivalently, column <i>j</i> corresponds to a <a href="#">numlist</a> for the <i>j</i> th weight.

where *tablespect* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N_per_group</code>	number of subjects per group	$N/N_g$
<code>N_avg</code>	average number of subjects per group	$N_{avg}$
<code>N#</code>	number of subjects in group #	$N_{\#}$
<code>delta</code>	effect size	$\delta$
<code>N_g</code>	number of groups	$N_g$
<code>N_rep</code>	number of repeated measurements	$N_{rep}$
<code>m#1_#2</code>	cell mean (#1, #2): group #1, occasion #2	$\mu_{\#1, \#2}$
<code>Var_b</code>	between-subjects variance	$\sigma_b^2$
<code>Var_w</code>	within-subject variance	$\sigma_w^2$
<code>Var_bw</code>	between–within (group-by-occasion) variance	$\sigma_{bw}^2$
<code>Var_be</code>	between-subjects error variance	$\sigma_{be}^2$
<code>Var_we</code>	within-subject error variance	$\sigma_{we}^2$
<code>Var_bwe</code>	between–within (group-by-occasion) error variance	$\sigma_{bwe}^2$
<code>Var_e</code>	error variance	$\sigma_e^2$
<code>corr</code>	correlation between repeated measures	$\rho$
<code>grwgt#</code>	group weight #	$w_{\#}$
<code>target</code>	target parameter; synonym for target effect variance	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if specified.

Column `N_per_group` is available and is shown in the default table only for balanced designs.

Columns `N_avg` and `N#` are shown in the default table only for unbalanced designs.

Columns `m#1_#2` are not shown in the default table.

Columns `Var_b` and `Var_be` are shown in the default table for the between-subjects test, `Var_w` and `Var_we` for the within-subjects test, and `Var_bw` and `Var_bwe` for the between–within test.

Columns `grwgt#` are not shown in the default table.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**.

`npergroup(numlist)` specifies the group size. Only positive integers are allowed. This option implies a balanced design. `npergroup()` cannot be specified with `n()`, `n#()`, or `grweights()`.

`n#(numlist)` specifies the size of the #th group. Only positive integers are allowed. All group sizes must be specified. For example, all three options `n1()`, `n2()`, and `n3()` must be specified for a design with three groups. `n#()` cannot be specified with `n()`, `npergroup()`, or `grweights()`.

`grweights(wgtspec)` specifies  $J$  group weights for an unbalanced design. The weights may be specified either as a list of values or as a matrix, and multiple sets of weights are allowed; see [wgtspec](#) for details. The weights must be positive and must also be integers unless the `nfractional` option is specified. `grweights()` cannot be specified with `npergroup()` or `n#()`.

`ngroups(#)` specifies the number of groups. This option is required if `meanspec` is not specified.

This option is also required for effect-size determination unless `grweights()` is specified. For a one-way repeated-measures ANOVA, specify `ngroups(1)`.

`nrepeated(#)` specifies the number of repeated measurements within each subject. At least two repeated measurements must be specified. This option is required if the `corr()` option is specified and `meanspec` is not specified. This option is also required for effect-size determination unless `covmatrix()` is specified.

`corr(numlist)` specifies the correlation between repeated measurements. `corr()` cannot be specified with `covmatrix()`. This option requires the `nrepeated()` option unless `meanspec` is specified.

`covmatrix(matname)` specifies the covariance matrix between repeated measurements. `covmatrix()` cannot be specified with `corr()` or `varerror()`.

`factor(between | within | bwithin)` specifies the effect of interest for which power and sample-size analysis is to be performed. For a one-way repeated-measures ANOVA, only `factor(within)` is allowed and is implied when only one group is specified. In a two-way repeated-measures ANOVA, the tested effects include the between effect or main effect of a between-subjects factor, the within effect or main effect of a within-subject factor, and the between-within effect or interaction effect of the between-subjects factor and the within-subject factor. The default for a two-way repeated design is `factor(between)`.

`vareffect(numlist)` specifies the variance explained by the tested effect specified in `factor()`. For example, if `factor(between)` is specified, `vareffect()` specifies the variance explained by the between-subjects factor. This option is required if the `factor()` option is specified and `meanspec` is not specified. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varbetween()`, `varwithin()`, or `varbwithin()` may be specified.

`varbetween(numlist)` specifies the variance explained by the between-subjects factor. This option is equivalent to specifying `factor(between)` and `vareffect(numlist)` and thus cannot be combined with `factor()`. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varbetween()`, `varwithin()`, or `varbwithin()` may be specified. This option is not allowed when only one group is specified.

`varwithin(numlist)` specifies the variance explained by the within-subject factor. This option is equivalent to specifying `factor(within)` and `vareffect(numlist)` and thus cannot be combined with `factor()`. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varbetween()`, `varwithin()`, or `varbwithin()` may be specified.

`varbwithin(numlist)` specifies the variance explained by the interaction between a between-subjects factor and a within-subject factor. This option is equivalent to specifying `factor(bwithin)` and `vareffect(numlist)` and thus cannot be combined with `factor()`. This option is not allowed with the effect-size determination. Only one of `vareffect()`, `varbetween()`, `varwithin()`, or `varbwithin()` may be specified. This option is not allowed when only one group is specified.

`varerror(numlist)` specifies the error variance if `covmatrix()` is not specified. This option is allowed only if `corr()` is specified. When `corr()` is specified, the default is `varerror(1)`.

`showmatrices` specifies that the cell-means matrix and the covariance matrix be displayed, when applicable.

`showmeans` specifies that the cell means be reported. For a text or graphical output, this option is equivalent to `showmatrices` except only the cell-mean matrix will be reported. For a tabular output, the columns containing cell means will be included in the default table.

`parallel`; see [PSS] [power](#).

## Table

`table`, `table()`, `notable`; see [PSS] **power**, **table**.

`saving()`; see [PSS] **power**.

## Graph

`graph`, `graph()`; see [PSS] **power**, **graph**. Also see the *column* table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value of the sample size for the sample-size determination or the initial value of the effect size  $\delta$  for the effect-size determination. The default uses a bisection algorithm to bracket the solution.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power repeated` but is not shown in the dialog box:

`notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power repeated*

*Computing sample size*

*Computing power*

*Computing effect size and target variance explained by the tested effect*

*Testing hypotheses about means from multiple dependent populations*

This entry describes the `power repeated` command and the methodology for power and sample-size analysis for one-way and two-way repeated-measures ANOVA. See [PSS] **intro** for a general introduction to power and sample-size analysis and [PSS] **power** for a general introduction to the `power` command using hypothesis tests.

## Introduction

Repeated-measures ANOVA models are popular among experimenters because of their increased power compared with regular ANOVA models. Repeated-measures designs allow multiple measurements on the same subject. The repeated measurements often correspond to outcomes measured over time for each subject, but they can also correspond to different measurements of the same treatment or measurements of different treatments. The key point is that multiple measurements are made on the same subject.

One example of a repeated-measures design is a longitudinal study that offers an important alternative to a cross-sectional study because of its ability to establish a temporal relationship between the treatment and the outcome. For example, patients with hypertension might be randomized to receive a new experimental drug or standard care and have their systolic blood pressure measured at baseline and each year for two years.

	Baseline	Year 1	Year 2
Old drug	145	135	130
New drug	145	130	120

What makes repeated-measures designs more powerful? Using each subject as his or her own control reduces subject-to-subject variability that is explained by anything other than the effect of the treatment under study. This may dramatically increase power for detecting the effect of the treatment of interest.

Two classes of methods can be used to analyze repeated-measures data: univariate methods such as regular  $F$  tests and multivariate methods such as Wilks's lambda test, Pillai's trace test, and the Lawley–Hotelling trace test. The multivariate methods are more flexible in terms of the assumptions about the repeated-measures covariance structure, but they have lower power than regular  $F$  tests. In this entry, we concentrate on the univariate methods.

A repeated-measures design belongs to a class of within-subject designs, designs that contain one or more within-subject factors. A within-subject factor is a factor for which each subject receives several or all levels. A between-subjects factor, on the other hand, is any factor for which each subject receives only one level. In what follows, we consider designs with one fixed within-subject factor—one-way fixed-effects repeated-measures ANOVA models—or designs with one fixed between-subjects factor and one fixed within-subjects factor—two-way fixed-effects repeated-measures ANOVA models.

In a one-way repeated-measures model, the within-subject effect is the effect of interest. In a two-way repeated-measures model, you can choose between the three effects of interest: a main between-subjects effect or the between effect, a main within-subject effect or the within effect, and an interaction effect between the between-subjects factor and within-subject factor or the between–within effect. `power repeated` provides power and sample-size computations for the tests of all of these effects.

Repeated-measures ANOVA assumes that errors are normally distributed with zero mean and constant variance. The measurements between subjects are independent, but the measurements within a subject are assumed to be correlated. The within-subject covariance matrices must be constant between groups defined by the levels of between-subjects factors. The validity of the regular  $F$  test also relies on the so-called *sphericity* assumption (or, more generally, the circularity assumption). You can think of this assumption as all differences between levels of the within-subject factor having the same variances. A special case of this assumption is compound symmetry, a less stringent assumption. A covariance matrix is said to have a compound-symmetry structure if all the variances are equal and all the covariances are equal.

The assumption of sphericity is rather restrictive. When it is violated, the distribution of the test statistic of the regular  $F$  test of within and between–within effects is no longer an exact  $F$  distribution. To compensate for this, several nonsphericity corrections such as the Greenhouse–Geisser correction or Huynh–Feldt correction for the degrees of freedom of the regular  $F$  test are proposed (for example, Geisser and Greenhouse [1958]; Huynh and Feldt [1976]).

The distribution of the test statistic under the alternative hypothesis is a noncentral  $F$  distribution for all the considered tests. Thus power is a function of the noncentrality parameter, and the noncentrality parameter is a function of the ratio of the variance of the tested effect to the comparison error variance used in the denominator of the corresponding  $F$  test. For example, for a test of the within effect, the comparison error variance is the within-effect error variance. In what follows, by comparison error variance, we will imply one of the between-effect, within-effect, or between–within-effect error variance, whichever is appropriate for the considered test. The effect size for each of the  $F$  tests is defined as the square root of the ratio of the variance of the tested effect to the comparison error variance.

This entry describes power and sample-size analysis of repeated-measures ANOVA using the univariate  $F$  test with Greenhouse–Geisser correction for the nonsphericity.

## Using power repeated

`power repeated` computes sample size, power, or effect size for one-way and two-way fixed-effects repeated-measures ANOVA models. A one-way repeated-measures ANOVA model includes one fixed within-subject factor. The supported two-way repeated-measures ANOVA includes one fixed between-subjects factor and one fixed within-subject factor. A one-way model is available as a special case of a two-way model with one group. At least one group and two repeated measures must be specified.

All computations are performed assuming a significance level of 0.05. You may change the significance level by specifying the `alpha()` option.

The computations are performed for an  $F$  test of the effect of interest. In a one-way model, the only effect of interest is a within-subject effect. In a two-way model, you can choose between the three effects of interest: between-subjects effect with `factor(between)` (the default), within-subject effect with `factor(within)`, and between-within effect with `factor(bwithin)`.

All computations require that you specify a residual covariance between repeated measures. You can either specify any unstructured covariance matrix in `covmatrix()` or specify the correlation between repeated measures in `corr()` and the error variance in `varerror()`. If `corr()` is specified, `varerror(1)` is assumed. The latter specification implies a residual covariance with compound-symmetry structure.

To compute the total sample size, you must also specify the alternative *meanspec* and, optionally, the power of the test in `power()`. The default power is set to 0.8.

To compute power, you must specify the total sample size in the `n()` option and the alternative *meanspec*.

Instead of the alternative cell means, you can specify the number of groups (rows) in the `ngroups()` option, the number of repeated measures (columns) in the `nrepeated()` option, and the variance explained by the tested effect in the `vareffect()` option when computing sample size or power. See [Alternative ways of specifying effect in \[PSS\] power twoway](#); substitute `ngroups()` for `nrows()`, `nrepeated()` for `ncols()`, `varbetween()` for `varrow()`, `varwithin()` for `varcolumn()`, and `varbwithin()` for `varrowcolumn()`. If `covmatrix()` is specified, the `nrepeated()` option is not required—the number of repeated measures is determined by the dimensionality of the specified covariance matrix.

To compute effect size, the square root of the ratio of the variance explained by the tested factor to the comparison error variance, and the target variance explained by the tested factor, in addition to the residual covariance, you must specify the total sample size in `n()`, the power in `power()`, the number of groups in `ngroups()`, and the number of repeated measures in `nrepeated()` if `corr()` is specified.

By default, all computations assume a balanced- or an equal-allocation design. You can use `grweights()` to specify an unbalanced design for power, sample-size, or effect-size computations. For power and effect-size computations, you can specify individual group sizes in `n1()`, `n2()`, and so on, instead of a combination of `n()` and `grweights()` to accommodate an unbalanced design. For a balanced design, you can also specify `npergroup()` to specify a group size instead of a total sample size in `n()`.

In repeated-measures ANOVA, sample size and effect size depend on the noncentrality parameter of the  $F$  distribution, and their estimation requires iteration. The default initial values are obtained from a bisection search that brackets the solution. If you desire, you may change this by specifying your own value in the `init()` option. See [\[PSS\] power](#) for the descriptions of other options that control the iteration procedure.



## Computing sample size

To compute sample size, you must specify a repeated-measures covariance, an alternative cell means, or the variance of the tested effect and, optionally, the power of the test in the `power()` option. A power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a one-way repeated-measures ANOVA

Consider a version of the study described in [Winer, Brown, and Michels \(1991, 228\)](#). Suppose that researchers would like to conduct a similar study to investigate the effects of three drugs on reaction time to a series of standardized tasks. Per design, each subject will receive all three drugs, and a subject's score (mean reaction time to a task) will be recorded for each of the three drugs; that is, there will be three repeated measurements on each subject. This is a simple one-way repeated-measures design in which drug is the within-subject factor. See [Winer, Brown, and Michels \(1991\)](#) for other details of the design.

Before conducting the study, researchers would like to compute the required sample size to detect the effect of interest with 80% power and a 5% significance level. Suppose that the postulated means for the three drug levels are 26.4, 25.6, and 21; the correlation between repeated measurements is 0.6; and the error variance is 77. We use `power repeated` to compute the sample size:

```
. power repeated 26.4 25.6 21, corr(0.6) varerror(77)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.7426
      N_g =         1
      N_rep =         3
      means = <matrix>
      Var_w =    5.6622
      Var_we =   10.2667
      Var_e =   77.0000
      rho =    0.6000
Estimated sample sizes:
      N =          20
      N per group =    20
```

We need to recruit 20 subjects to detect the effect size of  $0.7426 = \sqrt{5.6622/10.2667}$  in this study.

Repeated-measures covariance in this study has a compound-symmetry structure by design, so the assumption of sphericity, underlying the  $F$  test of means for the within-subject factor, is automatically satisfied. Thus no correction to the degrees of freedom of the test is made.

◀

### ► Example 2: Alternative ways of specifying effect and repeated-measures covariance

Instead of specifying the alternative cell means as in [example 1](#), we can specify the variance between them. Here we also need to specify the number of groups and the number of repeated measures. From [example 1](#), the variance between the means was computed as 5.6622. We specify this value in `varwithin()`, the number of groups in `ngroups()`, and the number of repeated measures in `nrepeated()`:

```
. power repeated, ngroups(1) varwithin(5.6622) nrepeated(3) corr(0.6) varerror(77)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.7426
      N_g =           1
      N_rep =           3
      Var_w =    5.6622
      Var_we =   10.2667
      Var_e =   77.0000
      rho =    0.6000
Estimated sample sizes:
      N =           20
      N per group =    20
```

We obtain the exact same results as in [example 1](#).

Instead of specifying alternative means directly following the command line, we can define a matrix, say, *M*, containing these means and use it with `power repeated`:

```

. matrix M = (26.4,25.6,21)
. power repeated M, corr(0.6) varerror(77) showmatrices
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7426
    N_g =      1
    N_rep =     3
    means = <matrix>
    Var_w =    5.6622
    Var_we =   10.2667
    Var_e =    77.0000
    rho =     0.6000

```

Study matrices:

Cell means

		repeated		
		1	2	3
groups	1	26.4	25.6	21

Covariance

		repeated		
		1	2	3
repeated	1	77		
	2	46.2	77	
	3	46.2	46.2	77

Estimated sample sizes:

```

    N = 20
    N per group = 20

```

We used the `showmatrices` option to display the cell-means matrix and the covariance matrix.

We can also use the `covmatrix()` option to specify the repeated-measures covariance matrix. This option allows you to specify unstructured covariance matrices.

We could have typed the values of the covariance matrix displayed above, but instead, we simply retrieve it from the stored result `r(Cov)`. We then display the values of the covariance matrix to verify that we have the correct matrix.

```

. matrix Cov = r(Cov)
. matlist Cov

```

		repeated		
		1	2	3
repeated	1	77		
	2	46.2	77	
	3	46.2	46.2	77

We specify the covariance matrix in `covmatrix()`:

```
. power repeated M, covmatrix(Cov)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.7426
      N_g =      1
      N_rep =     3
      means = <matrix>
      Var_w =    5.6622
      Var_we =  10.2667
      Cov = <matrix>
      spherical = true
Estimated sample sizes:
      N =      20
      N per group = 20
```

We obtain the exact same results as before.

◀

### ▶ Example 3: Sample size for a two-way repeated-measures ANOVA—between effect

A group of researchers would like to design a study to determine whether a new antihypertension medication is more effective than the best medication currently available. They plan their study based on the design and results of the ALLHAT clinical trial (1996, 2002). Average systolic blood pressure (SBP) is assumed to be 145 mm/Hg at baseline in both treatment groups. Using the results of the ALLHAT study, the researchers expect a mean SBP of 135 at year 1 and 130 at year 2 in the old drug group. Using the results of pilot studies, the researchers expect a mean SBP of 130 at year 1 and 120 at year 2 in the new drug group.

	Baseline	Year 1	Year 2
Old drug	145	135	130
New drug	145	130	120

There are two factors in this experiment: treatment group is the between-subjects factor, and measurement time (baseline, year 1, and year 2) is the within-subject factor. Using the ALLHAT study and the pilot data, the researchers assume that the variance of SBP will be 225 for both groups at each of the three measurements. They also assume that the correlation between the repeated measurements is 0.7, so the covariance matrix is

$$\Sigma = \begin{bmatrix} 225 & 157.5 & 157.5 \\ 157.5 & 225 & 157.5 \\ 157.5 & 157.5 & 225 \end{bmatrix}$$

There are potentially three tests of interest here: the test of the main effect of treatment, the test of the main effect of time, and the test of the interaction effect between treatment and time.

Let's compute the required sample size for the test of the between effect, treatment. This is the default test in `power repeated` when there is more than one group.

We begin by defining a matrix of means and a covariance matrix.

```
. matrix M = (145,135,130\145,130,120)
. matrix Cov = (225,157.5,157.5\157.5,225,157.5\157.5,157.5,225)
```

We can use the `matlist` command to display these matrices to verify that we typed them correctly:

```
. matlist M
```

	c1	c2	c3
r1	145	135	130
r2	145	130	120

```
. matlist Cov
```

	c1	c2	c3
r1	225		
r2	157.5	225	
r3	157.5	157.5	225

For brevity, we use one of the alternative specifications from [example 2](#) to compute sample size. We specify the cell-means matrix `M` following the command name and the covariance matrix `Cov` in `covmatrix()`:

```
. power repeated M, covmatrix(Cov)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.1863
    N_g = 2
    N_rep = 3
    means = <matrix>
    Var_b = 6.2500
    Var_be = 180.0000
    Cov = <matrix>
Estimated sample sizes:
    N = 228
    N per group = 114
```

To detect the treatment effect of the specified magnitude,  $\delta = 0.1863 = \sqrt{6.25/180}$ , we need to enroll 228 subjects with 114 subjects per treatment. Note that the sphericity requirement is not needed for the  $F$  test of between effects, so no correction is done to the degrees of freedom of the test.

By default, `power repeated` does not display the specified matrices. If desired, we can use the `showmatrices` option to display them:

```
. power repeated M, covmatrix(Cov) showmatrices
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.1863
    N_g =      2
    N_rep =    3
    means = <matrix>
    Var_b =    6.2500
    Var_be = 180.0000
    Cov = <matrix>
```

Study matrices:

Cell means

		repeated		
		1	2	3
groups	1	145	135	130
	2	145	130	120

Covariance

		repeated		
		1	2	3
repeated	1	225		
	2	157.5	225	
	3	157.5	157.5	225

Estimated sample sizes:

```
    N =      228
    N per group =    114
```

Similar to the alternative specifications discussed in [example 2](#), all the specifications below will produce identical results:

```
. power repeated 145 135 130 \ 145 130 120, covmatrix(Cov)
(output omitted)
. power repeated M, corr(0.7) varerror(225)
(output omitted)
. power repeated, nrepeated(3) corr(0.7) varerror(225) ngroups(2) varbetween(6.25)
(output omitted)
```

### ▷ Example 4: Sample size for a two-way repeated-measures ANOVA—within effect

Continuing with [example 3](#), we now compute the required sample size for the test of the main effects of time, the within effects.

```
. power repeated M, covmatrix(Cov) factor(within)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    1.7392
    N_g =      2
    N_rep =     3
    means = <matrix>
    Var_w =   68.0556
    Var_we =  22.5000
    Cov = <matrix>
    spherical = true
Estimated sample sizes:
    N =      6
    N per group = 3
```

We only need a total of 6 subjects, 3 per group, to detect the within effect in this study.

We can also obtain identical results by using the following alternative specification:

```
. power repeated, covmatrix(Cov) ngroups(2) varwithin(68.0556)
(output omitted)
```

◀

### ▷ Example 5: Sample size for a two-way repeated-measures ANOVA—between–within effect

Continuing with [example 3](#), we can also compute the required sample size for the test of the between–within interaction effects, interaction between treatment and time.

```
. power repeated M, covmatrix(Cov) factor(bwithin)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between-within subjects with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.4303
    N_g =      2
    N_rep =     3
    means = <matrix>
    Var_bw =    4.1667
    Var_bwe =  22.5000
    Cov = <matrix>
    spherical = true
Estimated sample sizes:
    N =     54
    N per group = 27
```

For this test, we need a total of 54 subjects with 27 subjects per group.

If we are interested in performing all three tests (between, within, and between–within) during our analysis, we should pick the largest of the three sample sizes as our final sample size. In our examples, the largest sample size is 228 for the test of between effects.

We can also obtain results identical to the above by using the following alternative specification:

```
. power repeated, covmatrix(Cov) ngroups(2) varbwithin(4.1667)
(output omitted)
```

◀

## ▷ Example 6: Unbalanced design

Continuing with [example 2](#), suppose we anticipate that the first group will have twice as many subjects as the second group. We can accommodate this unbalanced design by specifying the corresponding group weights in `grweights()`:

```
. power repeated M, covmatrix(Cov) grweights(2 1)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.1757
      N_g = 2
      N_rep = 3
      means = <matrix>
      Var_b = 5.5556
      Var_be = 180.0000
      Cov = <matrix>
Estimated sample sizes:
      N = 258
      Average N = 129.0000
      N1 = 172
      N2 = 86
```

The required total sample size for this unbalanced design is 258 with 172 subjects in the first group and 86 subjects in the second group. The average number of subjects per group is 129.

We can compute results for multiple sets of group weights. The specification of group weights within `grweights()` is exactly the same as the specification of group means described in [Alternative ways of specifying effect](#). Suppose that we would like to compute sample sizes for two unbalanced designs. The first design has twice as many subjects in the first group, and the second design has twice as many subjects in the second group. We specify multiple group weights for the first and second groups in parentheses. We also specify the `parallel` option to treat multiple weight values in parallel instead of computing results for all possible combinations of these values that would have been done by default.



```
. local columns alpha power N N1 N2 grwgt1 grwgt2 delta N_rep Var_b Var_be
. power repeated M, covmatrix(Cov) grweights((2 1) (1 2)) parallel
> table('columns', formats("%6.0g"))
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0
means = <matrix>
Cov = <matrix>
```

alpha	power	N	N1	N2	grwgt1	grwgt2	delta	N_rep	Var_b	Var_be
.05	.8	258	172	86	2	1	.1757	3	5.556	180
.05	.8	258	86	172	1	2	.1757	3	5.556	180

The default table does not include group weights, so we request a table with custom columns containing group weights via `table()`. We also request a smaller format to make the table more compact.

◀

## Computing power

To compute power, you must specify a repeated-measures covariance, the total sample size in `n()`, and the alternative cell means or the variance of the tested effect.

### ▶ Example 7: Power for a two-way repeated-measures ANOVA

The team discovers that they are only able to recruit a maximum of  $n = 200$  participants. They would like to calculate the statistical power for the between-subjects effect given this constraint and assuming a balanced design.

```
. power repeated M, covmatrix(Cov) n(200)
Estimated power for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
alpha = 0.0500
N = 200
N per group = 100
delta = 0.1863
N_g = 2
N_rep = 3
means = <matrix>
Var_b = 6.2500
Var_be = 180.0000
Cov = <matrix>
Estimated power:
power = 0.7462
```

The power corresponding to this design is 75%.

◀

### ▷ Example 8: Multiple values of study parameters

Continuing with [example 7](#), suppose that the researchers would like to know whether randomizing 60% of the participants to the new drug group and 40% to the old drug group will have an effect on statistical power. For comparison, we will also include the results from a balanced design.

To accommodate this unbalanced design, we could use `grweights()`, as we demonstrated in [example 6](#). For variety, we instead use `n1()` and `n2()` to specify unequal group sizes directly. We also display only a subset of table columns, including power and sample sizes.

```
. power repeated M, covmat(Cov) n1(100 80) n2(100 120) parallel
> table(power N1 N2 N)

Estimated power for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0

means = <matrix>
Cov = <matrix>
```

power	N1	N2	N
.7462	100	100	200
.7289	80	120	200

For the specified unbalanced design, the power decreases slightly to 73% from 75%.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).



## Computing effect size and target variance explained by the tested effect

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, repeated-measures covariance, power, sample size, the numbers of groups, and possibly the number of repeated measurements must be specified.

The effect size in `power repeated` is defined as a square root of the ratio of the variance explained by the tested effect to the comparison error variance. The effect size and the target variance explained by the tested effect are computed.

### ▷ Example 9: Effect size for a two-way repeated-measures ANOVA

Continuing with [example 7](#), suppose that researchers would like to know how large the between-subjects variance must be to achieve a power of 80% with a total sample size of 200 using a balanced design.

```

. power repeated, covmat(Cov) n(200) power(0.8) ngroups(2)
Performing iteration ...
Estimated between-subjects variance for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      200
N per group =    100
       N_g =       2
      N_rep =       3
     Var_be = 180.0000
         Cov = <matrix>
Estimated effect size and between-subjects variance:
      delta =    0.1991
     Var_b =    7.1331

```

We see that to achieve a power of at least 80%, the between-subjects variance must increase to 7.1331 from 6.250, which achieved a power of 0.7462 in example 7. The effect size increases from 0.1863 to 0.1991.

◀

## Testing hypotheses about means from multiple dependent populations

After the data are collected, we can use Stata's `anova` command, for example, to perform inference for repeated-measures ANOVA. We show a quick example of how to do this here; see [R] [anova](#) for more examples and details.

### ▶ Example 10: One-way repeated-measures ANOVA

Suppose that researchers conduct their study and collect the data. Consider the data from [Winer, Brown, and Michels \(1991, 228\)](#), a version of which was discussed in [example 1](#).

`t43.dta` contains 20 observations of scores of 4 repeated measurements identified by the `drug` variable from 5 people identified by the `person` variable. We use the `anova` command to fit a one-way repeated-measures model to these data.

```
. use http://www.stata-press.com/data/r14/t43
(T4.3 -- Winer, Brown, Michels)
```

```
. anova score person drug, repeated(drug)
```

	Number of obs =	20	R-squared =	0.9244	
	Root MSE =	3.06594	Adj R-squared =	0.8803	
Source	Partial SS	df	MS	F	Prob>F
Model	1379	7	197	20.96	0.0000
person	680.8	4	170.2	18.11	0.0001
drug	698.2	3	232.73333	24.76	0.0000
Residual	112.8	12	9.4		
Total	1491.8	19	78.515789		

```
Between-subjects error term: person
                             Levels: 5      (4 df)
Lowest b.s.e. variable: person
```

```
Repeated variable: drug
```

```
Huynh-Feldt epsilon = 1.0789
*Huynh-Feldt epsilon reset to 1.0000
Greenhouse-Geisser epsilon = 0.6049
Box's conservative epsilon = 0.3333
```

Source	df	F	Prob > F			
			Regular	H-F	G-G	Box
drug	3	24.76	0.0000	0.0000	0.0006	0.0076
Residual	12					

We are interested in the test of the effect of *drug*. The regular  $F$  test reports a significant result. The *anova* output for the repeated variable *drug*, however, indicates that the sphericity assumption is not met in these data; for example, the Greenhouse–Geisser epsilon of 0.6049 is different from 1.

When the sphericity assumption is not met, the degrees of freedom of a regular  $F$  test must be adjusted. Even after the adjustment, the effect of a *drug* is still significant according to all tests, at least at the 1% level.

To design a new study based on the results of this experiment, we can use *power repeated* to compute the required sample size. To perform this computation, we will need the estimates of the repeated-measures covariance and within-drug score means.

*anova* saves the estimated repeated-measures covariance in *e(Srep)*. We save it to a new matrix *Cov* and display it:

```
. mat Cov = e(Srep)
. matlist Cov
```

	c1	c2	c3	c4
r1	76.8			
r2	53.2	42.8		
r3	29.2	15.8	14.8	
r4	69	47	27	64

We now use the *mean* command to estimate means for each of the four drug levels. We store the resulting matrix of means in *M*:

```
. mean score, over(drug)
Mean estimation          Number of obs   =          20
      1: drug = 1
      2: drug = 2
      3: drug = 3
      4: drug = 4
```

Over	Mean	Std. Err.	[95% Conf. Interval]	
score				
1	26.4	3.919184	18.19705	34.60295
2	25.6	2.925748	19.47634	31.72366
3	15.6	1.720465	11.99903	19.20097
4	32	3.577709	24.51177	39.48823

```
. mat M = e(b)
```

We now specify the obtained matrices with power repeated to compute the sample size:

```
. power repeated M, covmatrix(Cov)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0 versus Ha: delta != 0
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 3.8543
      N_g = 1
      N_rep = 4
      means = <matrix>
      Var_w = 34.9100
      Var_we = 2.3500
      Cov = <matrix>
      spherical = false
Estimated sample sizes:
      N = 4
      N per group = 4
```

We only need 4 subjects to detect the effect of a drug in a study with 80% power and a 5% significance level.

## Stored results

`power repeated` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N_avg)</code>	average sample size
<code>r(N#)</code>	number of subjects in group #
<code>r(N_per_group)</code>	number of subjects per group
<code>r(N_g)</code>	number of groups
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(balanced)</code>	1 for a balanced design; 0 otherwise
<code>r(grwgt#)</code>	group weight #
<code>r(N_rep)</code>	number of rows
<code>r(m#<sub>1</sub>_<sub>2</sub>)</code>	cell mean ( <sub>1</sub> , <sub>2</sub> )
<code>r(Var_b)</code>	between-subjects variance
<code>r(Var_w)</code>	within-subject variance
<code>r(Var_bw)</code>	between-within subjects, interaction variance
<code>r(Var_be)</code>	between-subjects error variance
<code>r(Var_we)</code>	within-subject error variance
<code>r(Var_bwe)</code>	between-within subjects, interaction error variance
<code>r(Var_e)</code>	error variance
<code>r(spherical)</code>	1 covariance is spherical; 0 otherwise
<code>r(epsilon)</code>	nonsphericity correction
<code>r(epsilon_m)</code>	mean nonsphericity correction
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for sample size or effect size
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>repeated</code>
<code>r(factor)</code>	<code>between</code> , <code>within</code> , or <code>bwithin</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrices

<code>r(pss_table)</code>	table of results
<code>r(means)</code>	cell-means matrix
<code>r(Cov)</code>	repeated-measures covariance

## Methods and formulas

Consider a sample of  $n$  units where each observation comprises  $q$  responses based on  $p$  predictors. A general linear multivariate model can then be expressed as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (1)$$

where  $\mathbf{Y}$  is an  $n \times q$  matrix of dependent variables,  $\mathbf{X}$  is an  $n \times p$  matrix of fixed predictor variables,  $\mathbf{B}$  is a  $p \times q$  matrix of coefficients, and the error  $\mathbf{E}$  is an  $n \times q$  matrix where each row is an independent and identically distributed random variable drawn from a  $p$ -dimensional multivariate normal with mean  $\mathbf{0}$  and a variance–covariance matrix  $\Sigma$ . In our repeated measures design,  $q = K$  is the number of repeated measures within  $p = J$  treatments or groups.

For expositional purposes, consider a two-way repeated-measures design with one between-subjects factor, treatment, and one within-subject factor, time. Suppose we wish to test the effect of a treatment with three levels. The response of each individual is measured at the beginning of the experiment and at three time periods after one of the three treatments is administered. To put this into perspective, we see that  $K = 4$  is the number of repeated measures and that  $J = 3$  is the number of columns of treatment levels. We can express this model as

$$\begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & y_{1,4} \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ y_{n,1} & y_{n,2} & y_{n,3} & y_{n,4} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{n,1} & x_{n,2} \end{bmatrix} \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} & \alpha_{1,4} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} & \alpha_{2,4} \end{bmatrix} + \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \\ \cdot \\ \cdot \\ \epsilon'_n \end{pmatrix}$$

where  $y_{i,k}$  is the response of the  $i$ th individual at time period  $k = 1, 2, 3, 4$  and

$$x_{i1} = \begin{cases} 1 & \text{if subject } i \text{ received treatment 1} \\ 0 & \text{if subject } i \text{ received treatment 2} \\ -1 & \text{if subject } i \text{ received treatment 3} \end{cases}$$

$$x_{i2} = \begin{cases} 0 & \text{if subject } i \text{ received treatment 1} \\ 1 & \text{if subject } i \text{ received treatment 2} \\ -1 & \text{if subject } i \text{ received treatment 3} \end{cases}$$

represent the effects of a treatment for individual  $i$ . The elements in the coefficient matrix  $\mathbf{B}$  have the following interpretation:  $\mu_k$  is the mean-treatment response at time period  $k$ ,  $\alpha_{j,k}$  is the  $j$ th treatment effect,  $j = 1, 2$ , at time period  $k$ , and  $\alpha_{3,k} = -\alpha_{1,k} - \alpha_{2,k}$ . The treatment-by-time means are  $\boldsymbol{\mu} = \mathbf{XB}$ . The  $\epsilon_i$  are independent normal vectors of length  $K$  with mean  $\mathbf{0}$  and variance–covariance  $\Sigma$ .

Methods and formulas are presented under the following headings:

- Hypothesis testing*
- Computing power*

## Hypothesis testing

A hypothesis test for a general linear multivariate model can be formed as

$$H_0: \boldsymbol{\Theta} = \mathbf{0} \quad H_a: \boldsymbol{\Theta} \neq \mathbf{0}$$

where  $\boldsymbol{\Theta} = \mathbf{CBU}$  is a  $d_c \times d_u$  matrix with arbitrary dimensions  $d_c$  and  $d_u$  that depend on the specified contrast matrices  $\mathbf{C}$  and  $\mathbf{U}$ .  $\mathbf{C}$  is a  $d_c \times p$  matrix of full rank, where  $\text{rank}(\mathbf{C}) = d_c \leq p$ , and  $\mathbf{U}$  is a  $q \times d_u$  matrix of full rank, where  $\text{rank}(\mathbf{U}) = d_u \leq q$ .

Each row of  $\mathbf{C}$  corresponds to the row of  $\boldsymbol{\Theta}$  and forms a contrast to test the between-subjects effects. Similarly, each column of  $\mathbf{U}$  corresponds to the column of  $\boldsymbol{\Theta}$  and forms a contrast to test the within-subject effect. Together,  $\mathbf{C}$  and  $\mathbf{U}$  can also be used to test for interaction effects.

The estimates are given by

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \hat{\boldsymbol{\theta}} &= \mathbf{C}\hat{\mathbf{B}}\mathbf{U}\end{aligned}\quad (2)$$

$$\hat{\mathbf{H}} = \hat{\boldsymbol{\theta}}' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \}^{-1} \hat{\boldsymbol{\theta}} \quad (3)$$

Define  $\hat{\mathbf{E}} = \mathbf{U}'\hat{\boldsymbol{\Sigma}}\mathbf{U}(n-p)$ . Then, under the assumption of sphericity, the test statistic is given by

$$F_{C,U} = \frac{\text{tr}(\hat{\mathbf{H}})/d_c d_u}{\text{tr}(\hat{\mathbf{E}})/\{d_u(n-p)\}} \quad (4)$$

where the statistic follows an  $F$  distribution with  $d_c d_u$  numerator and  $d_u(n-p)$  denominator degrees of freedom. However, if the assumption is not met, then the test statistic follows an  $F$  distribution with  $d_c d_u \varepsilon$  numerator and  $d_u(n-p)\varepsilon$  denominator degrees of freedom, where

$$\varepsilon = \frac{\text{tr}^2(\hat{\boldsymbol{\Sigma}})}{d_u \text{tr}(\hat{\boldsymbol{\Sigma}}^2)} = \frac{\left(\sum_{k=1}^{d_u} \lambda_k\right)^2}{d_u \sum_{k=1}^{d_u} \lambda_k^2}$$

Under the alternative hypothesis, the power is obtained using a noncentral  $F$  distribution with noncentrality parameter equal to

$$\lambda = d_c d_u \varepsilon F_{C,U}$$

## Computing power

To compute power, we make conjectures about the parameters of interest,  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$ , and rewrite (2), (3), and (4) as

$$\begin{aligned}\boldsymbol{\theta} &= \mathbf{C}\mathbf{B}\mathbf{U} \\ \mathbf{H} &= n\boldsymbol{\theta}' \{ \mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}' \}^{-1} \boldsymbol{\theta} \\ &= n\mathbf{H}_* \\ \mathbf{E} &= \mathbf{U}'\boldsymbol{\Sigma}\mathbf{U}(n-p) \\ &= \boldsymbol{\Sigma}_*(n-p)\end{aligned}$$

where  $\ddot{\mathbf{X}}$  is the  $p \times p$  model matrix containing all the unique rows of  $\mathbf{X}$  in a special order, and  $\mathbf{W}$  is a diagonal matrix containing  $n_j/n$ , the sample size for the  $j$ th treatment divided by the total sample size. In our three-treatment example, the matrix  $\ddot{\mathbf{X}}$  is

$$\ddot{\mathbf{X}} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$



The  $F_{C,U}$  statistic using the parameter matrices  $\mathbf{H}$  and  $\mathbf{E}$  is

$$\begin{aligned} F_{C,U} &= \frac{\text{tr}(\mathbf{H})/d_c d_u}{\text{tr}(\mathbf{E})/\{d_u(n-p)\}} \\ &= \frac{n \text{tr}(\mathbf{H}_*)}{d_c \text{tr}(\mathbf{\Sigma}_*)} \end{aligned}$$

from which we obtain the noncentrality parameter as

$$\begin{aligned} \lambda &= d_u d_c \epsilon F_{C,U} \\ &= n \epsilon \delta^2 \end{aligned} \tag{5}$$

where the effect size  $\delta$  is defined as  $\delta = \sqrt{d_u \text{tr}(\mathbf{H}_*)/\text{tr}(\mathbf{\Sigma}_*)}$ .

The effect variance (`Var_b`, `Var_w`, or `Var_bw`) reported by `power repeated` is computed as  $\text{tr}(\mathbf{H}_*)/d_c$ . The effect error variance (`Var_be`, `Var_we`, or `Var_bwe`) is computed as  $\text{tr}(\mathbf{\Sigma}_*)/(d_c d_u)$ .

Under the alternative hypothesis, the test statistic in (4) is distributed as a noncentral  $F$  distribution with  $d_c d_u \epsilon$  numerator and  $d_u(n-p)\epsilon$  denominator degrees of freedom and noncentrality parameter  $\lambda$  from (5).

The power of the overall  $F$  test is

$$1 - \beta = F_{d_c d_u \epsilon, d_u(n-p)\epsilon, \lambda} (F_{d_c d_u \epsilon_m, d_u(n-p)\epsilon_m, 1-\alpha}) \tag{6}$$

where  $F_{\cdot, \cdot, \lambda}(\cdot)$  is the cdf of a noncentral  $F$  distribution, and  $\epsilon_m = E(\epsilon)$  is computed as described in Muller and Barton (1989, 551).

Total sample size and effect size are obtained by iteratively solving the nonlinear equation (6). When the `grweights()` option is specified, a constant multiplier  $n_c$  is computed and rounded to an integer unless the `nfractional` option is specified. The group sizes are then computed as  $\tilde{w}_j n_c$ , where  $\tilde{w}$  is a standardized weight; see *Methods and formulas* of [PSS] `power oneway` for details. The actual sample size, `N_a`, is the sum of the group sizes.

See Muller et al. (1992) for details.

## References

- ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. 2002. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *Journal of the American Medical Association* 288: 2981–2997.
- Davis, B. R., J. A. Cutler, D. J. Gordon, C. D. Furberg, J. T. Wright, Jr, W. C.ushman, R. H. Grimm, J. LaRosa, P. K. Whelton, H. M. Perry, M. H. Alderman, C. E. Ford, S. Oparil, C. Francis, M. Proschan, S. Pressel, H. R. Black, and C. M. Hawkins, for the ALLHAT Research Group. 1996. Rationale and design for the antihypertensive and lipid lowering treatment to prevent heart attack trial (ALLHAT). *American Journal of Hypertension* 9: 342–360.
- Geisser, S., and S. W. Greenhouse. 1958. An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics* 29: 885–891.
- Huynh, H., and L. S. Feldt. 1976. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics* 1: 69–82.
- Muller, K. E., and C. N. Barton. 1989. Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association* 84: 549–555.
- Muller, K. E., L. M. LaVange, S. Landesman Ramey, and C. T. Ramey. 1992. Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association* 87: 1209–1226.
- Winer, B. J., D. R. Brown, and K. M. Michels. 1991. *Statistical Principles in Experimental Design*. 3rd ed. New York: McGraw-Hill.

## Also see

[PSS] **power** — Power and sample-size analysis for hypothesis tests

[PSS] **power oneway** — Power analysis for one-way analysis of variance

[PSS] **power pairedmeans** — Power analysis for a two-sample paired-means test

[PSS] **power twoway** — Power analysis for two-way analysis of variance

[PSS] **power, graph** — Graph results from the power command

[PSS] **power, table** — Produce table of results from the power command

[PSS] **Glossary**

[R] **anova** — Analysis of variance and covariance

# Title

**power cmh** — Power and sample size for the Cochran–Mantel–Haenszel test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power cmh` computes sample size, power, or effect size (the minimum detectable odds ratio) for a Cochran–Mantel–Haenszel (CMH) test of association in stratified  $2 \times 2$  tables. The command accommodates unbalanced stratum sizes and unbalanced group sizes within each stratum.

## Quick start

Sample size for a two-sided CMH test with success probabilities in the control group of 0.2, 0.3, and 0.4 and hypothesized common odds ratio of 3.5 given default power of 0.8 and significance level  $\alpha = 0.05$

```
power cmh .2 .3 .4, oratio(3.5)
```

Sample size for an unbalanced design with numbers of subjects in the experimental group as 25% of the subjects in each stratum

```
power cmh .2 .3 .4, oratio(3.5) grratios(.25 .25 .25)
```

Sample size for a one-sided test with the continuity correction applied

```
power cmh .2 .3 .4, oratio(3.5) onesided continuity
```

Power for a total sample size of 120 subjects

```
power cmh .2 .3 .4, oratio(3.5) n(120)
```

Plot of power against total sample size for samples of 100 to 150 in increments of 10

```
power cmh .2 .3 .4, oratio(3.5) n(100(10)150) graph
```

As above, but display results in a table instead of a graph

```
power cmh .2 .3 .4, oratio(3.5) n(100(10)150)
```

Minimum detectable odds ratio with 80% power

```
power cmh .2 .3 .4, power(.8) n(120)
```

As above, but with 50% of subjects from the first stratum, 30% of subjects from the second stratum, and 20% of subjects from the third stratum

```
power cmh .2 .3 .4, power(.8) n(120) strweights(50 30 20)
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power cmh probspec, oratio(numlist) [power(numlist) options]
```

Compute power

```
power cmh probspec, oratio(numlist) n(numlist) [options]
```

Compute target odds ratio

```
power cmh probspec, n(numlist) power(numlist) [options]
```

where *probspec* is either a matrix *matname* containing the probability of a success in a control group for each stratum or a list of individual stratum probabilities:

$$p_{11} \ p_{12} \ \dots \ p_{1K}$$

$p_{1k}$ , where  $k = 1, 2, \dots, K$ , is the control-group probability of a success in the  $k$ th stratum. Each  $p_{1k}$  may be specified either as one number or as a list of values in parentheses; see [U] 11.1.8 **numlist**.

*matname* is the name of a Stata matrix with  $K$  columns containing control-group success probabilities. Multiple rows are allowed, in which case each row corresponds to a different set of  $K$  strata probabilities or, equivalently, column  $k$  corresponds to a *numlist* for the control-group success probability in the  $k$ th stratum.

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample sizes
* <u>nperstratum</u> ( <i>numlist</i> )	number of subjects per stratum; implies balanced design
* <u>n#</u> ( <i>numlist</i> )	number of subjects in stratum #
<u>strweights</u> ( <i>wgtspec</i> )	stratum weights; default is one for each stratum, meaning equal stratum sizes
* <u>grratios</u> ( <i>grspec</i> )	stratum-specific group ratios of the experimental-group size to the stratum size, $n_{2k}/n_k$
* <u>oratio</u> ( <i>numlist</i> )	common odds ratio of the experimental group to the control group; required to compute power or sample size
<u>continuity</u>	apply continuity correction; default is no continuity correction
<u>direction</u> ( <u>upper</u>   <u>lower</u> )	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<u>saving</u> ( <i>filename</i> [ , <u>replace</u> ])	save the table data to <i>filename</i> ; use <u>replace</u> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<u>init</u> (#)	initial value for the sample sizes or the effect size
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
[ <u>no</u> ] <u>log</u>	suppress or display iteration log
[ <u>no</u> ] <u>dots</u>	suppress or display iterations as dots
Reporting	
[ <u>no</u> ] <u>showgrstrsizes</u>	suppress or display group-per-stratum sizes
<u>showasmatrix</u>	display all sample sizes in a matrix
<u>notitle</u>	suppress the title

\* Starred options may be specified either as one number or as a list of values; see [U] [11.1.8 numlist](#).

`notitle` does not appear in the dialog box.

<i>wgtspec</i>	Description
<code>#1 #2 ... #K</code>	$K$ stratum weights. Weights must be positive and must be integers unless option <code>nfractional</code> is specified. Multiple values for each stratum weight $\#_k$ can be specified as a <i>numlist</i> enclosed in parentheses.
<i>matname</i>	matrix with $K$ columns containing $K$ stratum weights. Multiple rows are allowed, in which case each row corresponds to a different set of $K$ weights or, equivalently, column $k$ corresponds to a <i>numlist</i> for the $k$ th weight.

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N_per_stratum</code>	number of subjects per stratum	$N/N_s$
<code>N_avg</code>	average number of subjects per stratum	$N_{\text{avg}}$
<code>N#</code>	number of subjects in stratum #	$N_{\#}$
<code>N_per_group</code>	number of subjects per group	$N/2$
<code>G1</code>	number of subjects in control group	$G_1$
<code>G2</code>	number of subjects in experimental group	$G_2$
<code>N_per_grstr</code>	number of subjects per group and stratum	$N/(2N_s)$
<code>G1_#</code>	number of subjects in control group and stratum #	$G_{1,\#}$
<code>G2_#</code>	number of subjects in experimental group and stratum #	$G_{2,\#}$
<code>delta</code>	effect size	$\delta$
<code>N_s</code>	number of strata	$N_s$
<code>oratio</code>	odds ratio	$\theta$
<code>p1_#</code>	#th stratum control-group success probability	$p_{1,\#}$
<code>strwgt#</code>	weight for stratum #	$w_{\#}$
<code>grratio#</code>	ratio of experimental-group size to stratum size for stratum #	$G_{2,\#}/N_{\#}$
<code>target</code>	target parameter; synonym for <code>oratio</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in place of column `power` in the default table if option `beta()` is specified.

Column `N_per_stratum` is shown in the default table only for equal-strata designs; otherwise, columns `N#` are displayed.

Column `N_per_group` is shown in the default table only when group sizes are the same; otherwise, columns `G1` and `G2` are displayed.

Column `N_per_grstr` is shown in the default table only for balanced designs.

Columns `G1_#` and `G2_#` are shown in the table only if requested or if option `showgrstrsizes` is specified.

Columns `strwgt#` are shown in the default table only if option `strweights()` is specified.

Columns `grratio#` are shown in the default table only if option `grratios()` is specified.

## Options

## Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] [power](#).

`nperstratum(numlist)` specifies the stratum size. Only positive integers are allowed. This option implies a balanced-strata design with equal strata sizes. `nperstratum()` cannot be specified with `n()`, `n#()`, or `strweights()`.

`n#(numlist)` specifies the size of the *#*th stratum. Only positive integers are allowed. All stratum sizes must be specified. For example, all three options `n1()`, `n2()`, and `n3()` must be specified for a design with three strata. `n#()` cannot be specified with `n()`, `nperstratum()`, or `strweights()`.

`strweights(wgtspec)` specifies *K* stratum weights for an unequal-strata design. The weights may be specified either as a list of values or as a matrix, and multiple sets of weights are allowed; see [wgtspec](#) for details. The weights must be positive and must also be integers unless the `nfractional` option is specified. `strweights()` cannot be specified with `nperstratum()` or `n#()`.

`grratios(grspec)` specifies *K* ratios, one for each stratum, of the number of subjects in the experimental group to the number of subjects in the corresponding stratum,  $n_{2k}/n_k$ . By default, a balanced group design (or equal numbers of subjects in the control and experimental groups in each stratum) is assumed, which corresponds to setting the *K* ratios to 0.5.

*grspec* is similar to [wgtspec](#) but allows noninteger numbers.

`oratio(numlist)` specifies the alternative value of the common odds ratio of the experimental group to the control group. This option specifies the magnitude of an effect size. It is required to compute power or sample size.

`continuity` requests that the continuity correction be applied. By default, no continuity correction is applied.

`direction()`, `onesided`, `parallel`; see [PSS] [power](#).

## Table

`table`, `table()`, `notable`; see [PSS] [power, table](#).

`saving()`; see [PSS] [power](#).

## Graph

`graph`, `graph()`; see [PSS] [power, graph](#). Also see the [column](#) table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value for the estimated sample size or effect size when an iterative search is required. When computing the sample size for the two-sided test, the sample-size estimate from the one-sided test is used. The initial estimate for computing the effect size is obtained from a bisection search.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

## Reporting

`showgrstrsizes` and `noshowgrstrsizes` displays or suppresses the display of sample sizes in each group and stratum. The default for general output is to display group-per-stratum sizes in a matrix.

The default for table output is to suppress the display of group-per-stratum sizes. If you specify this option with table output, group-per-stratum sizes will be displayed in a table as columns. This option has no effect on graphical output.

`showasmatrix` requests that reported sample sizes be displayed as a matrix containing group-per-stratum sizes as cells, and total strata sizes, total group sizes, and a total sample size as marginal totals. This option is not allowed with table or graphical output.

The following option is available with `power cmh` but is not shown in the dialog box:

`notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power cmh*

*Alternative ways of specifying probabilities*

*Motivating example*

*Computing sample size*

*Computing power*

*Computing effect size*

*Testing hypotheses about association in  $2 \times 2 \times K$  tables*

This entry describes the `power cmh` command and the methodology for power and sample-size analysis for a CMH test of association in  $2 \times 2 \times K$  tables. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

Many studies are designed to ascertain the relationship between a binary exposure (exposed or unexposed) and a binary outcome (success or failure). Sometimes the relationship between the two binary variables is influenced by another variable (or variables). One way to adjust for such influence is to stratify on that variable and perform stratified analysis.

For example, a health researcher might conduct a case–control study of birth defects (cases) and mother’s medication use during pregnancy (exposure) stratified by mother’s age group. A law firm might want to know if there is a relationship between a defendant receiving the death penalty and the defendant’s race stratified by the murder victim’s race. An education researcher might like to know if students’ promotion to the next grade level is more likely after participating in a remedial reading program after stratifying by grade level.

In stratified analysis, a separate  $2 \times 2$  contingency table is constructed for each stratum  $k = 1, 2, \dots, K$ .

	Group		Total
	Control	Experimental	
Failure	$a_k$	$b_k$	$m_{0k}$
Success	$c_k$	$d_k$	$m_{1k}$
Total	$n_{1k}$	$n_{2k}$	$n_k$



The corresponding stratum-specific probabilities of success are  $\pi_{1k}$  and  $\pi_{2k}$  in the control and experimental groups, respectively, and their respective estimates are  $p_{1k} = c_k/n_{1k}$  and  $p_{2k} = d_k/n_{2k}$ . The strength of the association in each table is described by the stratum-specific odds ratio of the experimental group to control group  $\theta_k$ , estimated as  $p_{2k}/(1-p_{2k})/\{p_{1k}/(1-p_{1k})\} = (a_k d_k)/(b_k c_k)$ . An overall measure of association between the exposure groups and the outcome across all strata is formed to provide stratified-adjusted inference about the association.

The CMH test (Cochran 1954; Mantel and Haenszel 1959) is commonly used to test for association in stratified  $2 \times 2$  tables or, equivalently, in  $2 \times 2 \times K$  tables. It forms a common odds ratio  $\theta$ , which can be viewed as a weighted aggregate of stratum-specific odds ratios  $\theta_k$  to quantify the strength of the association between the exposure and outcome in such stratified analysis. The null hypothesis for this test is  $H_0: \pi_{1k} = \pi_{2k}$  for all  $k = 1, 2, \dots, K$ . This is equivalent to  $H_0: \theta_k = 1$  for all  $k$ . The alternative hypothesis is such that there is common odds ratio that is different from one, that is,  $H_a: \theta_1 = \theta_2 = \dots = \theta_K = \theta \neq 1$ . Thus, we can rewrite our hypotheses simply as  $H_0: \theta = 1$  versus the two-sided alternative  $H_a: \theta \neq 1$ , an upper one-sided alternative  $H_a: \theta > 1$ , or a lower one-sided alternative  $H_a: \theta < 1$ . Statistical significance is assessed using a weighted CMH  $\chi^2$  test statistic. Under the null, the distribution of the test statistic is approximately a  $\chi^2$  distribution with one degree of freedom. See, for example, Lachin (2011) for more details about this test.

`power cmh` provides power and sample-size computations based on the asymptotic distribution of the test statistic according to Nam (1992); see *Methods and formulas* for details.

## Using power cmh

`power cmh` computes sample size, power, or effect size (the minimum detectable odds ratio) for the CMH test of association in  $2 \times 2 \times K$  tables. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying option `alpha()`. You can specify the `onesided` option to request a one-sided test.

To compute the total sample size, you must specify the probabilities of success  $p_{1k}$  in the control group in each of the  $K$  strata following the command name; the common odds ratio in the `oratio()` option; and, optionally, the power of the test in option `power()`. The default power is set to 0.8.

To compute power, you must specify the total sample size in the `n()` option, the common odds ratio in the `oratio()` option, and the control-group success probabilities  $p_{1k}$  following the command name.

To compute effect size, the minimum detectable odds ratio, or the target odds ratio, you must specify the total sample size in `n()`; the power in `power()`; the control-group success probabilities following the command name; and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the common odds ratio is assumed to be larger than one. You can change the direction to be lower, which means that the common odds ratio is assumed to be smaller than one, by specifying the `direction(lower)` option.

There are multiple ways to specify the control-group success probabilities; see *Alternative ways of specifying probabilities*.

By default, all computations assume a design with equal numbers of subjects in each group and each stratum. To change group-specific proportions of subjects in each stratum, use the `grratios(gspec)` option. A common proportion may be used for all strata or you may specify different proportions for each stratum. Regardless of whether you choose a common proportion or choose to let the proportions vary across strata, you must specify  $K$  ratios of cases to the stratum sample size within `grratios()`.

To accommodate unequal stratum sizes for power and effect-size computations, you can specify either individual stratum sizes in options `n1()`, `n2()`,  $\dots$ , `nK()` or a combination of the total sample

size in `n()` and integer stratum weights in `strweights(wgtpec)`. For equal stratum sizes, you can also specify the `nperstratum()` option to specify a common stratum size instead of a total sample size in `n()`.

By default, all computations assume no continuity correction. Use the `continuity` option to change that.

`power cmh` reports group-per-stratum sizes as a matrix in the output. To suppress this matrix, use the `noshowgrstrsizes` option. Alternatively, for the table output, you can use the `showgrstrsizes` option to include columns containing group-per-stratum sizes in the default table.

To make the output of `power cmh` more compact, you may consider using the `showasmatrix` option to display all sample sizes in a matrix.

Sample-size determination for the two-sided test and effect-size determination require iteration. The default initial value for the sample size uses the estimate for the one-sided test. The initial estimate for the effect size is obtained using a bisection search. To specify a different starting value, you may use option `init()`. For more options used to control the iteration process, see [PSS] [power](#).

## Alternative ways of specifying probabilities

There are multiple ways in which you can supply the success probabilities in the control group to `power cmh`.

You may specify each  $p_{1k}$  following the command line as

```
power cmh p11 p12 ... p1K, ...
```

At least two probabilities must be specified.

When you have many strata, you may find it more convenient to first define a Stata matrix as a column vector and use it with `power cmh`. The dimension of the matrix must be at least 2. For example,

```
matrix define probmat = (p11, p12, ..., p1K)
```

```
power cmh probmat, ...
```

In some cases, you may wish to examine multiple control-group success probabilities in one or more strata. To do this, you can specify multiple values or a *numlist* for each of the stratum probabilities in parentheses:

```
power cmh (p11,1 p11,2 ... p11,L1) (p12,1 p12,2 ... p12,L2) ... , ...
```

Each of the *numlists* may contain different numbers of values,  $L_1 \neq L_2 \neq \dots \neq L_K$ . `power cmh` will produce results for all possible combinations of values across *numlists*. Results are presented in a table. If instead you would like to treat each specification as a separate scenario, you may specify the `parallel` option.

You can accommodate multiple sets of stratum probabilities in a matrix form by adding a row for each specification. The columns of a matrix with multiple rows correspond to  $K$  strata probabilities, and values within each column  $k$  correspond to multiple specifications of the  $k$ th stratum probability or a *numlist* for the  $k$ th stratum probability.

For example, the following two specifications for three strata with two scenarios each are the same:

```
power cmh (p11,1 p11,2) (p12,1 p12,2) (p13,1 p13,2), ...
```

and

```
matrix define probmat = (p11,1, p12,1, p13,1 \ p11,2, p12,2, p13,2)
```

```
power cmh probmat, ...
```

In the above specification, if you wish to specify a *numlist* only for the first stratum, you may define your matrix as

```
matrix define probmat = (p11,1, p12, p13 \ p11,2, ., .)
```

and the results of

```
power cmh probmat, ...
```

will be the same as the results of

```
power cmh (p11,1 p11,2) p12 p13, ...
```

In the following sections, we describe the use of `power cmh` accompanied by examples for computing sample size, power, and the minimum detectable odds ratio.

## Motivating example

Consider example 4.1 of a clinical trial in duodenal ulcers from [Lachin \(2011\)](#). [Blum \(1982\)](#) describes a hypothetical clinical trial studying the effectiveness of a new drug for treating an ulcer versus placebo. The outcome is whether an ulcer is healed or not, that is, whether the excretion of gastric juices that lead to ulceration of the duodenum is retarded. Three classes of ulcers are considered: acid-dependent (caused by excessive gastric secretion), drug-dependent (caused by excessive use of drugs), and ulcers of intermediate origin. The third class contains ulcers for which the cause is difficult to determine.

The research objective is to evaluate the effectiveness of the treatment on healing ulcers adjusted for ulcer types.

The data are summarized in three  $2 \times 2$  tables, one for each ulcer type.

```
. use http://www.stata-press.com/data/r14/ulcer
(Duodenal Ulcers Data)
. table healed treatment [fw=weight], by(ulcer) contents(freq) row col
```

Ulcer type and Healing status	Treatment		
	Placebo	Drug	Total
Acid-Dependent			
Not healed	27	26	53
Healed	20	16	36
Total	47	42	89
Drug-Dependent			
Not healed	5	3	8
Healed	4	9	13
Total	9	12	21
Intermediate			
Not healed	28	18	46
Healed	16	28	44
Total	44	46	90

The stratum-specific proportions of healed ulcers in each group are as follows:

```
. table treatment [fw=weight], by(ulcer) contents(mean healed) format(%8.3g)
```

Ulcer type and Treatment	mean(healed)
Acid-Dependent	
Placebo	.426
Drug	.381
Drug-Dependent	
Placebo	.444
Drug	.75
Intermediate	
Placebo	.364
Drug	.609

The common (marginal) odds ratio can be computed by using the `cc` command. (See [R] [epitab](#) and [Testing hypotheses about association in 2×2×K tables](#) for details about this command.)

```
. cc treatment healed [fw=weight]
```

	Exposed	Unexposed	Total	Proportion Exposed
Cases	53	47	100	0.5300
Controls	40	60	100	0.4000
Total	93	107	200	0.4650
	Point estimate		[95% Conf. Interval]	
Odds ratio	1.691489		.9299035	3.08095 (exact)
Attr. frac. ex.	.408805		-.0753804	.6754248 (exact)
Attr. frac. pop	.2166667			

```
chi2(1) = 3.40 Pr>chi2 = 0.0653
```

The estimate of the marginal odds ratio for these data is 1.691.

Suppose that we would like to conduct a similar study for evaluating a new treatment for healing ulcers. We would like to perform power and sample-size analysis for this new study. In what follows, we demonstrate how to use `power cmh` to perform these analyses, and we use the results from the current study as our pilot estimates.

## Computing sample size

To compute sample size, you must specify the control-group probability of success in each stratum following the command name; the common odds ratio in option `oratio()`; and, optionally, the power of the test in option `power()`. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-sided CMH test

Consider a study of drug effectiveness for healing ulcers from [Motivating example](#). The data are stratified by an ulcer type. We would like to conduct a new study to evaluate another ulcer treatment versus placebo. We need to compute the required sample size for this study. We use the estimates from the previous ulcer study as our pilot estimates.

The proportions of healed ulcers in the placebo (control) group for each of the three strata were estimated to be 0.426, 0.444, and 0.364, respectively. We anticipate an improvement in the new treatment, and we would like to detect an odds ratio of at least 2.5 with 80% power using a two-sided 5%-level test.

To compute the required sample size, we specify the control-group proportions after the command name and an odds ratio of 2.5 in option `oratio()`. We omit options `alpha(0.05)` and `power(0.8)`, because the specified values are their defaults.

```
. power cmh 0.426 0.444 0.364, or(2.5)
Performing iteration ...
Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
Ho: OR = 1 versus Ha: OR != 1; OR = OR_1 = OR_2 = OR_3
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 2.5000
      N_s = 3
      p1_1 = 0.4260
      p1_2 = 0.4440
      p1_3 = 0.3640
      oratio = 2.5000
Estimated sample sizes:
      N = 156
      N per stratum = 52
      N per group = 78
      N per group/stratum = 26
```

A total sample of 156 subjects (78 per group, 52 per stratum, and 26 per each group and stratum) must be obtained to detect an odds ratio of at least 2.5 with 80% power using a 5%-level two-sided CMH test.

◀

## ► Example 2: Unbalanced design, unequal stratum sizes

In [example 1](#), we assumed equal numbers of subjects in each stratum. Based on our pilot data, there are 89 subjects in stratum 1, 21 subjects in stratum 2, and 90 subjects in stratum 3. In other words, the first and the third strata have about 4 times as many subjects as the second stratum.

Let's see how unequal stratum sizes affect our required sample size. We specify strata weights in the `strweights()` option.

```
. power cmh 0.426 0.444 0.364, or(2.5) strweights(4 1 4)
Performing iteration ...
Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
Ho: OR = 1 versus Ha: OR != 1; OR = OR_1 = OR_2 = OR_3
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 2.5000
      N_s = 3
      p1_1 = 0.4260
      p1_2 = 0.4440
      p1_3 = 0.3640
      oratio = 2.5000
Estimated sample sizes:
      N = 162
      N1 = 72
      N2 = 18
      N3 = 72
      N per group = 81
Group-per-stratum sample sizes:
```

		stratum		
		1	2	3
group	1	36	9	36
	2	36	9	36

The required total sample size increases slightly from 156 to 162 with 81 subjects now in each group, which is a slightly larger number than before. Because of unbalanced stratum sizes, the number of subjects per group differs across strata with 36 subjects per group in the first and the third strata and 9 subjects per group in the second stratum.

◀

### ► Example 3: Unbalanced design, unequal group ratios

Continuing with [example 2](#) and our *Motivating example*, we notice that our pilot study also had different numbers of subjects in each group across strata. The actual ratios of the experimental group sizes to the stratum sizes were  $42/89 = 0.47$  in stratum 1,  $12/21 = 0.57$  in stratum 2, and  $46/90 = 0.51$  in stratum 3.

To accommodate unequal group sizes across strata, we can specify stratum-specific group ratios in the `grratios()` option.

```
. power cmh 0.426 0.444 0.364, or(2.5) strweights(4 1 4) grratios(0.47 0.57 0.51)
```

```
Performing iteration ...
```

```
Estimated sample size for a test of independence in stratified 2x2 tables
```

```
Cochran-Mantel-Haenszel test
```

```
Ho: OR = 1 versus Ha: OR != 1; OR = OR_1 = OR_2 = OR_3
```

```
Study parameters:
```

```
alpha = 0.0500
power = 0.8000
delta = 2.5000
N_s = 3
p1_1 = 0.4260
p1_2 = 0.4440
p1_3 = 0.3640
oratio = 2.5000
```

```
Estimated sample sizes:
```

```
N = 162
N1 = 72
N2 = 18
N3 = 72
N group 1 = 80
N group 2 = 82
```

```
Group-per-stratum sample sizes:
```

		stratum		
		1	2	3
group	1	38	7	35
	2	34	11	37

Because all the group ratios are close enough to 0.5 (equal-sized groups), the estimate of the total sample size does not change, and the group sizes change only slightly.

We can investigate the impact of group ratios that are appreciably different from 0.5. For example, for the following three group ratios, the total sample size increases to 207 with 88 subjects in the control group and 119 subjects in the experimental group.

```
. power cmh 0.426 0.444 0.364, or(2.5) strweights(4 1 4) grratios(0.8 0.7 0.3)
Performing iteration ...
Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
Ho: OR = 1 versus Ha: OR != 1; OR = OR_1 = OR_2 = OR_3
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 2.5000
      N_s = 3
      p1_1 = 0.4260
      p1_2 = 0.4440
      p1_3 = 0.3640
      oratio = 2.5000
Estimated sample sizes:
      N = 207
      N1 = 92
      N2 = 23
      N3 = 92
      N group 1 = 88
      N group 2 = 119
Group-per-stratum sample sizes:
```

		stratum		
		1	2	3
group	1	18	6	64
	2	74	17	28

◀

## Computing power

To compute power, you must specify the total sample size in the `n()` option, the common odds ratio in option `oratio()`, and the probabilities of success in the control group following the command name.

### ▷ Example 4: Power of a two-sided CMH test

Returning to [example 1](#), suppose that we have resources to recruit twice as many subjects, say, 300. To compute power, we add the total number of subjects, 300, in the `n()` option to the syntax from example 1:



```
. power cmh 0.426 0.444 0.364, or(2.5) n(300)
Estimated power for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
Ho: OR = 1 versus Ha: OR != 1; OR = OR_1 = OR_2 = OR_3
Study parameters:
      alpha = 0.0500
      N = 300
      N per stratum = 100
      N per group = 150
      N per group/stratum = 50
      delta = 2.5000
      N_s = 3
      p1_1 = 0.4260
      p1_2 = 0.4440
      p1_3 = 0.3640
      oratio = 2.5000
Estimated power:
      power = 0.9759
```

For such an increase in sample size, the power increases dramatically to 0.976 to detect an odds ratio of 2.5.

◀

### ► Example 5: Multiple values of study parameters

We may want to check powers for several sample sizes. Continuing with [example 4](#), we simply list desired sample-size values in option `n(numlist)`; see [\[U\] 11.1.8 numlist](#).

```
. power cmh 0.426 0.444 0.364, or(2.5) n(150(25)300) table(power N)
Estimated power for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
Ho: OR = 1 versus Ha: OR != 1; OR = OR_1 = OR_2 = OR_3
```

power	N
.7904	150
.8473	175
.8902	200
.9253	225
.9475	250
.9634	275
.9759	300

To shorten our default table, we selected only two varying columns, power and total sample size `N`, by specifying them within the `table()` option.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

◀

## Computing effect size

Sometimes, we may be interested in determining the smallest odds ratio that yields a statistically significant result for prespecified sample size and power. In this case, we must specify power in option `power()`, sample size in option `n()`, and control-group success probabilities following the command name. We may also choose the direction of the effect by specifying the `direction()` option. The default is `direction(upper)`, meaning a common odds ratio greater than one, which corresponds to the improvement of the experimental group over the control group. You can use `direction(lower)` to request an odds ratio less than one.

### ► Example 6: Minimum detectable odds ratio

In [example 4](#), we learned that with a sample of 300 subjects, the power to detect a common odds ratio of 2.5 is very high. We now want to identify the minimum detectable odds ratio for that study with 80% power.

```
. power cmh 0.426 0.444 0.364, power(0.8) n(300)
Performing iteration ...
Estimated odds ratio for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
Ho: OR = 1 versus Ha: OR != 1, OR > 1; OR = OR_1 = OR_2 = OR_3
Study parameters:
      alpha =      0.0500
      power =      0.8000
        N =      300
  N per stratum =      100
    N per group =      150
N per group/stratum =      50
      N_s =         3
      p1_1 =      0.4260
      p1_2 =      0.4440
      p1_3 =      0.3640
Estimated effect size and odds ratio:
      delta =      1.9192
  odds ratio =      1.9192
```

With a sample of 300 subjects, we can detect an odds ratio of 1.92 with 80% power using a two-sided 5%-level CMH test.



## Testing hypotheses about association in $2 \times 2 \times K$ tables

You can test hypotheses about association in  $2 \times 2 \times K$  tables using the `cc` command; see [\[R\] epitab](#). `cc` conducts a Mantel–Haenszel test, which is asymptotically equivalent to a Cochran test; see, for example, *Methods and formulas*.

### ► Example 7: Performing Mantel–Haenszel test

Consider the ulcer data from [Motivating example](#). We would like to test for association between the treatment and healing of an ulcer stratified by a type of ulcer.

```
. use http://www.stata-press.com/data/r14/ulcer
(Duodenal Ulcers Data)
. describe
Contains data from http://www.stata-press.com/data/r14/ulcer.dta
  obs:                12                Duodenal Ulcers Data
  vars:                4                3 Oct 2014 16:11
  size:                48
```

variable name	storage type	display format	value label	variable label
treatment	byte	%9.0g	treatlab	Treatment
healed	byte	%10.0g	heallab	Healing status
ulcer	byte	%14.0g	ulcerlab	Ulcer type
weight	byte	%9.0g		Frequency weights

Sorted by:

The binary variable `healed` indicates whether an ulcer healed. The binary variable `treatment` indicates whether a subject received a treatment (drug) or placebo. The categorical variable `ulcer` is a stratification variable, which records an ulcer type. The `weight` variable contains the numbers of subjects for each combination of values for `healed` and `treatment` for each of the three strata. This variable will be used as a frequency weight on our analysis.

Here are our data:

```
. list, sepby(ulcer) nolabel
```

	treatm~t	healed	ulcer	weight
1.	1	1	1	16
2.	1	0	1	26
3.	0	1	1	20
4.	0	0	1	27
5.	1	1	2	9
6.	1	0	2	3
7.	0	1	2	4
8.	0	0	2	5
9.	1	1	3	28
10.	1	0	3	18
11.	0	1	3	16
12.	0	0	3	28

We use the `cc` command to test for association between `healed` and `treatment` adjusted for ulcer types. We specify `healed` as the dependent variable and `treatment` as the exposure variable after the command name and use the `by(ulcer)` option to stratify on `ulcer`. We also specify the `weight` variable as the frequency weight.

```
. cc healed treatment [fweight=weight], by(ulcer)
```

Ulcer type	OR	[95% Conf. Interval]		M-H Weight
Acid-Dependent	.8307692	.32516	2.112608	5.842697 (exact)
Drug-Dependent	3.75	.4283095	35.61362	.5714286 (exact)
Intermediate	2.722222	1.069202	6.992856	3.2 (exact)
Crude	1.691489	.9299035	3.08095	(exact)
M-H combined	1.633836	.934329	2.857044	

```
Test of homogeneity (M-H)      chi2(2) =      4.58  Pr>chi2 = 0.1012
Test that combined OR = 1:
Mantel-Haenszel chi2(1) =      3.00
Pr>chi2 =      0.0830
```

The combined Mantel–Haenszel odds ratio equals 1.63, which is not significantly different from 1 at the 5% significance level; the  $p$ -value is 0.0830. Thus, we do not have sufficient evidence to reject the null hypothesis of no association between the treatment and healing of ulcer after adjusting for ulcer type, at least at the 5% significance level.

◀

## Stored results

`power cmh` stores the following in `r()`:

Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N_avg)</code>	average sample size
<code>r(N#)</code>	number of subjects in stratum #
<code>r(N_per_stratum)</code>	number of subjects per stratum
<code>r(N_s)</code>	number of strata
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(balanced)</code>	1 for a balanced design; 0 otherwise
<code>r(strwgt#)</code>	stratum weight #
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(N_per_group)</code>	number of subjects per group
<code>r(G1)</code>	number of subjects in the control group
<code>r(G2)</code>	number of subjects in the experimental group
<code>r(N_per_grstr)</code>	number of subjects per group and stratum
<code>r(G1_#)</code>	number of subjects in the control group in stratum #
<code>r(G2_#)</code>	number of subjects in the experimental group in stratum #
<code>r(grratio#)</code>	ratio of the experimental-group size to stratum size for stratum #
<code>r(p1_#)</code>	control-group probability of success in stratum #
<code>r(oratio)</code>	odds ratio of the experimental group to control group
<code>r(continuity)</code>	1 if continuity correction is used; 0 otherwise
<code>r(c)</code>	continuity-correction value
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size or effect size
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

Macros	
r(type)	test
r(method)	cmh
r(direction)	upper or lower
r(columns)	displayed table columns
r(labels)	table column labels
r(widths)	table column widths
r(formats)	table column formats
Matrix	
r(pss_table)	table of results

## Methods and formulas

Consider definitions and a  $2 \times 2$  contingency table for stratum  $k = 1, 2, \dots, K$  from [Introduction](#).

Assume a common odds ratio among all the  $K$  tables; that is,  $\theta_1 = \theta_2 = \dots = \theta_K = \theta$  ([Woolson, Bean, and Rojas 1986](#); [Nam 1992](#)). To test the hypothesis  $H_0: \theta = 1$  versus  $H_a: \theta > 1$ , Cochran's test statistic may be formed as

$$C = \frac{\sum_{k=1}^K \omega_k (p_{2k} - p_{1k})}{\sqrt{\sum_{k=1}^K \omega_k \bar{p}_k \bar{q}_k}} = \frac{W}{\sqrt{\widehat{\text{Var}}_0(W)}}$$

where

$$\omega_k = \frac{n_{1k}n_{2k}}{n_k}$$

$$\bar{p}_k = \frac{n_{1k}p_{1k} + n_{2k}p_{2k}}{n_k} \quad \text{and} \quad \bar{q}_k = 1 - \bar{p}_k$$

Under the null hypothesis, the statistic  $C$  converges to a standard normal distribution. We therefore reject the null hypothesis if  $C > z_{1-\alpha}$ , where  $z_{1-\alpha}$  is the  $(1 - \alpha)$ th quantile of a standard normal distribution.

A similar statistic is the Mantel–Haenszel statistic, which considers a hypergeometric distribution for all fixed marginals for the variance of  $W$  under the null hypothesis. Asymptotically, the Cochran test is equivalent to the Mantel–Haenszel test ([Armitage, Berry, and Matthews 2002](#)), and thus we refer to this test jointly as the Cochran–Mantel–Haenszel (CMH) test.

Following [Nam \(1992\)](#), we note that the expected value and the variance of  $W$  is given, respectively, by

$$E(W) = \sum_{k=1}^K \omega_k (\pi_{2k} - \pi_{1k})$$

$$\text{Var}(W) = \sum_{k=1}^K \omega_k^2 \left\{ \frac{\pi_{1k}(1 - \pi_{1k})}{n_{1k}} + \frac{\pi_{2k}(1 - \pi_{2k})}{n_{2k}} \right\}$$

The variance under the null hypothesis is given by

$$\text{Var}(W|H_0) = \text{Var}_0(W) = \sum_{k=1}^K \omega_k \bar{\pi}_k (1 - \bar{\pi}_k)$$

For the upper one-sided alternative hypothesis,  $H_a: \theta > 1$ , the asymptotic power is given by

$$1 - \beta \approx P(C > z_{1-\alpha}) = 1 - \Phi(U_u) \quad (1)$$

where

$$U_u = \frac{z_{1-\alpha} \sqrt{\text{Var}_0(W)} - E(W)}{\sqrt{\text{Var}(W)}}$$

and  $\Phi(\cdot)$  is the cumulative of the standard normal distribution.

For the lower one-sided alternative hypothesis,  $H_a: \theta < 1$ , the asymptotic power is given by

$$1 - \beta \approx P(C \leq z_\alpha) = \Phi(U_l) \quad (2)$$

where

$$U_l = \frac{z_\alpha \sqrt{\text{Var}_0(W)} - E(W)}{\sqrt{\text{Var}(W)}}$$

When the `onesided` option is used, the direction of the test is determined by the sign of  $E(W)$ —when  $E(W) > 0$ , then  $H_a: \theta > 1$  and when  $E(W) < 0$ , then  $H_a: \theta < 1$ .

Then, the power of the test is given by

$$\pi = \begin{cases} 1 - \Phi(U_u) & \text{for an upper one-sided test} \\ \Phi(U_l) & \text{for a lower one-sided test} \\ 1 - \Phi(U_u) + \Phi(U_l) & \text{for a two-sided test} \end{cases}$$

The two-sided test uses  $\alpha/2$  in the definitions of  $U_u$  and  $U_l$ .

The sample size for a two-sided test and the minimum detectable odds ratio  $\theta$  must be computed iteratively from the corresponding power equations. The starting value for the sample-size computation is the sample-size estimate for the one-sided test. For the minimum detectable odds ratio, the starting value is obtained from a bisection algorithm.

When strata weights  $w_k$  are specified in the `strweights()` option, a constant multiplier  $n_c$  is computed and rounded to an integer unless the `nfractional` option is specified. The stratum sizes are then computed as  $\tilde{w}_j n_c$ , where  $\tilde{w}_k$  are normalized strata weights. The actual sample size, `N_a`, is the sum of the stratum sizes. The numbers of subjects in each group and stratum are computed as  $n_k/2$  for equal-group designs and as  $n_{2k} = n_k \times \#_k$  and  $n_{1k} = n_k - n_{2k}$  when group ratios  $\#_k$ s are supplied in `gratios()`. Unless the `nfractional` option is specified, the experimental-group sizes corresponding to unbalanced strata are rounded up. For strata that have equal numbers of subjects in each group, the group-per-stratum sizes are not rounded.

Woolson, Bean, and Rojas (1986) provide a formula for the sample size that is required to attain a specific power by substituting  $U_u = z_\beta = \Phi^{-1}(\beta)$  and  $U_l = z_{1-\beta} = \Phi^{-1}(1 - \beta)$  in (1) and (2) and solving for  $n$ :

$$n = \begin{cases} \left( \frac{z_{1-\alpha}\sqrt{X} + z_{1-\beta}\sqrt{Y}}{Z} \right)^2 & \text{if } E(W) > 0 \\ \left( \frac{z_{\alpha}\sqrt{X} + z_{\beta}\sqrt{Y}}{Z} \right)^2 & \text{if } E(W) < 0 \end{cases}$$

where

$$X = \sum_{k=1}^K w_k s_k (1 - s_k) \bar{\pi}_k (1 - \bar{\pi}_k)$$

$$Y = \sum_{k=1}^K w_k s_k (1 - s_k) \{ (1 - s_k) \pi_{2k} (1 - \pi_{2k}) + s_k \pi_{1k} (1 - \pi_{1k}) \}$$

$$Z = \sum_{k=1}^K w_k s_k (1 - s_k) (\pi_{2k} - \pi_{1k})$$

and  $w_k = n_k/n$  is the fraction of the individual-stratum size to the total sample size, and  $s_k = n_{2k}/n_k$  is the fraction of the number of subjects in the experimental group to the sample size of each stratum. Sample size for the two-sided alternative hypothesis is solved iteratively.

The continuity-corrected version of  $U_u$  and  $U_l$  is

$$U_u = \frac{z_{1-\alpha}\sqrt{\text{Var}_0(W)} - E(W) + \frac{1}{2}}{\sqrt{\text{Var}(W)}} \quad \text{if } E(W) > 0$$

$$U_l = \frac{z_{\alpha}\sqrt{\text{Var}_0(W)} - E(W) - \frac{1}{2}}{\sqrt{\text{Var}(W)}} \quad \text{if } E(W) < 0$$

Nam (1992) derives the sample size for the continuity-corrected version of the CMH test,  $n_c$ , as follows:

$$n_c = \begin{cases} \frac{n}{4} \left( 1 + \sqrt{1 + \frac{2}{nZ}} \right)^2 & \text{if } E(W) > 0 \\ \frac{n}{4} \left( 1 + \sqrt{1 - \frac{2}{nZ}} \right)^2 & \text{if } E(W) < 0 \end{cases}$$

where  $n$  is the uncorrected sample size.

For the computation,  $\pi_{1k}$  and  $\pi_{2k}$  are replaced with their estimates  $p_{1k}$  and  $p_{2k}$  in the above formulas.  $p_{1k}s$  are supplied directly to `power cmh`, and  $p_{2k}s$  are computed using the specified  $p_{1k}s$  and odds ratio in `oratio()`.

## References

- Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.
- Blum, A. L. 1982. Principles for selection and exclusion. In *The Randomized Clinical Trial and Therapeutic Decisions*, ed. N. Tygstrup, J. M. Lachin, and E. Juhl, 43–58. New York: Dekker.
- Cochran, W. G. 1954. Some methods for strengthening the common chi-squared tests. *Biometrics* 10: 417–451.
- Lachin, J. M. 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.
- Mantel, N., and W. Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719–748. Reprinted in *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods*, ed. S. Greenland, pp. 112–141. Newton Lower Falls, MA: Epidemiology Resources.
- Nam, J. 1992. Sample size determination for case–control studies and the comparison of stratified and unstratified analyses. *Biometrics* 48: 389–395.
- Woolson, R. F., J. A. Bean, and P. B. Rojas. 1986. Sample size for case–control studies using Cochran’s statistic. *Biometrics* 42: 927–932.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **epitab** — Tables for epidemiologists



# Title

**power mcc** — Power analysis for matched case–control studies

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power mcc` computes sample size, power, or effect size (the minimum detectable odds ratio) for a test of association between a risk factor and a disease in  $1:M$  matched case–control studies.

## Quick start

Number of cases for a test with exposure probability for controls of 0.2 and odds ratio for exposure of 1.4 from a 1:1 matched design using default power of 0.8 and significance level  $\alpha = 0.05$

```
power mcc .2, oratio(1.4)
```

As above, but for a 1:2 matched design and compute the ratio of the number of cases for a 1:2 matched design relative to a 1:1 matched design

```
power mcc .2, oratio(1.4) m(2) compare
```

Number of cases when correlation of exposure between matched pairs is 0.3

```
power mcc .2, oratio(1.4) corr(.3)
```

Power for 500 cases and a 1:1 matched design

```
power mcc .2, oratio(1.4) n(500)
```

Plot of power against the number of cases for 450, 475, 500, 525, and 550 cases

```
power mcc .2, oratio(1.4) n(450(25)550) graph
```

Minimum detectable odds ratio with 80% power and 500 cases using a 1:1 matched design

```
power mcc .2, power(.8) n(500)
```

As above, but for an upper one-sided test

```
power mcc .2, power(.8) n(500) onesided direction(upper)
```

Same as above

```
power mcc .2, power(.8) n(500) onesided
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power mcc  $p_0$ , oratio(numlist) [power(numlist) options]
```

Compute power

```
power mcc  $p_0$ , oratio(numlist) n(numlist) [options]
```

Compute target odds ratio

```
power mcc  $p_0$ , power(numlist) n(numlist) [options]
```

where  $p_0$  is the probability of exposure among control patients.  $p_0$  must satisfy the condition  $0 < p_0 < 1$  and may be specified either as one number or as a list of values in parentheses; see [U] **11.1.8 numlist**.

<i>options</i>	Description
Main	
* <u>a</u> lpha( <i>numlist</i> )	significance level; default is alpha(0.05)
* <u>p</u> ower( <i>numlist</i> )	power; default is power(0.8)
* <u>b</u> eta( <i>numlist</i> )	probability of type II error; default is beta(0.2)
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>n</u> fractional	allow fractional sample size
* <u>o</u> ratio( <i>numlist</i> )	odds ratio of exposure in cases relative to controls; required to compute power or sample size
* <u>m</u> ( <i>numlist</i> )	number of matched controls per case; default is m(1)
<u>c</u> ompare	ratio of the required number of cases for a 1: <i>M</i> design relative to a paired 1:1 design
* <u>c</u> orr( <i>numlist</i> )	correlation of exposure between cases and controls; default is corr(0)
<u>d</u> irection( <u>u</u> pper  <u>l</u> ower)	direction of the effect for effect-size determination; default is direction( <u>u</u> pper), which means that the postulated value of the parameter is larger than the hypothesized value
<u>o</u> nesided	one-sided test; default is two sided
<u>p</u> arallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>n</u> o] <u>t</u> able [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <a href="#">power, table</a>
<u>s</u> aving( <i>filename</i> [ , replace ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>g</u> raph [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <a href="#">power, graph</a>
Iteration	
<u>i</u> nit(#)	initial value for the sample size or the effect size
<u>i</u> terate(#)	maximum number of iterations; default is iterate(500)
<u>t</u> olerance(#)	parameter tolerance; default is tolerance(1e-12)
<u>f</u> tolerance(#)	function tolerance; default is ftolerance(1e-12)
[ <u>n</u> o] <u>l</u> og	suppress or display iteration log
[ <u>n</u> o] <u>d</u> ots	suppress or display iterations as dots
<u>n</u> otitle	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] [11.1.8 numlist](#).

notitle does not appear in the dialog box.

where *tablespec* is

```
column[:label] [column[:label] [...]] [ , tableopts]
```

*column* is one of the columns defined [below](#), and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	number of cases	$N$
<code>delta</code>	effect size	$\delta$
<code>M</code>	number of matched controls	$M$
<code>F_M</code>	ratio of the number of cases with $M$ controls relative to one control	$F_M$
<code>p0</code>	probability of exposure among controls	$p_0$
<code>p1</code>	probability of exposure among cases	$p_1$
<code>oratio</code>	odds ratio	$\theta$
<code>corr</code>	correlation of exposure between cases and controls	$\rho$
<code>target</code>	target parameter; synonym for <code>oratio</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if option `beta()` is specified.

Column `F_M` is shown in the default table only if option `compare` is specified.

Column `p1` is not shown in the default table.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**. The sample size in `n()` is the number of matched case–control sets or, equivalently, the number of cases. The `nfractional` option is allowed only for sample-size determination.

`oratio(numlist)` specifies the odds ratio of exposure in cases relative to controls. This option is required for `power` or sample-size determination and may not be specified for effect-size determination.

`m(numlist)` specifies the number of matched controls per case. Only positive integers are allowed. The default is `m(1)`, which implies a paired design.

`compare` specifies that the ratio,  $F_M$ , of the required number of cases for a 1: $M$  design relative to a paired 1:1 design be computed. `compare` can be specified only when computing sample size and when a value of 2 or greater is specified in option `m()`.

`corr(numlist)` specifies the correlation coefficient for exposure  $\rho$  between matched cases and controls. `corr()` must contain numbers between  $-1$  and  $1$ . The default is `corr(0)`, meaning no correlation between matched cases and controls. This assumption may not be realistic in practice; see [example 3](#) for discussion.

`direction()`, `onesided`, `parallel`; see [PSS] **power**.

### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

## Graph

`graph`, `graph()`; see [PSS] [power](#), [graph](#). Also see the *column* table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies the initial value for the estimated sample size or effect size when an iterative search is required. When computing the sample size for the two-sided test, the closed-form sample-size computation for the one-sided test is used. The initial estimate for computing the minimum detectable odds ratio is obtained from a bisection search.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power mcc` but is not shown in the dialog box:

`notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power mcc*

*Computing sample size*

*Computing power*

*Computing target odds ratio*

*Testing hypotheses in matched case–control studies*

This entry describes the `power mcc` command and the methodology for power and sample size analysis for  $1:M$  matched case–control studies. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

Matched case–control studies investigate the relationship between disease and exposure, controlling for the effect of confounding variables. Cases are observations that have the outcome of interest; controls are observations that do not. Cases are matched to the controls on the basis of similar values of the variable or variables thought to confound the relationship between exposure and disease.

Matched case–control studies are used to investigate a variety of outcomes. A pediatrician might be interested in the relationship between low birthweight (case) and mother’s smoking status during pregnancy (exposure), where case and control mothers are matched on the basis of age, race, alcohol consumption, and history of hypertension (confounding variables). An oncologist might want to know if women with endometrial cancer are more likely to have taken estrogen, where cases and controls are matched on age, marital status, and time living in the community. A psychologist might design a study to see if suicide is more prevalent among patients who used a particular antidepressant, where cases and controls are matched on age, sex, race, severity of depression symptoms, and history of head injury.

This entry describes power and sample-size analysis for inference about correlated binary outcomes. In a  $1:M$  matched case–control study, we first randomly select cases from a population of cases and observe their exposure status. The population is then stratified by the confounding (matching) variables, and for each selected case,  $M$  matched controls are randomly selected from the same

stratum as the selected case. This creates a series of  $2 \times 2$  contingency tables within the stratum defined by the matching covariates. Each contingency table summarizes the probability of observing each exposure–outcome combination within a given stratum, and the probabilities are assumed to be equal across strata (or, equivalently, odds ratios are assumed to be equal across strata) so that for any study we need to analyze only one table.

Case	Control		Total
	Exposed	Unexposed	
Exposed	$p_{11}$	$p_{10}$	$p_1$
Unexposed	$p_{01}$	$p_{00}$	$q_1$
Total	$p_0$	$q_0$	1

The concordant probabilities lie on the diagonal;  $p_{11}$  is the probability that an exposed case subject is matched to an exposed control subject, and  $p_{00}$  is the probability that an unexposed case subject is matched to an unexposed control subject. The target parameter is the odds ratio  $\theta$  of developing the disease in exposed and unexposed subjects who have equal values of matching variables. It can be calculated from the discordant probabilities as  $p_{10}/p_{01}$ .

The probability of exposure for controls,  $p_0$ , is the probability that the sampled control subject is exposed and is simply the sum of  $p_{11}$  and  $p_{01}$ . The probability of exposure for cases,  $p_1$ , is the probability that the sampled case subject is exposed and is simply the sum of  $p_{11}$  and  $p_{10}$ .

The two-sided hypothesis test for association between disease and exposure can be formally stated in terms of the odds ratio as  $H_0: \theta = 1$  versus  $H_a: \theta \neq 1$ . We can equivalently state the hypothesis test in terms of marginal homogeneity:  $H_0: p_0 = p_1$  versus  $H_a: p_0 \neq p_1$ .

In a matched case–control study,  $n$  cases are sampled and then matched to  $M$  controls. When  $M = 1$ , a 1:1 matched design or paired design, there are  $n$  matched pairs and  $n \times 2$  total subjects. When  $M > 1$ , there are  $n$  matched sets and  $n \times (M + 1)$  total subjects. Unlike a study without matching, a matched case–control study does not have  $n \times 2$  independent observations [or  $n \times (M + 1)$  for matched designs with multiples controls].

All calculations performed by `power mcc` treat  $n$  as the relevant sample size. Throughout the remainder of this entry, when we refer to the sample size, we mean  $n$ , the number of cases and thus the number of matched pairs (or sets).

## Using `power mcc`

`power mcc` computes sample size, power, or effect size (the minimum detectable odds ratio) for 1: $M$  matched case–control studies, in which one case is matched to  $M$  controls. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

To compute sample size, you must specify the probability of exposure for the control group  $p_0$ ; the odds ratio for exposure  $\theta$  in option `oratio()`; and, optionally, the power of the test in the `power()` option. The default power is set to 0.8. The sample-size estimate returned is the number of matched pairs or, if option `m()` was specified, the number of matched sets. This is equivalent to the number of cases. Hereafter, we simply refer to the number of cases.

To compute power, you must specify the sample size in option `n()`, the probability of exposure for the control group  $p_0$ , and the odds ratio in option `oratio()`.

To compute the minimum detectable odds ratio, you must specify the sample size in option `n()`; the power in option `power()`; the probability of exposure for the control group  $p_0$ ; and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the probability of exposure among cases is assumed to be larger than the specified control-group value. You can change the direction to be lower, which means that the probability of exposure among cases is assumed to be lower than the specified control-group value, by specifying the `direction(lower)` option. `power mcc` defines the effect size as the target odds ratio.

By default, all computations assume a 1:1 or paired design, in which one case is matched to one control. You may specify the `m()` option to accommodate multiple matches per case.

The correlation between the matched case–control subjects is set to 0 by default but may be changed by specifying option `corr()`.

For sample-size determination, you can specify the `compare` option to compute the ratio of the required number of cases when using  $M$  matched controls rather than one.

Sample-size determination for the two-sided test and effect-size determination for  $M > 1$  require iteration. The default initial sample-size value is set to the closed-form one-sided sample size. The initial value for the effect size is computed using a bisection algorithm. You can use the `init()` option to specify your own value. See [PSS] `power` for a description of other options that control the iteration process.

In the following sections, we describe the use of `power mcc` accompanied by examples for computing sample size, power, and the minimum detectable odds ratio.

## Computing sample size

To compute sample size, you must specify the probability of exposure among control patients  $p_0$  after the command name; the odds ratio  $\theta$  in option `oratio()`; and, optionally, the power in option `power()`.

### ► Example 1: Sample size for a 1:1 matched case–control study

Consider a study comparing the odds of developing lung cancer among smokers compared with the odds among nonsmokers. Suppose that previous studies matching smokers and nonsmokers on the basis of age, gender, race, and alcohol consumption found the following proportions:

Lung Cancer (Case)	No Lung Cancer (Control)		Total
	Smoker	Nonsmoker	
Smoker	0.180	0.144	0.324
Nonsmoker	0.040	0.636	0.676
Total	0.220	0.780	1

If we wish to plan a new case–control study, we might assume that these proportions represent population probabilities and, therefore, let  $p_0 = 0.22$ . Under the assumption of no correlation of exposure in matched pairs,  $\theta = (0.324 \times 0.78)/(0.22 \times 0.676) = 1.7$ .

We would like to determine the number of case–control pairs that we will need to achieve 80% power to detect an odds ratio of 1.7 with a 5%-level two-sided test.

To compute the required sample size, we specify 0.22 as the probability of exposure for the control group after the command name and specify 1.7 as the odds ratio in option `oratio()`. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults.

```

. power mcc .22, oratio(1.7)
Performing iteration ...
Estimated sample size for a matched case-control study
Asymptotic z test, 1:1 matched design
Ho: OR = 1 versus Ha: OR != 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    1.7000
      p0 =     0.2200
      oratio =    1.7000
      corr =    0.0000
      M =         1
Estimated sample size:
      N cases =     285

```

Our calculation indicates that we will need a sample of 285 cases to detect an odds ratio of 1.7 with 80% power using a 5%-level test.

◀

### ▷ Example 2: Sample size for a 1: $M$ matched case–control study

Multiple controls are often matched with one case to increase the efficiency of the study. Continuing with [example 1](#), we note that we have access to many more control participants than case participants.

We specify option `m(2)` to recalculate our sample size assuming that we will match two controls with each case (a 1:2 matched design). We also specify the `compare` option to calculate the ratio of the number of required cases relative to the 1:1 paired design.

```

. power mcc .22, oratio(1.7) m(2) compare
Performing iteration ...
Estimated sample size for a matched case-control study
Asymptotic z test, 1:2 matched design
Ho: OR = 1 versus Ha: OR != 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    1.7000
      p0 =     0.2200
      oratio =    1.7000
      corr =    0.0000
      M =         2
Estimated sample size:
      N cases =     210
      F_M =     0.7368

```

We obtain a new sample-size estimate of 210 cases, and the reduction in the number of cases relative to the paired design is approximately 26% ( $1 - F\_M = 0.2632$ ).

◀



### ▷ Example 3: Sample size with correlated exposures

Matching based on confounders will typically lead to correlation of the exposure between cases and controls. For example, smoking status is known to be correlated with alcohol consumption. Matching on alcohol consumption might, therefore, result in the cases and controls being more similar with regard to smoking status. Ignoring this correlation will lead to underestimation of the required sample size or overestimation of power.

The correlation coefficient  $\rho$  for exposure between cases and controls can be computed from the probabilities in our contingency table; see (1) in *Methods and formulas*.

Returning to [example 1](#), we compute the correlation of exposure:

$$\rho = (0.180 \times 0.636 - 0.144 \times 0.040) / \sqrt{0.324 \times 0.676 \times 0.220 \times 0.780} = 0.56$$

We then specify its value in the `corr()` option.

```
. power mcc .22, oratio(1.7) corr(.56)
Performing iteration ...
Estimated sample size for a matched case-control study
Asymptotic z test, 1:1 matched design
Ho: OR = 1 versus Ha: OR != 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    1.7000
    p0 =      0.2200
    oratio =    1.7000
    corr =    0.5600
    M =        1
Estimated sample size:
    N cases =      703
```

With such a high level of correlation between exposure and the matching variables, our sample size increases dramatically to 703 cases.

This examples demonstrates the importance of taking into account the correlation of exposure between matched cases and controls. For this reason, [Dupont \(1988\)](#) recommends using the value of, say, 0.2 in computations instead of making an independence assumption, which is unlikely to hold in practice.

◀

### ▷ Example 4: Multiple values of study parameters

Continuing with [example 3](#), suppose that we believe that the previously gathered data may provide a good estimate of exposure probability for controls  $p_0$  and the odds ratio  $\theta$ , but that the individual cell proportions are not precise enough estimates of the population probabilities to produce a good estimate of  $\rho$ .

In this case, we may want to use a range of plausible values and consider how it affects the sample size for our study. We can specify a range of correlations between 0.4 and 0.6 with a step size of 0.05 using standard *numlist* (see [\[U\] 11.1.8 numlist](#)) notation in option `corr()`.

```
. power mcc .22, oratio(1.7) corr(.4(.05).6)
Performing iteration ...
Estimated sample size for a matched case-control study
Asymptotic z test, 1:1 matched design
Ho: OR = 1 versus Ha: OR != 1
```

alpha	power	N	delta	M	p0	oratio	corr
.05	.8	503	1.7	1	.22	1.7	.4
.05	.8	553	1.7	1	.22	1.7	.45
.05	.8	613	1.7	1	.22	1.7	.5
.05	.8	687	1.7	1	.22	1.7	.55
.05	.8	779	1.7	1	.22	1.7	.6

For a given power, the required sample size increases as the correlation  $\rho$  increases. In this example, the choice of the correlation has a large effect on the required sample size, which suggests that we should carefully consider the choice of matching variables.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## Computing power

To compute power, you must specify the number of cases in option `n()`, the exposure probability among controls  $p_0$  following the command name, and the odds ratio in option `oratio()`.

### ▶ Example 5: Power for matched case–control studies

Returning to [example 1](#), we discover that we are able to recruit 300 cases for our study. To compute the corresponding power, we specify 300 as the sample size in option `n()`.

```
. power mcc .22, oratio(1.7) n(300)
Estimated power for a matched case-control study
Asymptotic z test, 1:1 matched design
Ho: OR = 1 versus Ha: OR != 1
Study parameters:
    alpha =    0.0500
    N cases =     300
    delta =    1.7000
    p0 =     0.2200
    oratio =    1.7000
    corr =    0.0000
    M =         1
Estimated power:
    power =    0.8204
```

As expected, with a larger sample size, this example has a higher power (about 82%) than [example 1](#).

◀

### ► Example 6: Power for a one-sided test

Continuing with [example 5](#), suppose that we are interested in testing whether the odds ratio is greater than 1 because we hypothesize that a history of smoking will lead to an increased incidence of lung cancer. In this case, we can specify the `onesided` option to calculate power for a one-sided test.

```
. power mcc .22, oratio(1.7) n(300) onesided
Estimated power for a matched case-control study
Asymptotic z test, 1:1 matched design
Ho: OR = 1 versus Ha: OR > 1
Study parameters:
    alpha =    0.0500
    N cases =    300
    delta =    1.7000
    p0 =    0.2200
    oratio =    1.7000
    corr =    0.0000
    M =    1
Estimated power:
    power =    0.8931
```

As expected, the power of the one-sided is higher (89%) than the power of the corresponding two-sided test.

◀

## Computing target odds ratio

Sometimes, we may be interested in determining the smallest effect that will yield a statistically significant result for a prespecified sample size and power. In this case, power, sample size, and the exposure probability among controls must be specified. The effect size in `power mcc` is expressed as an odds ratio of exposure in cases relative to controls.

### ► Example 7: Minimum detectable odds ratio with 1:1 matching

Continuing with [example 5](#), we now would like to calculate the minimum detectable odds ratio that we can identify with the study design we have planned and knowing that we will be able to recruit 300 cases. We again specify 300 cases in option `n()` and also specify 80% power by using option `power(0.8)`:

```
. power mcc .22, n(300) power(.8)
Performing iteration ...
Estimated odds ratio for a matched case-control study
Asymptotic z test, 1:1 matched design
Ho: OR = 1 versus Ha: OR != 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    N cases =    300
    p0 =    0.2200
    corr =    0.0000
    M =    1
Estimated effect size and odds ratio:
    delta =    1.6783
    odds ratio =    1.6783
```

Our minimum detectable odds ratio is about 1.68 for this study design.

◀

▷ Example 8: Minimum detectable odds ratio with 1: $M$  matching

We can specify other options to tailor our estimates of the minimum detectable odds ratio to accommodate other study design parameters. Continuing with [example 7](#), we may wish to calculate the minimum detectable odds ratio if we adopt a 1:2 matched design.

```
. power mcc .22, n(300) power(.8) m(2)
Performing iteration ...
Estimated odds ratio for a matched case-control study
Asymptotic z test, 1:2 matched design
Ho: OR = 1 versus Ha: OR != 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      N cases =     300
      p0 =    0.2200
      corr =    0.0000
      M =      2
Estimated effect size and odds ratio:
      delta =    1.5656
      odds ratio = 1.5656
```

Our minimum detectable odds ratio estimate decreases to 1.57 from 1.68 when the number of cases is held constant at 300.

Comparing results with [example 7](#), we see that increasing  $M$  while holding power and sample size constant decreases the minimum detectable odds ratio. Consequently, increasing  $M$  while holding sample size and the odds ratio constant increases power.

◀

## Testing hypotheses in matched case–control studies

Matched case–control data can be organized in two ways: long and wide format. In long format, each row corresponds to a person; this is the format used by `clogit` and `mhodds` (see [\[R\] clogit](#) and [\[R\] epitab](#)). In wide format, each row corresponds to a matched pair or set; this is the format used by `mcc` (see [\[R\] epitab](#)).

▷ Example 9: Analysis of matched case–control data in long format

[Hosmer, Lemeshow, and Sturdivant \(2013\)](#) describe a study in which low birthweight infants were matched with normal weight infants on the basis of the age of the mother. The mothers were then asked whether or not they smoked during pregnancy. We will list a subset of these data to illustrate long format.

```
. use http://www.stata-press.com/data/r14/lowbirth2
(Appplied Logistic Regression, Hosmer & Lemeshow)
. list pairid low smoke in 1/6, sepby(pairid)
```

	pairid	low	smoke
1.	1	0	0
2.	1	1	1
3.	2	0	0
4.	2	1	0
5.	3	0	0
6.	3	1	0

The first column is the case–control identifier for each pair of infants. The second column identifies each infant as a case (`low==1`, indicating low birthweight) or a control (`low==0`). The third column identifies each infant as exposed (`smoke==1`, indicating that the mother smoked) or not exposed (`smoke==0`). We can estimate the odds ratio and test the null hypothesis that it equals one by using `clogit`.

We must use option `group()` to specify the identifier for our case–control matched pairs. We also specify option `nolog` to suppress the iteration log and `or` to view the result as an odds ratio.

```
. clogit low smoke, group(pairid) nolog or
```

Conditional (fixed-effects) logistic regression

```
Number of obs      =      112
LR chi2(1)         =       6.79
Prob > chi2        =      0.0091
Pseudo R2          =      0.0875
Log likelihood = -35.419282
```

	low	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
	smoke	2.75	1.135369	2.45	0.014	1.224347 6.176763

The estimated odds ratio is 2.75. We reject the null hypothesis that mothers who smoke and mothers who do not smoke have equal odds of giving birth to an infant with low birthweight at the 5%-level ( $p = 0.014$ ).

◀

### ▶ Example 10: Analysis of paired case–control data in wide format

Continuing with [example 9](#), because the [Hosmer, Lemeshow, and Sturdivant \(2013\)](#) data are for a study with matched pairs, we could also conduct a classic McNemar’s test. In Stata, McNemar’s test is calculated by the `mcc` command; see [\[R\] `epitab`](#). The `mcc` command, however, requires that the data be in wide form. For details, see the [technical note](#) in [\[R\] `clogit`](#).

We can reshape our low birthweight data using the `reshape` command.

```
. keep low smoke pairid
. reshape wide smoke, i(pairid) j(low 0 1)
Data                                long  ->  wide
-----
Number of obs.                      112  ->   56
Number of variables                  3    ->   3
j variable (2 values)                low  ->  (dropped)
xij variables:
                                smoke ->  smoke0 smoke1
```

```
. list pairid smoke1 smoke0 in 1/3
```

	pairid	smoke1	smoke0
1.	1	1	0
2.	2	0	0
3.	3	0	0

The variable `smoke1` indicates whether or not the case mother smoked, and the variable `smoke0` indicates whether or not the control mother smoked. We can now use `mcc` to estimate the overall odds ratio.

```
. mcc smoke1 smoke0
```

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	8	22	30
Unexposed	8	18	26
Total	16	40	56

```
McNemar's chi2(1) = 6.53 Prob > chi2 = 0.0106
```

```
Exact McNemar significance probability = 0.0161
```

```
Proportion with factor
```

```
Cases .5357143
```

```
Controls .2857143
```

```
[95% Conf. Interval]
```

```
difference .25 .0519726 .4480274
```

```
ratio 1.875 1.148685 3.060565
```

```
rel. diff. .35 .1336258 .5663742
```

```
odds ratio 2.75 1.179154 7.143667 (exact)
```

Again, the estimated odds ratio  $\theta$  is 2.75, and McNemar's  $\chi^2$  is 6.53. As with the analysis using `clogit`, we reject the null hypothesis of equal odds at the 5% significance level ( $p = 0.0106$ ). The confidence interval for the odds ratio calculated by `clogit` and `mcc` differ slightly due to different estimation methods.

We could also calculate the same test statistic based on symmetry and marginal homogeneity; see [\[R\] symmetry](#) for further details.

## Stored results

power mcc stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(p0)</code>	probability of exposure among controls
<code>r(M)</code>	number of matched controls per case
<code>r(F_M)</code>	ratio of the number of cases relative to the 1:1 paired design
<code>r(oratio)</code>	odds ratio
<code>r(corr)</code>	correlation of exposure between matched cases and controls
<code>r(init)</code>	initial value for the sample size or effect size
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	mcc
<code>r(direction)</code>	upper or lower
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Consider a 1: $M$  matched case–control study, where  $n$  cases with a disease are matched to  $M$  controls without the disease on the basis of similar values of confounding variables such as age, gender, etc. Some patients in the study have prior exposure to a certain risk factor of interest. Below we provide power and sample-size formulas based on Dupont (1988). Also see Breslow and Day (1980, sec. 5.3) for background on the analysis of 1: $M$  matched data.

Referring to the table in the [Introduction](#), let  $p_0$  denote the probability of exposure among the control patients. Under the null hypothesis, the odds ratio  $\theta$  of developing the disease in exposed and unexposed subjects is constant for equal values of the confounding variables.

The value of  $p_{ij}$  for each cell represents the joint probability of exposure status of the matched case–control pair for  $i, j = 0, 1$ . (With  $M$  matches, the first control is used.) Let  $p_0$  and  $p_1$  represent the probability of exposure of the control and case patient, respectively. Let  $\rho$  denote the correlation coefficient for exposure in matched pairs of case–control patients. Let  $q_0 = 1 - p_0$  and  $q_1 = 1 - p_1$ , then the odds ratio is given by  $\theta = p_1q_0/p_0q_1$  if  $\rho = 0$  and  $\theta = p_{10}/p_{01}$  if  $\rho \neq 0$ .

The correlation coefficient can be expressed as

$$\rho = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_1q_1p_0q_0}} \quad (1)$$

The individual cell probabilities can be expressed in terms of exposure probabilities as

$$\begin{aligned} p_{11} &= p_1p_0 + \rho\sqrt{p_1q_1p_0q_0} \\ p_{10} &= p_1q_0 - \rho\sqrt{p_1q_1p_0q_0} \\ p_{01} &= q_1p_0 - \rho\sqrt{p_1q_1p_0q_0} \\ p_{00} &= q_1q_0 + \rho\sqrt{p_1q_1p_0q_0} \end{aligned}$$

Let  $p_{0+}$  and  $p_{0-}$  denote the probability that a control patient is exposed given the corresponding matched case is or is not exposed, respectively. Then

$$\begin{aligned} p_{0+} &= \frac{p_{11}}{p_1} \\ p_{0-} &= \frac{p_{01}}{q_1} \\ q_{0+} &= 1 - p_{0+} \\ q_{0-} &= 1 - p_{0-} \end{aligned}$$

The probability of observing  $m$  exposed subjects among a case matched with  $M$  controls is given by

$$t_m = p_1 \binom{M}{m-1} p_{0+}^{m-1} q_{0+}^{M-m+1} + q_1 \binom{M}{m} p_{0-}^m q_{0-}^{M-m}$$

for  $m = 1, \dots, M$ .

Let  $n_{i,j}$  denote the number of matched sets such that  $n_{1,j}$  is the number of exposed cases matched with  $j$  of  $M$  exposed controls and  $n_{0,j}$  is the number of nonexposed cases matched with  $j$  of  $M$  exposed controls. Then, the number of matched sets in which  $m$  subjects were exposed is

$$T_m = n_{1,m-1} + n_{0,m}$$

Define a matched set as a discordant matched set if there is a least one discordant pair between the case and the  $M$  controls. Then, the number of discordant matched sets in which the case patients were exposed is

$$y = \sum_{m=1}^M n_{1,m-1}$$

Denote  $E_\theta$  and  $s_\theta$  the conditional mean and standard deviation of  $y$ , respectively, given  $T_m = E(T_m) = nt_m$ , for  $m = 1, \dots, M$ . Then  $E_\theta = ne_\theta$  and  $s_\theta = \sqrt{n\nu_\theta}$ , where

$$e_\theta = \sum_{m=1}^M \frac{mt_m\theta}{m\theta + M - m + 1} \quad \text{and} \quad \nu_\theta = \sum_{m=1}^M \frac{mt_m\theta(M - m + 1)}{(m\theta + M - m + 1)^2}$$

(Breslow and Day 1980, eq. 5.19).



Let  $L_\alpha = (E_1 - E_\theta - z_{1-\alpha}s_1)/s_\theta$  and  $U_\alpha = (E_1 - E_\theta + z_{1-\alpha}s_1)/s_\theta$ , where  $z_{1-\alpha}$  is the quantile of a standard normal distribution such that  $P(Z \geq z_{1-\alpha}) = \alpha$  and  $\Phi(\cdot)$  is the cumulative of a standard normal distribution.

Then, the power of the test is given by

$$1 - \beta = \begin{cases} \Phi(L_{\alpha/2}) + 1 - \Phi(U_{\alpha/2}) & \text{for a two-sided test} \\ \Phi(L_\alpha) & \text{for a lower one-sided test} \\ 1 - \Phi(U_\alpha) & \text{for an upper one-sided test} \end{cases} \quad (2)$$

The sample size for a one-sided test can be obtained from the inverse of the power computation,  $n = (z_{1-\beta}\sqrt{V_\theta} + z_{1-\alpha}\sqrt{V_1})^2/(e_1 - e_\theta)^2$ . The sample size for a two-sided test and the minimum detectable odds ratio  $\theta$  must be computed iteratively from (2). The starting value for the sample-size computation is the sample-size estimate for the one-sided test. For the minimum detectable odds ratio, the starting value is obtained from a bisection algorithm.

Let  $F_M$  denote the ratio of the sample sizes for a study with a 1: $M$  matched design relative to a study with a 1:1 matched design. If  $n_1$  is the sample size required for a study with 1 control matched to 1 case and  $n_M$  is the sample size required for a study with  $M$  controls matched to 1 case, then  $F_M = n_M/n_1$ .

## References

- Breslow, N. E., and N. E. Day. 1980. *Statistical Methods in Cancer Research: Vol. 1—The Analysis of Case–Control Studies*. Lyon: IARC.
- Dupont, W. D. 1988. Power calculations for matched case–control studies. *Biometrics* 44: 1157–1168.
- Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power pairedproportions** — Power analysis for a two-sample paired-proportions test
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **clomit** — Conditional (fixed-effects) logistic regression
- [R] **epitab** — Tables for epidemiologists
- [R] **symmetry** — Symmetry and marginal homogeneity tests

# Title

**power trend** — Power analysis for the Cochran–Armitage trend test

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power trend` computes sample size or power for the Cochran–Armitage trend test, a test for a linear trend in a probability of response in  $J \times 2$  tables. It can accommodate unbalanced designs and unequally spaced exposure levels (doses). With equally spaced exposure levels, a continuity correction is available.

## Quick start

Sample size for a test with alternative probabilities of 0.2, 0.3, and 0.4 using default power of 0.8 and significance level  $\alpha = 0.05$

```
power trend .2 .3 .4
```

As above, but for power of 0.9

```
power trend .2 .3 .4, power(.9)
```

As above, but for power of 0.7, 0.75, 0.8, 0.85, and 0.9

```
power trend .2 .3 .4, power(.7(.05).9)
```

Sample size for a one-sided test

```
power trend .2 .3 .4, onesided
```

As above, and apply continuity correction

```
power trend .2 .3 .4, onesided continuity
```

Power for a total sample size of 240 subjects

```
power trend .2 .3 .4, n(240)
```

Same as above, specified as 3 groups of 80 subjects each

```
power trend .2 .3 .4, npergroup(80)
```

Power for 100 subjects in group 1, 80 in group 2, and 60 in group 3

```
power trend .2 .3 .4, n1(100) n2(80) n3(60)
```

Graph of power against group sample size for group sizes of 70, 80, 90, and 100

```
power trend .2 .3 .4, npergroup(70(10)100) graph
```

## Menu

Statistics > Power and sample size

## Syntax

Compute sample size

```
power trend probspec [ , power(numlist) options ]
```

Compute power

```
power trend probspec, n(numlist) [options ]
```

where *probspec* is either a matrix *matname* containing group probabilities or a list of individual group probabilities:

$$p_1 \ p_2 \ [ \ p_3 \ \dots \ p_J \ ]$$

$p_j$ , where  $j = 1, 2, \dots, J$ , is the alternative group probability of observing a success for subjects with the  $j$ th level of exposure. Each  $p_j$  may be specified either as one number or as a list of values in parentheses; see [U] 11.1.8 **numlist**.

*matname* is the name of a Stata matrix with  $J$  columns containing values of alternative group probabilities. Multiple rows are allowed, in which case each row corresponds to a different set of  $J$  group probabilities or, equivalently, column  $j$  corresponds to a *numlist* for the  $j$ th group probabilities.

Alternative probabilities should be strictly monotonic: all increasing or all decreasing.

<i>options</i>	Description
Main	
* <u>a</u> lpha( <i>numlist</i> )	significance level; default is alpha(0.05)
* <u>p</u> ower( <i>numlist</i> )	power; default is power(0.8)
* <u>b</u> eta( <i>numlist</i> )	probability of type II error; default is beta(0.2)
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power
<u>n</u> fractional	allow fractional sample sizes
* <u>n</u> pergroup( <i>numlist</i> )	number of subjects per group; implies balanced design
* <u>n</u> #( <i>numlist</i> )	number of subjects in group #
<u>g</u> rweights( <i>wgtspec</i> )	group weights; default is one for each group, meaning equal group sizes
<u>e</u> xposure( <i>expospec</i> )	strictly increasing exposure levels; default is equally spaced ordinal values
<u>c</u> ontinuity	apply the continuity correction; default is no continuity correction
<u>o</u> nesided	one-sided test; default is two sided
<u>p</u> arallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>n</u> o ] <u>t</u> able [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>s</u> aving( <i>filename</i> [ , replace ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>g</u> raph [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
Iteration	
<u>i</u> nit(#)	initial value for the sample size for a two-sided test; default is to use a sample-size estimate for a one-sided test
<u>i</u> terate(#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>t</u> olerance(#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>f</u> tolerance(#)	function tolerance; default is <code>ftolerance(1e-12)</code>
[ <u>n</u> o ] <u>l</u> og	suppress or display iteration log
[ <u>n</u> o ] <u>d</u> ots	suppress or display iterations as dots
<u>n</u> otitle	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 **numlist**.

`notitle` does not appear in the dialog box.

<i>wgtspec</i>	Description
<i>#<sub>1</sub> #<sub>2</sub> ... #<sub>J</sub></i>	<i>J</i> group weights. Weights must be positive and must be integers unless option <code>nfractional</code> is specified. Multiple values for each group weight <i>#<sub>j</sub></i> can be specified as a <i>numlist</i> enclosed in parentheses.
<i>matname</i>	matrix with <i>J</i> columns containing <i>J</i> group weights. Multiple rows are allowed, in which case each row corresponds to a different set of <i>J</i> weights or, equivalently, column <i>j</i> corresponds to a <i>numlist</i> for the <i>j</i> th weight.
<i>exposspec</i>	Description
<i>#<sub>1</sub> #<sub>2</sub> ... #<sub>J</sub></i>	<i>J</i> exposure levels. By default, equally spaced exposure levels of 1, 2, ..., <i>J</i> are used. Multiple values for each exposure level <i>#<sub>j</sub></i> can be specified as a <i>numlist</i> enclosed in parentheses.
<i>matname</i>	matrix with <i>J</i> columns containing <i>J</i> exposure levels. Multiple rows are allowed, in which case each row corresponds to a different set of <i>J</i> exposure levels or, equivalently, column <i>j</i> corresponds to a <i>numlist</i> for the <i>j</i> th exposure level.

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N_per_group</code>	number of subjects per group	$N/N_g$
<code>N_avg</code>	average number of subjects per group	$N_{avg}$
<code>N#</code>	number of subjects in group #	$N_{\#}$
<code>N_g</code>	number of groups	$N_g$
<code>p#</code>	probability of outcome for group #	$p_{\#}$
<code>x#</code>	exposure level #	$x_{\#}$
<code>grwgt#</code>	group weight #	$w_{\#}$
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if option `beta()` is specified.

Column `N_per_group` is shown in the default table only for balanced designs.

Columns `N_avg` and `N#` are shown in the default table only for unbalanced designs.

Columns `x#` are shown only when exposure levels are specified using the `exposure()` option.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS] **power**.

`n#(numlist)` specifies the number of subjects in the `#`th group to be used for power determination. Only positive integers are allowed. All group sizes must be specified. `n#()` cannot be specified with `n()`, `npergroup()`, or `grweights()`.

`npergroup(numlist)` specifies the group size. Only positive integers are allowed. This option implies a balanced design. `npergroup()` cannot be specified with `n()`, `n#()`, or `grweights()`.

`grweights(wgtspec)` specifies  $J$  group weights for an unbalanced design. The weights may be specified either as a list of values or as a matrix, and multiple sets of weights are allowed; see *wgtspec* for details. The weights must be positive and must also be integers unless the `nfractional` option is specified. `grweights()` cannot be specified with `npergroup()` or `n#()`.

`exposure(expospec)` specifies the  $J$  strictly increasing exposure levels. The default is to use equally spaced values of 1, 2, ...,  $J$ .

`continuity` requests that the continuity correction be applied. This option can be specified only for equally spaced exposure levels. By default, no continuity correction is applied.

`onesided`, `parallel`; see [PSS] **power**.

### Table

`table`, `table()`, `notable`; see [PSS] **power, table**.

`saving()`; see [PSS] **power**.

### Graph

`graph`, `graph()`; see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

### Iteration

`init(#)` specifies the initial value of the sample size for the sample-size computation for a two-sided test. The default initial value is the sample size for the corresponding one-sided test.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] **power**.

The following option is available with `power trend` but is not shown in the dialog box:

`notitle`; see [PSS] **power**.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power trend*

*Alternative ways of specifying probabilities*

*Computing sample size*

*Computing power*

*Testing hypotheses about a trend in  $J \times 2$  tables*

This entry describes the `power trend` command and the methodology for power and sample-size analysis for Cochran–Armitage test for a linear trend in probability of response in  $J \times 2$  tables. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests.

## Introduction

Studies that examine the relationship between an exposure and a binary outcome have many biomedical and social science applications. When exposure can be treated as ordinal levels, researchers are often interested in whether there is a trend, or dose–response relationship, in exposure and a binary outcome. The data are typically summarized in an ordered  $J \times 2$  table,

Exposure level	Binary response	
	Success	Failure
$x_1$	$m_1$	$n_1 - m_1$
$x_2$	$m_2$	$n_2 - m_2$
$\vdots$	$\vdots$	$\vdots$
$x_J$	$m_J$	$n_J - m_J$

where  $x_j$  is an ordinal level (table score) or rank score associated with the exposure (dose) received by group  $j$  such that  $x_{j-1} < x_j$ ,  $m_j$  is the number of successes in group  $j$  and  $n_j$  is the number of subjects in group  $j$  for each  $j = 1, 2, \dots, J$ . For equally spaced exposure levels, the levels are often assigned ordinal numbers;  $x_j = j$ ,  $j = 1, 2, \dots, J$ .

A “success” simply means observing an event of interest. A dermatologist might wish to identify dosage of a topical antibiotic (ordinal exposure) necessary to cure a skin infection (binary outcome). A oncologist might conduct a case–control study to see if the number of first- and second-degree relatives with a BRCA1 gene mutation is associated with the occurrence of breast cancer. An education researcher might want to know if a higher number of unexcused absences from school is associated with failing a school grade.

This entry describes power and sample-size analysis for inference using hypothesis testing about the presence of a linear trend in probability of response in  $J \times 2$  tables. The Cochran–Armitage trend test (Cochran 1954 and Armitage 1955) is commonly used to test for trend in  $J \times 2$  tables. It is based on the linear logit model,

$$\text{logit}(p_j) = a + bx_j$$

where  $p_j$  is the hypothesized probability of a success in group  $j$ , and  $a$  and  $b$  are unknown coefficients. These group probabilities can be estimated from our contingency table as  $\hat{p}_j = m_j/n_j$ .

The null hypothesis of interest is  $H_0: p_1 = p_2 = \dots = p_J$  against the one-sided increasing-trend alternative  $H_a: p_1 < p_2 < \dots < p_J$ , the one-sided decreasing-trend alternative  $H_a: p_1 > p_2 > \dots > p_J$ , or the two-sided alternative  $H_a: p_1 < p_2 < \dots < p_J$  or  $p_1 > p_2 > \dots > p_J$ . For the linear logit model, these hypothesis are equivalent to the null  $H_0: b = 0$  against the two-sided alternative  $H_a: b \neq 0$ . If we believe that the probability of a success increases with exposure, the alternative hypothesis is upper one-sided,  $H_a: b > 0$ . If we believe that the probability of a success decreases with exposure, the alternative hypothesis is lower one-sided,  $H_a: b < 0$ . The test statistic for testing  $H_0: b = 0$  is asymptotically normal under the null hypothesis.

Power and sample-size computations are based on the asymptotic distribution of the test statistic.

## Using power trend

`power trend` computes sample size or power for a Cochran–Armitage trend test in  $J \times 2$  tables. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.

To compute the total and individual group sample sizes, you must specify the alternative probabilities of a success for  $J$  levels of exposure and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the total sample size in the `n()` option and the alternative probabilities.

There are multiple ways to specify the alternative probabilities; see [Alternative ways of specifying probabilities](#).

By default, all computations assume a balanced- or equal-allocation design. You can use the `grweights()` option to specify an unbalanced design for power or sample-size computations. For power computations, you can specify individual group sizes in options `n1()`, `n2()`,  $\dots$ , `nJ()` instead of a combination of `n()` and `grweights()` to accommodate an unbalanced design. For a balanced design, you can also specify the `npergroup()` option to specify a group size instead of a total sample size in `n()`.

Computations also assume that exposure levels are equally spaced and no continuity correction is applied. When the exposure levels are equally spaced, you can use option `continuity` to request that the continuity correction be applied. You may specify specific exposure levels in the `exposure()` option. There are multiple ways of specifying the levels; any method described in [Alternative ways of specifying probabilities](#) can also be applied to the specification of exposure levels.

Sample-size determination for a two-sided test requires iteration. The default initial values are sample-size estimates for the corresponding one-sided test. You can use the `init()` option to specify your own value. See [\[PSS\] power](#) for a description of other options that control the iteration process.

## Alternative ways of specifying probabilities

There are multiple ways in which you can supply the group probabilities of success to `power trend`.

You may specify each  $p_j$  following the command line as

```
power trend p1 p2 ... pJ [, ...]
```

At least two probabilities must be specified.

When you have many groups, you may find it more convenient to first define a Stata matrix as a column vector and use it with `power trend`. The dimension of the matrix must be at least 2. For example,

```
matrix define probmat = (p1, p2, ..., pJ)
power trend probmat [, ...]
```

In some cases, you may wish to examine multiple alternative probabilities for one or more groups. To do this, you can specify multiple values or a *numlist* for each of the group probabilities in parentheses.

```
power trend (p1,1 p1,2 ... p1,K1) (p2,1 p2,2 ... p2,K2) ... [, ...]
```



Each of the *numlists* may contain different numbers of values,  $K_1 \neq K_2 \neq \dots \neq K_J$ . `power trend` will produce results for all possible combinations of values across *numlists*. Results are presented in a table. If instead you would like to treat each specification as a separate scenario, you may specify the `parallel` option.

You can accommodate multiple sets of group probabilities in a matrix form by adding a row for each specification. The columns of a matrix with multiple rows correspond to  $J$  group probabilities, and values within each column  $j$  correspond to multiple specifications of the  $j$ th group probability of a success or a *numlist* for the  $j$ th group probability.

For example, the following two specifications for three groups defined by their exposure levels with two scenarios each are the same:

```
power trend (p1,1 p1,2) (p2,1 p2,2) (p3,1 p3,2) [, ...]
```

and

```
matrix define probmat = (p1,1, p2,1, p3,1 \ p1,2, p2,2, p3,2)
```

```
power trend probmat [, ...]
```

In the above specification, if you wish to specify a *numlist* only for the first group, you may define your matrix as

```
matrix define probmat = (p1,1, p2,1, p3,1 \ p1,2, ., .)
```

and the results of

```
power trend probmat [, ...]
```

will be the same as the results of

```
power trend (p1,1 p1,2) p2,1 p3,1 [, ...]
```

In the following sections, we describe the use of `power trend` accompanied by examples for computing sample size and power.

## Computing sample size

To compute sample size, you must specify the alternative probabilities of a success for each of the  $J$  exposure levels and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

### ► Example 1: Sample size for a two-sided trend test

Consider a study investigating the effectiveness of a new topical antibiotic for the treatment of skin infections.

Suppose that in previous studies of the treatment, we observed the following proportions of successfully treated cases at different doses. We may hypothesize that these represent the probability of a successful treatment for each dose.

Doses/day	Proportion Successes
1	0.80
2	0.85
3	0.90

We wish to determine the minimum sample size required for a clinical trial designed to detect a dose–response trend with 80% power using a two-sided 5%-level test.

To compute the required sample size, we specify the values 0.80, 0.85, and 0.90 as the alternative probabilities after the command name. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults.

```
. power trend .80 .85 .90
note: exposure levels are assumed to be equally spaced
Performing iteration ...
Estimated sample size for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b != 0; logit(p) = a + b*x
Study parameters:
      alpha =    0.0500
      power =    0.8000
      N_g   =         3
      p1    =    0.8000
      p2    =    0.8500
      p3    =    0.9000
Estimated sample sizes:
      N =      597
      N per group =    199
```

A total sample of 597 individuals, 199 individuals per group, must be obtained to detect a linear trend in probability of a successful treatment with 80% power using a two-sided 5%-level Cochran–Armitage test.

◀

### ► Example 2: Sample size for a one-sided test

Continuing with [example 1](#), suppose that the relevant research question is whether the probability of a successful treatment increases with the number of doses. In this case, we would choose to use a one-sided test because we are only interested in an increasing trend.

By specifying the `onesided` option, we obtain the sample size needed to detect a positive trend with 80% power using a one-sided 5%-level Cochran–Armitage test.

```

. power trend 0.80 0.85 0.90, onesided
note: exposure levels are assumed to be equally spaced
Estimated sample size for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b > 0; logit(p) = a + b*x
Study parameters:
    alpha =    0.0500
    power =    0.8000
    N_g =      3
    p1 =    0.8000
    p2 =    0.8500
    p3 =    0.9000
Estimated sample sizes:
    N =      471
    N per group =    157

```

Switching to a one-sided hypothesis decreased our total sample-size requirement to 471 individuals. Because the doses are equally spaced, it is possible to also perform a continuity correction. If we wanted this correction, we would have also specified the `continuity` option.

◀

### ▶ Example 3: Unbalanced design

Continuing with [example 1](#), we have assumed that the participants will be equally divided among the treatment groups. Thus, we would randomize 199 participants to each treatment level. Suppose that we instead plan to have twice as many subjects at the lowest treatment level. We can accommodate this unbalanced design by specifying the corresponding group weights in the `grweights()` option.

```

. power trend 0.80 0.85 0.90, grweights(2 1 1)
note: exposure levels are assumed to be equally spaced
Performing iteration ...
Estimated sample size for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b != 0; logit(p) = a + b*x
Study parameters:
    alpha =    0.0500
    power =    0.8000
    N_g =      3
    p1 =    0.8000
    p2 =    0.8500
    p3 =    0.9000
Estimated sample sizes:
    N =      600
    Average N = 200.0000
    N1 =      300
    N2 =      150
    N3 =      150

```

The required total sample size for this unbalanced design is 600 with 300 subjects in the group receiving a single dose per day and 150 in the groups receiving two and three doses per day. The average number of subjects per group is 200.

◀

## ▷ Example 4: Sample size for unequally spaced exposures

It is possible that the groups do not represent equally spaced exposures. Consider an example from [Agresti \(2013, 89\)](#) on the association between maternal drinking and congenital malformations, which shows data originally from [Graubard and Korn \(1987\)](#). The data are reported in ranges rather than actual values.

```
. use http://www.stata-press.com/data/r14/infants
(Congenital Malformation Data)
. list, noobs abbreviate(12)
```

	consump	cscore	infants	cases	prmalform
	0	0.0	17,114	48	0.0028
	< 1	0.5	14,502	38	0.0026
	1 to 2	1.5	793	5	0.0063
	3 to 5	4.0	127	1	0.0079
	>= 6	7.0	38	1	0.0263

Suppose we wish to use these data to design a new study to test whether there is an increasing trend in the probability of congenital malformation as the average number of alcoholic beverages consumed each week by the mother increases. We will use the `onesided` option as we did in [example 2](#) to obtain sample size with 80% power using a one-sided 5%-level Cochran–Armitage test.

To accurately compute the sample size that we will need, we should use the score associated with the reported range. [Agresti](#) recommends using the midpoint of the range for a Cochran–Armitage trend test and adopts an arbitrary value of 7 for the last interval in this case.

We can relax the assumption that the exposure levels for alcohol consumption are evenly spaced by adding the `exposure()` option. Because we have 5 exposure levels, we use a matrix specification of alternative probabilities and exposure levels. We can use the `mkmat` command to create the matrices rather than retyping the values that we already have in our dataset; see [\[P\] matrix mkmat](#).

```
. mkmat prmalform, matrix(p)
. matrix list p
p[5,1]
    prmalform
r1  .00280472
r2  .00262033
r3  .00630517
r4  .00787402
r5  .02631579
. mkmat cscore, matrix(exposed)
. matrix list exposed
exposed[5,1]
    cscore
r1      0
r2     .5
r3    1.5
r4     4
r5     7
```

Now, we specify the matrix of probabilities `p` after the command `power trend` and the matrix of exposure levels `exposed` in option `exposure()`.

```

. power trend p, onesided exposure(exposed)
Estimated sample size for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b > 0; logit(p) = a + b*x
Study parameters:
      alpha = 0.0500
      power = 0.8000
      N_g = 5
      p1 = 0.0028      x1 = 0.0000
      p2 = 0.0026      x2 = 0.5000
      p3 = 0.0063      x3 = 1.5000
      p4 = 0.0079      x4 = 4.0000
      p5 = 0.0263      x5 = 7.0000
Estimated sample sizes:
      N = 1030
      N per group = 206
Warning: Alternative probabilities are not monotonic.

```

To conduct this study, we would need to recruit a total sample of 1,030 mothers with 206 mothers at each level of consumption.

For this example, `power trend` reported a warning message that the specified alternative probabilities are not monotonic. The power for the test depends on the specific alternative, which is  $p_1 < p_2 < \dots < p_J$  in this example. Gross departures from the monotonicity assumption may lead to invalid results. In our example, the violation of this assumption is very mild—the offending probabilities are  $p_1 = 0.0028$  and  $p_2 = 0.0026$ .

◀

## Computing power

To compute power, you must specify the alternative probabilities after the command name and the total sample size in the `n()` option.

### ▶ Example 5: Power of a two-sided Cochran–Armitage trend test

Returning to [example 1](#), suppose that we anticipate obtaining a sample size of only 540 participants. To compute the corresponding power, we specify the sample size of 540 in `n()`:

```

. power trend 0.80 0.85 0.90, n(540)
note: exposure levels are assumed to be equally spaced
Estimated power for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b != 0; logit(p) = a + b*x
Study parameters:
      alpha = 0.0500
      N = 540
      N per group = 180
      N_g = 3
      p1 = 0.8000
      p2 = 0.8500
      p3 = 0.9000
Estimated power:
      power = 0.7592

```

Power decreases to 75.9% with the smaller sample of 540 subjects.

◀

### ► Example 6: Multiple values of study parameters

We may want to check powers for several sample sizes. Continuing with [example 5](#), we simply list sample-size values in the option `n(numlist)`; see [\[U\] 11.1.8 numlist](#).

```
. power trend 0.80 0.85 0.90, n(540 570 600 630 660)
> table(, labels(N_per_group "N/N_g") formats("%6.2g"))
note: exposure levels are assumed to be equally spaced

Estimated power for a trend test
Cochran-Armitage trend test
Ho: b = 0 versus Ha: b != 0; logit(p) = a + b*x
```

alpha	power	N	N/N_g	N_g	p1	p2	p3
.05	.76	540	180	3	.8	.85	.9
.05	.78	570	190	3	.8	.85	.9
.05	.8	600	200	3	.8	.85	.9
.05	.82	630	210	3	.8	.85	.9
.05	.84	660	220	3	.8	.85	.9

To shorten our default table, we specified a shorter label for the `N_per_group` column and reduced the default display format for all table columns by specifying the corresponding options within the `table()` option.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

◀

## Testing hypotheses about a trend in $J \times 2$ tables

There are several ways to conduct a trend test in Stata. We demonstrate one method here. For more examples showing the Cochran–Armitage trend test in Stata, see [Sribney \(1996\)](#).

### ► Example 7: Testing hypotheses about trends

Returning to [example 4](#), let's test whether the proportion of infants with congenital malformations increases as maternal alcohol consumption per week increases using Agresti's data. (We converted these data from wide to long as needed by further analysis.)

```
. use http://www.stata-press.com/data/r14/infants2
(Congenital Malformation Data, long form)
. list, noobs sepby(consump)
```

consump	cscore	malform	n
0	0.0	Has malformation	17066
0	0.0	No malformation	48
< 1	0.5	Has malformation	14464
< 1	0.5	No malformation	38
1 to 2	1.5	Has malformation	788
1 to 2	1.5	No malformation	5
3 to 5	4.0	Has malformation	126
3 to 5	4.0	No malformation	1
>= 6	7.0	Has malformation	37
>= 6	7.0	No malformation	1

In Stata, we can conduct a trend test by using the `tabodds` command; see [R] [epitab](#).

```
. tabodds malform cscore [fweight=n]
```

cscore	cases	controls	odds	[95% Conf. Interval]	
0	48	17066	0.00281	0.00212	0.00373
.5	38	14464	0.00263	0.00191	0.00361
1.5	5	788	0.00635	0.00263	0.01529
4	1	126	0.00794	0.00111	0.05678
7	1	37	0.02703	0.00371	0.19698

```
Test of homogeneity (equal odds): chi2(4) = 12.08
Pr>chi2 = 0.0168
```

```
Score test for trend of odds: chi2(1) = 6.57
Pr>chi2 = 0.0104
```

`tabodds` reports a  $\chi^2$  value of 6.57 for the trend test. We reject the null hypothesis of no trend at the 5% significance level; the  $p$ -value is 0.0104.

Note that the  $\chi^2$  statistic reported by `tabodds` is for a score test for trend of odds. This test is asymptotically equivalent to the Cochran–Armitage trend test. See [Lachin \(2011\)](#) and [Agresti \(2013\)](#) for details.

## Stored results

`power trend` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N_avg)</code>	average sample size
<code>r(N#)</code>	number of subjects in group #
<code>r(N_per_group)</code>	number of subjects per group
<code>r(N_g)</code>	number of groups
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(balanced)</code>	1 for a balanced design; 0 otherwise
<code>r(grwgt#)</code>	group weight #
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(p#)</code>	probability of a success in group #
<code>r(x#)</code>	exposure level for group #
<code>r(continuity)</code>	1 if continuity correction is used; 0 otherwise
<code>r(c)</code>	continuity-correction value
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for the sample size for a two-sided test
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>trend</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Assume that the probability of a success or a positive response,  $p_j$ , follows a linear trend on the logistic scale

$$p_j = \frac{e^{a+bx_j}}{1 + e^{a+bx_j}}$$

such that  $\text{logit}(p_j) = a + bx_j$ , where  $x_j$  denotes the exposure level (dose) for each of the  $J$  groups and  $x_{j-1} < x_j$  for  $j = 1, 2, \dots, J$ .  $b$  is the trend parameter about which we form our hypotheses. Under the null hypothesis,  $H_0$ :  $b = 0$ .

Power and sample-size computations for this test are based on [Nam \(1987\)](#).

Methods and formulas are presented under the following headings:

[Computing power](#)  
[Computing sample size](#)



## Computing power

Let  $n_j$  be the sample size at each exposure level for  $j = 1, 2, \dots, J$  and  $n = \sum_{j=1}^J n_j$  denote the total sample size. The observed number of successes  $m_j$  at each of the  $J$  exposure levels follows a binomial distribution; the  $m_j$ s are assumed to be independent. Let  $M = \sum_{j=1}^J m_j$  denote the total response. Then, the average response rate is given by  $\bar{p} = M/n$  and the average nonresponse rate by  $\bar{q} = 1 - \bar{p}$ .

Let  $U = \sum_{j=1}^J m_j x_j$  denote the total exposure-weighted response and  $\bar{x} = \sum_{j=1}^J n_j x_j / n$  be the average exposure level in the sample. Denote the conditional mean of  $U$  given  $M$  as  $E_0(U|M)$ , where  $E_0(U|M) = \bar{p}(\sum_{j=1}^J n_j x_j)$ . Let  $U' = U - E_0(U|M) = \sum_{j=1}^J m_j(x_j - \bar{x})$ .

Given the significance level  $\alpha$  and the probability of a type II error  $\beta$ , the power  $\pi = 1 - \beta$  is computed as

$$\pi = \begin{cases} \Phi(u_l) & \text{for a lower one-sided test} \\ 1 - \Phi(u_u) & \text{for an upper one-sided test} \\ 1 - \Phi(u_u) + \Phi(u_l) & \text{for a two-sided test} \end{cases} \quad (1)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution. For a one-sided test,

$$u_l = \frac{-E(U' + c) + z_\alpha \sqrt{\text{Var}_0(U')}}{\sqrt{\text{Var}(U')}} \quad (2a)$$

and

$$u_u = \frac{-E(U' - c) - z_\alpha \sqrt{\text{Var}_0(U')}}{\sqrt{\text{Var}(U')}} \quad (2b)$$

In (2a) and (2b),  $z_\alpha$  is the ( $\alpha$ )th quantile of the standard normal distribution and  $c = (x_{j+1} - x_j)/2$  in the presence of continuity correction or  $c = 0$  in the absence of continuity correction. The correction is not available with unequally spaced exposure levels. The power of the two-sided test uses  $\alpha/2$  to determine the value of  $z$  when computing  $u_l$  in (2a) and  $u_u$  in (2b).

Under the null hypothesis,  $p_1 = p_2 = \dots = p_J$ . Therefore, we can define a common  $p = \sum_{j=1}^J n_j p_j / N$  and  $q = 1 - p$ . The variance of  $U'$  under the null is

$$\text{Var}_0(U') = \text{Var}(U'|H_0) = pq \sum_{j=1}^J n_j (x_j - \bar{x})^2$$

and the variance of  $U'$  given  $p_1, p_2, \dots, p_J$  is

$$\text{Var}(U') = \sum_{j=1}^J n_j p_j q_j (x_j - \bar{x})^2$$

## Computing sample size

Let  $n_1$  denote the sample size of the reference group, the group with the lowest exposure level, and  $r_j = n_j/n_1$  be the known ratio of sample size of the  $j$ th group to that of the reference group. Define  $A = \sum_{j=1}^J r_j p_j (x_j - \bar{x})$ , and then sample size for a one-sided test without continuity correction is given by

$$n_1 = \frac{1}{A^2} \left[ -z_\alpha \sqrt{pq \left\{ \sum_{j=1}^J r_j (x_j - \bar{x})^2 \right\}} + z_{1-\beta} \sqrt{\sum_{j=1}^J r_j p_j q_j (x_j - \bar{x})^2} \right]^2 \quad (3)$$

where  $z_\alpha$  is the  $(\alpha)$ th quantile of the standard normal distribution.

The total sample size is computed as  $n = \sum_{j=1}^J n_j = n_1 \sum_{j=1}^J r_j$ .

The sample-size estimate  $n_{1c}$  for the continuity-corrected statistic is found according to Nam (1987), as follows:

$$n_{1c} = \begin{cases} \frac{n_1}{4} \left( 1 + \sqrt{1 - \frac{4c}{An_1}} \right)^2 & \text{for a lower one-sided test} \\ \frac{n_1}{4} \left( 1 + \sqrt{1 + \frac{4c}{An_1}} \right)^2 & \text{for an upper one-sided test} \end{cases} \quad (4)$$

For a two-sided hypothesis,  $n$  is computed by iteratively solving the two-sided power equation given in (1) using the one-sided estimates as starting values. Without continuity correction, (3) is used to obtain the one-sided sample estimates for the starting values; with a continuity correction, (4) is used.

## References

- Agresti, A. 2013. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- Armitage, P. 1955. Tests for linear trends in proportions and frequencies. *Biometrics* 11: 375–386.
- Cochran, W. G. 1954. Some methods for strengthening the common chi-squared tests. *Biometrics* 10: 417–451.
- Graubard, B. I., and E. L. Korn. 1987. Choice of column scores for testing independence in ordered  $2 \times K$  contingency tables. *Biometrics* 43: 471–476.
- Lachin, J. M. 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.
- Nam, J. 1987. A simple approximation for calculating sample sizes for detecting linear trend in proportions. *Biometrics* 43: 701–705.
- Sribney, W. M. 1996. FAQ: Does Stata provide a test for trend? <http://www.stata.com/support/faqs/statistics/test-for-trend>.

## Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **epitab** — Tables for epidemiologists
- [R] **logit** — Logistic regression, reporting coefficients

# Title

**power cox** — Power analysis for the Cox proportional hazards model

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`power cox` computes sample size, power, or effect size for survival analyses that use Cox proportional hazards (PH) models. The results are obtained for the test of the effect of one covariate (binary or continuous) on time to failure adjusted for other predictors in a PH model. Effect size can be expressed as a regression coefficient (or log hazard-ratio) or as a hazard ratio. The command can account for the dependence between the covariate of interest and other model covariates, and it can adjust computations for censoring and for withdrawal of subjects for the study.

## Quick start

Sample size for a test of  $H_0: \beta_1 = 0$  versus  $H_a: \beta_1 \neq 0$  of a binary covariate in a Cox PH model given alternative coefficient `bl` of 0.4, no censoring, and default power of 0.8 and significance level  $\alpha = 0.05$

```
power cox .4
```

Same as above, specified using hazard ratio of 1.492

```
power cox, hratio(1.492)
```

Sample size for a test of a continuous independent variable with `bl` of 1.2 and standard deviation of 0.47

```
power cox 1.2, sd(.47)
```

As above, but for a model with multiple covariates and a squared multiple-correlation coefficient  $R^2 = 0.2$  when the covariate of interest is regressed on all other covariates

```
power cox 1.2, sd(.47) r2(.2)
```

Sample size and estimated number of events for a study with censoring, assuming an 80% event rate

```
power cox .4, eventprob(.8)
```

Plot required sample size versus coefficient values of 0.2, 0.3, 0.4, 0.5, and 0.6

```
power cox (.2(.1).6), graph
```

Sample size assuming 12% dropout of subjects in the study

```
power cox .4, wdprob(.12)
```

Power of a test of a binary independent variable given sample size of 200

```
power cox .4, n(200)
```

Power of a one-sided test of a continuous independent variable with `bl` of 1.2 and standard deviation of 0.47

```
power cox 1.2, sd(.47) n(200) onesided
```

Effect size given power of 0.8 and sample sizes of 120, 130, 140, 150, and 160

```
power cox, power(.8) n(120(10)160)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

```
power cox [b1] [, power(numlist) options]
```

*Compute power*

```
power cox [b1], n(numlist) [options]
```

*Compute effect size (target regression coefficient)*

```
power cox, n(numlist) power(numlist) [options]
```

where *b1* is the hypothesized regression coefficient (effect size) of a covariate of interest in a Cox PH model desired to be detected by a test with a prespecified power. *b1* may be specified either as one number or as a list of values in parentheses (see [U] [11.1.8 numlist](#)).

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample size
* <u>hratio</u> ( <i>numlist</i> )	hazard ratio (exponentiated <i>b1</i> ) associated with a one-unit increase in covariate of interest; specify instead of the regression coefficient <i>b1</i> ; default is <code>hratio(0.5)</code>
* <u>sd</u> ( <i>numlist</i> )	standard deviation of covariate of interest; default is <code>sd(0.5)</code>
* <u>r2</u> ( <i>numlist</i> )	squared coefficient of multiple correlation with other covariates; default is <code>r2(0)</code>
* <u>eventprob</u> ( <i>numlist</i> )	overall probability of an event (failure) of interest; default is <code>eventprob(1)</code> , meaning no censoring
* <u>failprob</u> ( <i>numlist</i> )	synonym for <code>eventprob()</code>
* <u>wdprob</u> ( <i>numlist</i> )	proportion of subjects anticipated to withdraw from the study; default is <code>wdprob(0)</code>
<u>effect</u> ( <i>effect</i> )	specify the type of effect to display; default is <code>effect(coefficent)</code>
<u>direction</u> ( <u>lower</u>   <u>upper</u> )	direction of the effect for effect-size determination; default is <code>direction(lower)</code> , which means that the postulated value of the parameter is smaller than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [ , <u>replace</u> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>
<u>notitle</u>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 **numlist**.

`notitle` does not appear in the dialog box.

<i>effect</i>	Description
<u>coefficient</u>	regression coefficient, <i>b1</i> ; the default
<u>hratio</u>	hazard ratio, $\exp(b1)$
<u>lnhratio</u>	log hazard-ratio; synonym for <code>coefficient</code>

where *tablespec* is

```
column[:label] [column[:label] [...]] [, tableopts]
```

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<b>alpha</b>	significance level	$\alpha$
<b>power</b>	power	$1 - \beta$
<b>beta</b>	type II error probability	$\beta$
<b>N</b>	number of subjects	$N$
<b>delta</b>	effect size	$\delta$
<b>E</b>	total number of events (failures)	$E$
<b>b1</b>	regression coefficient	$\beta_{1a}$
<b>hratio</b>	hazard ratio	$\Delta$
<b>lnhratio</b>	log hazard-ratio	$\ln(\Delta)$
<b>sd</b>	standard deviation	$\sigma$
<b>R2</b>	squared multiple-correlation coefficient	$R^2$
<b>Pr_E</b>	overall probability of an event (failure)	$p_E$
<b>Pr_w</b>	probability of withdrawals	$p_w$
<b>target</b>	target parameter; synonym for <b>b1</b>	
<b>_all</b>	display all supported columns	

Column **beta** is shown in the default table in place of column **power** if option **beta()** is specified.

Column **b1** is shown in the default table in place of column **hratio** when a regression coefficient is specified.

Columns **R2** and **Pr\_w** are shown in the default table only if specified.

## Options

### Main

**alpha()**, **power()**, **beta()**, **n()**, **nfractional**; see [PSS] **power**. The **nfractional** option is allowed only for sample-size determination.

**hratio(numlist)** specifies the hazard ratio (or exponentiated regression coefficient) associated with a one-unit increase in the covariate of interest when other covariates are held constant. This value defines an effect size or the minimal clinically significant effect of a covariate on the response to be detected by a test with a certain power in a Cox PH model.

You can specify an effect size either as the regression coefficient *b1*, which is the command argument, or as the hazard ratio in **hratio()**. The default is **hratio(0.5)**. If you specify **hratio(#)**, the regression coefficient is computed as  $b1 = \ln(\#)$ . If you specify a regression coefficient *b1*, the hazard ratio is computed as  $\exp(b1)$ .

This option is not allowed with the effect-size determination.

**sd(numlist)** specifies the standard deviation of the covariate of interest. The default is **sd(0.5)**.

**r2(numlist)** specifies the squared multiple-correlation coefficient between the covariate of interest and other predictors in a Cox PH model. The default is **r2(0)**, meaning that the covariate of interest is independent of other covariates. This option defines the proportion of variance explained by the regression of the covariate of interest on other covariates used in the Cox model (see [R] **regress**).

`eventprob(numlist)` specifies the overall probability of a subject experiencing an event of interest (or failing, or not being censored) in the study. The default is `eventprob(1)`, meaning that all subjects experience an event (or fail) in the study; that is, no censoring of subjects occurs.

`failprob(numlist)` is a synonym for `eventprob()`.

`wdprob(numlist)` specifies the proportion of subjects anticipated to withdraw from a study. The default is `wdprob(0)`. `wdprob()` is allowed only with sample-size computation.

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. `effect` is one of `coefficient`, `hratio`, or `lnhratio`. By default, the effect size `delta` is the regression coefficient, `effect(coefficient)`.

`direction()`, `onesided`, `parallel`; see [PSS] [power](#). `direction(lower)` is the default.

#### Table

`table`, `table()`, `notable`; see [PSS] [power, table](#).

`saving()`; see [PSS] [power](#).

#### Graph

`graph`, `graph()`; see [PSS] [power, graph](#). Also see the [column](#) table for a list of symbols used by the graphs.

The following option is available with `power cox` but is not shown in the dialog box:

`notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

[Introduction](#)

[Using power cox](#)

[Computing sample size](#)

[Computing sample size in the absence of censoring](#)

[Computing sample size in the presence of censoring](#)

[Link to the sample-size and power computation for the log-rank test](#)

[Computing power](#)

[Computing effect size](#)

[Performing analyses using a Cox PH model](#)

This entry describes the `power cox` command and the methodology for power and sample-size analysis for survival analyses that use Cox PH models. See [PSS] [intro](#) for a general introduction to power and sample-size analysis, and see [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests. See *Survival data* in [PSS] [intro](#) for an introduction to power and sample-size analysis for survival data.

## Introduction

Consider a survival study for which the goal is to investigate the effect of a covariate of interest,  $x_1$ , on time to failure, possibly adjusted for other predictors,  $x_2, x_3, \dots, x_p$ , using the Cox PH model (Cox 1972). The effect is commonly measured as a hazard ratio  $\Delta$  associated with a one-unit increase in  $x_1$  when the other covariates  $x_2, x_3, \dots, x_p$  are held constant. For a binary predictor  $x_1$ , hazard ratio  $\Delta$  corresponds to the two categories of  $x_1$  when the other covariates are held constant.

In a Cox PH model, the hazard function is assumed to be

$$h(t) = h_0(t)\exp(\beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p)$$

where no distributional assumption is made about the baseline hazard  $h_0(t)$ . Under this assumption, the regression coefficient  $\beta_1$  is the log hazard-ratio  $\ln(\Delta)$  associated with a one-unit increase in  $x_1$  when the other predictors are held constant, and the exponentiated regression coefficient  $\exp(\beta_1)$  is the hazard ratio  $\Delta$ . Therefore, the effect of  $x_1$  on time to failure can be investigated by performing an appropriate test based on the partial likelihood (Hosmer, Lemeshow, and May 2008; Klein and Moeschberger 2003) for the regression coefficient  $\beta_1$  from a Cox model. Negative values of  $\beta_1$  correspond to the reduction in hazard for a one-unit increase in  $x_1$ , and, conversely, positive values correspond to the increase in hazard for a one-unit increase in  $x_1$ .

`power cox` provides power and sample-size analysis for a test of the regression coefficient  $\beta_1$  in a Cox model with the null hypothesis  $H_0: (\beta_1, \beta_2, \dots, \beta_p) = (0, \beta_2, \dots, \beta_p)$  against the alternative  $H_a: (\beta_1, \beta_2, \dots, \beta_p) = (\beta_{1a}, \beta_2, \dots, \beta_p)$ . The postulated or null coefficient value is zero and a hypothesized or alternative coefficient value, specified as `b1` with `power cox`, is  $\beta_{1a}$ . (Similarly,  $\Delta_a$  is a hypothesized or alternative value of the hazard ratio specified in the `hratio()` option.) The methods used are derived for the score test of  $H_0$  versus  $H_a$ . In practice, however, the obtained results may be used in the context of the Wald test as well because the two tests usually lead to the same conclusions about the significance of the regression coefficient. Refer to [The conditional versus unconditional approaches](#) in [PSS] `power exponential` for more details about the results based on conditional and unconditional tests. From now on, we will refer to  $H_a$  as  $H_a: \beta_1 = \beta_{1a}$  for simplicity.

`power cox` implements the method of Hsieh and Lavori (2000) for the sample-size and power computation, which reduces to the method of Schoenfeld (1983) for a binary covariate. The sample size is related to the power of a test through the number of events observed in the study; that is, for a fixed number of events, the power of a test is independent of the sample size. As a result, the sample size is estimated as the number of events divided by the overall probability of a subject failing in the study.

You can use `power cox` to

- compute required number of events and sample size when you know power and effect size expressed as a hazard ratio or regression coefficient (log hazard-ratio);
- compute power when you know sample size (number of events) and effect size expressed as a hazard ratio or regression coefficient (log hazard-ratio); or
- compute effect size and display it as a regression coefficient, a log hazard-ratio, or a hazard ratio when you know sample size (number of events) and power.

You can also account for the dependence between the covariate of interest and other model covariates in your analysis, adjust results for censoring, and adjust results for withdrawal of subjects from the study.

## Using `power cox`

`power cox` computes sample size, power, or effect size for a test of one regression coefficient in a Cox PH model, holding coefficients of the other covariates constant. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test.



To compute sample size, you specify an effect size and, optionally, power of the test in the `power()` option. The default power is set to 0.8. By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS] **unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

To compute power, you must specify the sample size in the `n()` option and an effect size.

An effect size may be specified either as a regression coefficient supplied as the command argument `b1` or as a hazard ratio supplied in the `hratio()` option. If neither is specified, a hazard ratio of 0.5 is assumed.

To compute effect size, which may be expressed either as a regression coefficient (log hazard-ratio) or hazard ratio, you must specify the sample size in the `n()` option; the power in the `power()` option; and, optionally, the direction of the effect. The direction is lower by default, `direction(lower)`, which means, for example, that the target regression coefficient is assumed to be negative. This is equivalent to the hazard ratio being less than one. You can change the direction to upper, which means that the target regression coefficient is assumed to be positive, by specifying the `direction(upper)` option. This is equivalent to the hazard ratio being greater than one.

As we mentioned above, the effect size for `power cox` may be expressed as a regression coefficient or, equivalently, a log hazard-ratio, or as a hazard ratio. By default, the effect size, which is labeled as `delta` in the output, corresponds to the regression coefficient. You can change this by specifying the `effect()` option: `effect(coefficient)` (the default) reports the regression coefficient, `effect(hratio)` reports the hazard ratio, and `effect(lnhratio)` reports the log hazard-ratio.

The standard deviation of the covariate of interest is set to 0.5 by default and may be changed by specifying the `sd()` option. In the presence of additional covariates in a Cox model, you can use the `r2()` option to specify the correlation between the covariate of interest and other covariates in the model.

All computations assume no censoring. In the presence of censoring, you can use the `eventprob()` option to specify an overall probability of an event or failure. When computing sample size, you can also adjust for withdrawal of subjects from the study by specifying the anticipated proportion of withdrawals in the `wdprob()` option.

In the following sections, we describe the use of `power cox` accompanied by examples for computing sample size, power, and effect size.

## Computing sample size

To compute sample size and number of events, you must specify an effect size (a regression coefficient or a hazard ratio) and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified. A hazard ratio of 0.5 is assumed if an effect size is not specified.

## Computing sample size in the absence of censoring

First, consider a **type I study** in which all subjects fail by the end of the study (no censoring). Then the required sample size is the same as the number of events required to be observed in a study.

### ▷ Example 1: Sample size for a model with a binary covariate of interest

Consider a survival study for which the goal is to investigate the effect of a treatment on survival times of subjects. The covariate of interest is binary with levels defining whether a subject receives the treatment (the experimental group) or a placebo (the control or placebo group). Prior to conducting the study, investigators need an estimate of the sample size that ensures that a ratio of hazards of the experimental group to the control group of 0.5 ( $\beta_{1a} = \ln(0.5) = -0.6931$ ) can be detected with a power of 80% with a two-sided, 5%-level test. Under 1:1 randomization, a subject has a 50% chance of receiving the treatment. The corresponding binary covariate follows a Bernoulli distribution with the probability of a subject receiving a treatment,  $p$ , equal to 0.5. As such, the standard deviation of the covariate is  $\{p(1-p)\}^{1/2} = 0.5$ . Because these study parameters correspond to default values of `power cox`, to obtain the sample size for the above study, we simply type

```
. power cox
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.6931 (coefficient)
    hratio =    0.5000
    sd =      0.5000

Censoring:
    Pr_E =    1.0000

Estimated number of events and sample size:
    E =      66
    N =      66
```

Recall that, by default, a hazard ratio of 0.5, corresponding to the regression coefficient of  $\ln(0.5) = -0.6931$ , is assumed. From the output, we see that 66 events (failures) are required to be observed in the study to ensure a power of 80% to detect an alternative  $H_a: \beta_1 = -0.6931$  using a two-sided test with a 0.05 significance level. Because we have no censoring ( $\text{Pr\_E} = 1.0000$ ), a total of 66 subjects is needed in the study to observe 66 events.

◀

### ▷ Example 2: Alternative ways of specifying effect

In [example 1](#), the effect size `delta` corresponds to the regression coefficient, the default. We can use the `effect(hratio)` option to redefine `delta` as a hazard ratio.

```
. power cox, effect(hratio)
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.5000 (hazard ratio)
    hratio =    0.5000
    sd =      0.5000

Censoring:
    Pr_E =    1.0000

Estimated number of events and sample size:
    E =      66
    N =      66
```

We can also obtain the same results as in [example 1](#) by directly specifying the value of the coefficient.

```
. power cox -0.6931
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.6931 (coefficient)
    b1 =     -0.6931
    sd =      0.5000
Censoring:
    Pr_E =    1.0000
Estimated number of events and sample size:
    E =       66
    N =       66
```

The specified regression coefficient,  $b_1$ , is reported in the output instead of the hazard ratio `hratio`. ◀

Suppose now that the covariate of interest,  $x_1$ , is continuous. [Hsieh and Lavori \(2000\)](#) extend the formula of [Schoenfeld \(1983\)](#) for the number of events to the case when a covariate is continuous. They also relax the assumption of [Schoenfeld \(1983\)](#) about the independence of  $x_1$  of other covariates and provide an adjustment to the estimate of the number of events for possible correlation.

### ▶ Example 3: Sample size for a model with a continuous covariate of interest

Consider an example from [Hsieh and Lavori \(2000\)](#) of a study of multiple-myeloma patients treated with alkylating agents ([Krall, Uthoff, and Harley 1975](#)). Although in the original study of multiple-myeloma patients, 17 of a total of 65 patients are censored; here we assume that all patients die by the end of the study (a type I study, no censoring). Suppose that the covariate of interest,  $x_1$ , is the log of the amount of blood urea nitrogen (BUN) measured in a patient. The sample size for a one-sided, 5%-level test to detect a coefficient (log hazard-ratio) of 1 for a unit increase in  $x_1$  with a power of 80% is required. The standard deviation of  $x_1$  is 0.3126. To obtain an estimate of the sample size, we supply `b1`, 1, as an argument, the `sd(0.3126)` option, and the `onesided` option to request a one-sided test.

```
. power cox 1, sd(0.3126) onesided
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    1.0000 (coefficient)
    b1 =      1.0000
    sd =      0.3126
Censoring:
    Pr_E =    1.0000
Estimated number of events and sample size:
    E =       64
    N =       64
```

The estimate of the required number of events and the sample size is 64.

◀

Based on the derivation in [Schoenfeld \(1983\)](#) and [Hsieh and Lavori \(2000\)](#), sample-size estimates in the above examples may be used if other covariates are also present in the model as long as these covariates are independent of the covariate of interest. The independence assumption holds for randomized studies, but it is not true for nonrandomized studies often encountered in practice. Also, in many studies, the main covariate of interest will often be correlated with other covariates. For example, age and gender will often be confounded with the covariate of interest, such as smoking. Below we investigate the effect of the confounding factor on the estimate of the required number of events.

#### ▶ Example 4: Sample size when covariates are not independent

Continuing with [example 3](#), suppose that we want to adjust the effect of the covariate BUN for eight other covariates in the model. [Hsieh and Lavori \(2000\)](#) report the coefficient of determination of  $R^2 = 0.1837$  from regression of the log of BUN,  $x_1$ , on the eight other covariates. We specify this value in the `r2()` option.

```
. power cox 1, sd(0.3126) onesided r2(0.1837)
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    1.0000 (coefficient)
        b1 =    1.0000
        sd =    0.3126
        R2 =    0.1837
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample size:
      E =    78
      N =    78
```

The number of events required to be observed in the study increases from 64 to 78 because of the correlation between BUN and the other covariates. The new estimate is equal to the original estimate multiplied by the variance inflation factor  $VIF = 1/(1 - R^2)$ . Likewise, the number of subjects increases from 64 to 78.

◀

### Computing sample size in the presence of censoring

In the previous section, we assumed that all subjects experience an event by the end of the study. In practice, the study often terminates after a fixed time,  $T$ . As a result, some subjects may not experience an event by the end of the study (a [type II study](#)). These subjects are censored. To obtain an estimate of the sample size in the presence of censoring, an estimate of the overall probability of a subject not being censored is required. The investigator may already have such an estimate from previous studies, or this probability may be computed as suggested in the literature ([Schoenfeld \[1983\]](#), [Lachin and Foulkes \[1986\]](#), [Barthel et al. \[2006\]](#), and [Barthel, Royston, and Babiker \[2005\]](#), also see [\[PSS\] power logrank](#) and [\[PSS\] power exponential](#)).

### ▷ Example 5: Sample size in the presence of censoring

Consider the study from [example 3](#). In reality, as mentioned earlier, 17 of a total of 65 patients survived until the end of the study. The overall death rate is estimated as  $1 - 17/65 = 0.738$ . We specify the overall probability of an event in the `eventprob()` option.

```
. power cox 1, sd(0.3126) onesided eventprob(0.738)
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    1.0000 (coefficient)
      b1 =    1.0000
      sd =    0.3126
Censoring:
    Pr_E =    0.7380
Estimated number of events and sample size:
    E =      64
    N =      86
```

In the presence of censoring, the number of subjects required in the study increases from 64 to 86. The number of events remains the same (64) because the only change in the study is the presence of censoring, and censoring is assumed to be independent of failure (event) times.

◀

### ▷ Example 6: Sample size in the presence of censoring adjusting for other covariates

Continuing with [example 5](#), if we also adjust for the correlation between the log of BUN and other covariates, we obtain the estimate of the sample size to be 106.

```
. power cox, hratio(2.7182) sd(0.3126) onesided eventprob(0.738) r2(0.1837)
Estimated sample size for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    1.0000 (coefficient)
  hratio =    2.7182
      sd =    0.3126
      R2 =    0.1837
Censoring:
    Pr_E =    0.7380
Estimated number of events and sample size:
    E =      78
    N =     106
```

In the above example, for no other reason than variety, rather than supplying the coefficient of 1, we used the `hratio()` option to specify the size of the effect expressed as the hazard ratio  $\exp(1) = 2.7182$ .

◀

## □ Technical note

Supplying the coefficient (log hazard-ratio) of 1 or  $-1$  [or, respectively, the hazard ratio of  $\exp(1) = 2.7182$  or  $\exp(-1) = 1/2.7182 = 0.36788$ ] is irrelevant for sample-size and power determination because it results in the same estimates of sample size and power. However, the sign of the coefficient (or the value of the hazard ratio being larger or smaller than one) is important at the analysis stage because it determines the direction of the effect associated with a one-unit increase of a covariate value. □

Often, in practice, subjects may withdraw from a study before it terminates. As a result, the information about the subjects' response is lost. The proportion of subjects anticipated to withdraw from a study may be specified by using `wdprob()`. Refer to *Survival data* in [PSS] [intro](#) and *Withdrawal of subjects from the study* in [PSS] [power logrank](#) for a more detailed description and an example.

### Link to the sample-size and power computation for the log-rank test

The score test of the regression coefficient of a binary covariate in a Cox model with one binary predictor is the same (in the absence of tied observations) as the log-rank test comparing survivor functions of two groups defined by this covariate. Powers of the two tests are the same and so are the formulas for the number of events (Schoenfeld 1983; Schoenfeld 1981). As such, the required number of events computed by `power cox` for the Cox model and the Schoenfeld method of `power logrank` for the log-rank test are the same.

### ▷ Example 7: Using power logrank for a binary covariate

Using `power logrank` (see [PSS] [power logrank](#)) for the study described in [example 1](#) yields the same estimates of 66 for both the required number of events and the required sample size.

```
. power logrank, schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.6931 (log hazard-ratio)
      hratio =    0.5000
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample sizes:
      E =        66
      N =        66
      N per group =    33
```

Schoenfeld (1983) demonstrates that the same formula can also be used to compute the required number of events when other covariates are present in a Cox model, provided that the binary covariate of interest is independent of these covariates and that the covariates are not extremely unbalanced. Although the formulas for the number of events are the same whether covariates are present or not, it is important to adjust for covariates when analyzing the data to avoid loss of power; see Schoenfeld (1983) for details. ◀

Vaeth and Skovlund (2004) demonstrate that for a continuous covariate of interest in a Cox model, the sample-size formula for the log-rank test with the value of the hazard ratio equal to  $\exp(2\beta_{1a}\sigma)$  and with equal-group allocation may be used to obtain the required sample size. Indeed, by substituting the above expression for the hazard ratio into the sample-size formula for the log-rank test, one obtains the sample-size formula derived in Hsieh and Lavori (2000) for a Cox model.

### ► Example 8: Using power logrank for a continuous covariate

For example, we obtain the same estimate of the total number of events as computed in [example 3](#) by using `power logrank` with the `schoenfeld` option and with the value of the hazard ratio equal to  $\exp(2\beta_{1a}\sigma) = \exp(2 \times 1 \times 0.3126) = 1.8686$ .

```
. power logrank, hratio(1.8686) onesided schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) > 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.6252 (log hazard-ratio)
      hratio =    1.8686
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample sizes:
      E =        64
      N =        64
      N per group =    32
```

◀

## Computing power

To compute power, you must specify the sample size in the `n()` option and an effect size (a regression coefficient or a hazard ratio). A hazard ratio of 0.5 is assumed if an effect size is not specified.

### ► Example 9: Power determination

In [example 6](#), the required number of patients was estimated to be 106 to ensure a power of 80% for a 0.05 one-sided test to detect a value of 1 in the regression coefficient. Suppose that we can recruit only 65 subjects. How does this reduction in sample size affect the power of the test to detect the alternative  $H_a: \beta_1 = 1$ ?

```

. power cox 1, sd(0.3126) onesided r2(0.1837) eventprob(0.738) n(65)
Estimated power for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0
Study parameters:
      alpha =    0.0500
        N =      65
      delta =    1.0000 (coefficient)
        b1 =    1.0000
        sd =    0.3126
        R2 =    0.1837

Number of events and censoring:
      E =      48
    Pr_E =    0.7380

Estimated power:
      power =    0.6222

```

When the sample size decreases from 106 to 65, the corresponding number of events decreases from 78 to 48, and power decreases from 80% to 62%.

◀

### ▷ Example 10: Multiple values of study parameters

Suppose we want to investigate the effect of a correlation between a covariate of interest and other model covariates on power. Continuing with [example 9](#), we can specify a list (see [\[U\] 11.1.8 numlist](#)) of correlations in the `r2()` option.

```

. power cox 1, sd(0.3126) onesided r2(0.1(0.1)0.5) eventprob(0.738) n(65)
Estimated power for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0

```

alpha	power	N	E	delta	b1	sd	R2	Pr_E
.05	.6588	65	48	1	1	.3126	.1	.738
.05	.6147	65	48	1	1	.3126	.2	.738
.05	.5662	65	48	1	1	.3126	.3	.738
.05	.5128	65	48	1	1	.3126	.4	.738
.05	.4547	65	48	1	1	.3126	.5	.738

As the correlation increases, the power decreases.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

◀



## Computing effect size

Effect size  $\delta$  for a test of a coefficient in a Cox PH model is defined as a coefficient (or, equivalently, a log hazard-ratio) or a hazard ratio, corresponding to a one-unit change in the tested covariate holding other model covariates constant.

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, both power and sample size must be specified in options `power()` and `n()`, respectively. Additionally, you may also choose the direction of the effect by specifying the `direction()` option. `direction(lower)`, the default, assumes  $\beta_{1a} < 0$  (or  $\Delta_a < 1$ ), corresponding to the reduction in hazard for a unit change in a covariate. `direction(upper)` assumes  $\beta_{1a} > 0$  (or  $\Delta_a > 1$ ).

### ► Example 11: Effect-size determination

Continuing with [example 9](#), if a power of 62% is unacceptable to investigators, they may want to determine the smallest value of the regression coefficient that can be detected with a preserved power of 80%. To obtain this estimate, we specify both the `n()` and the `power()` options.

```
. power cox, sd(0.3126) onesided r2(0.1837) eventprob(0.738) n(65) power(0.8)
> direction(upper)
Estimated target coefficient for Cox PH regression
Wald test
Ho: beta1 = 0 versus Ha: beta1 > 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      65
        sd =    0.3126
         R2 =    0.1837

Number of events and censoring:
      E =      48
     Pr_E =    0.7380

Estimated effect size and target coefficient:
      delta =    1.2711 (coefficient)
         b1 =    1.2711
```

With only 65 subjects, the smallest change in a coefficient (log hazards) for a one-unit increase in the log of BUN, which can be detected with a preserved 80% power, is roughly 1.27, corresponding to a 27% increase in the coefficient of 1 desired to be detected originally in [example 6](#).

◀

## Performing analyses using a Cox PH model

After the data are collected, one can use `stcox` and `test` to fit the Cox PH model and perform a Wald test, as we demonstrate below.

### ► Example 12: Performing a Wald test

We demonstrate how to perform a Wald test for the regression coefficient of the log of BUN from a Cox model using the data from [Krall, Uthoff, and Harley \(1975\)](#) described in [example 3](#). `myeloma.dta` consists of 11 variables, described below.

```
. use http://www.stata-press.com/data/r14/myeloma
(Multiple myeloma patients)
. describe
Contains data from http://www.stata-press.com/data/r14/myeloma.dta
  obs:           65                Multiple myeloma patients
  vars:          11                11 Feb 2014 19:26
  size:          1,690
```

variable name	storage type	display format	value label	variable label
time	float	%9.0g		Survival time from diagnosis to nearest month + 1
died	byte	%9.0g		0 - Alive, 1 - Dead
lnbun	float	%9.0g		log BUN at diagnosis
hemo	float	%9.0g		Hemoglobin at diagnosis
platelet	byte	%9.0g	normal	Platelets at diagnosis
age	byte	%9.0g		Age (complete years)
lnwbc	float	%9.0g		Log WBC at diagnosis
fracture	byte	%9.0g	present	Fractures at diagnosis
lnbm	float	%9.0g		log % of plasma cells in bone marrow
protein	byte	%9.0g		Proteinuria at diagnosis
scalcium	byte	%9.0g		Serum calcium (mgm%)

Sorted by:

Before using `stcox` to fit a Cox model, we need to set up the data by using `stset` (see [ST] `stset`). The analysis-time variable is `time`, and the failure variable is `died`.

```
. stset time, failure(died)
      failure event:  died != 0 & died < .
obs. time interval:  (0, time]
exit on or before:  failure
```

65	total observations		
0	exclusions		
65	observations remaining, representing		
48	failures in single-record/single-failure data		
1560.5	total analysis time at risk and under observation		
	at risk from t =		0
	earliest observed entry t =		0
	last observed exit t =		92

We include all nine covariates in the model and perform a fit by using `stcox`. Then we perform a Wald test of  $H_0: \beta_1 = 1$  for the coefficient of `lnbun` using `test`.

```

. stcox lnbun hemo platelet age lnwbc fracture lnbm protein scalcium, nohr
      failure _d: died
      analysis time _t: time
Iteration 0:   log likelihood = -154.85799
Iteration 1:   log likelihood = -146.68114
Iteration 2:   log likelihood = -146.29446
Iteration 3:   log likelihood = -146.29404
Refining estimates:
Iteration 0:   log likelihood = -146.29404
Cox regression -- Breslow method for ties
No. of subjects =           65           Number of obs   =           65
No. of failures =           48
Time at risk    =          1560.5
Log likelihood  = -146.29404           LR chi2(9)       =           17.13
                                           Prob > chi2     =           0.0468

```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnbun	1.798354	.6483293	2.77	0.006	.5276519 3.069056
hemo	-.1263119	.0718333	-1.76	0.079	-.2671026 .0144789
platelet	-.2505915	.5074656	-0.49	0.621	-1.245206 .7440228
age	-.0127949	.019475	-0.66	0.511	-.0509653 .0253755
lnwbc	.3537259	.7131935	0.50	0.620	-1.044108 1.75156
fracture	.3378767	.4072774	0.83	0.407	-.4603722 1.136126
lnbm	.3589346	.4860298	0.74	0.460	-.5936663 1.311535
protein	.0130672	.0261696	0.50	0.618	-.0382243 .0643587
scalcium	.1259479	.1034015	1.22	0.223	-.0767153 .3286112

```

. test lnbun = 1
( 1) lnbun = 1
      chi2( 1) =      1.52
      Prob > chi2 =      0.2182

```

By default, `stcox` reports estimates of hazard ratios and the two-sided tests of the equality of a coefficient to zero. We use the `nohr` option to request estimates of coefficients. From the output table, a one-sided test of  $H_0: \beta_1 = 0$  versus  $H_a: \beta_1 > 0$  is rejected at a 0.05 level (one-sided  $p$ -value is  $0.006/2 = 0.003 < 0.05$ ). The estimate of the log-hazard difference associated with a one-unit increase of `lnbun` is  $\hat{\beta}_1 = 1.8$ . From the `test` output, we cannot reject the hypothesis of  $H_0: \beta_1 = 1$ .

For these data, the observed effect size (coefficient) of 1.8 is large enough for the sample size of 65 to be sufficient to reject the null hypothesis of no effect of the BUN on the survival of subjects ( $H_0: \beta_1 = 0$ ). However, if the goal of the study were to ensure that the test detects the effect size corresponding to the coefficient of at least 1 with 80% power, a sample of approximately 106 subjects would have been required.

## Stored results

`power cox` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(E)</code>	total number of events (failures)
<code>r(hratio)</code>	hazard ratio under the alternative hypothesis
<code>r(b1)</code>	regression coefficient under the alternative hypothesis
<code>r(sd)</code>	standard deviation
<code>r(R2)</code>	squared multiple correlation, if specified
<code>r(Pr_E)</code>	probability of an event (failure), if specified
<code>r(Pr_w)</code>	proportion of withdrawals, if specified
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>cox</code>
<code>r(effect)</code>	<code>coefficient</code> , <code>hratio</code> , or <code>lnhratio</code>
<code>r(direction)</code>	<code>lower</code> or <code>upper</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
---------------------------	------------------

## Methods and formulas

Let  $\beta_1$  denote the regression coefficient of the covariate of interest,  $x_1$ , from a Cox PH model, possibly in the presence of other covariates,  $x_2, \dots, x_p$ ; and let  $\Delta$  denote the hazard ratio associated with a one-unit increase of  $x_1$  when other covariates are held constant. Under the PH model,  $\beta_1 = \ln(\Delta)$ , where  $\ln(\Delta)$  is the change in log hazards associated with a one-unit increase in  $x_1$  when other covariates are held constant.

Define  $E$  and  $n$  to be the total number of events (failures) and the total number of subjects required in the study;  $\sigma$  to be the standard deviation of  $x_1$ ;  $p_E$  to be the overall probability of an event (failure);  $R^2$  to be the proportion of variance explained by the regression of  $x_1$  on  $x_2, \dots, x_p$  (or squared multiple-correlation coefficient);  $p_w$  to be the proportion of subjects withdrawn from a study (lost to follow-up);  $\alpha$  to be the significance level;  $\beta$  to be the probability of a type II error; and  $z_{(1-\alpha/k)}$  and  $z_{(1-\beta)}$  to be the  $(1 - \alpha/k)$ th and the  $(1 - \beta)$ th quantiles of the standard normal distribution, respectively, with  $k = 1$  for the one-sided test and  $k = 2$  for the two-sided test.

The total number of events required to be observed in a study to ensure a power of  $1 - \beta$  of a test to detect the regression coefficient,  $\beta_1$ , with a significance level  $\alpha$ , according to Hsieh and Lavori (2000), is

$$E = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2}{\sigma^2 \beta_1^2 (1 - R^2)}$$

For the case of randomized study and a binary covariate  $x_1$ , this formula was derived in Schoenfeld (1983). The formula is an approximation and relies on a set of assumptions such as distinct failure times, all subjects completing the course of the study (no withdrawal), and a local alternative under which  $\ln(\Delta)$  is assumed to be of order  $O(n^{-1/2})$ . The formula is derived for the score test but may be applied to other tests (Wald, for example) that are based on the partial likelihood of a Cox model because all of these tests are asymptotically equivalent (Schoenfeld 1983; Hosmer, Lemeshow, and May 2008; Klein and Moeschberger 2003).

The total sample size required to observe the total number of events,  $E$ , is given by

$$n = \frac{E}{p_E}$$

If the `nfractional` option is not specified, the computed sample size is rounded up.

To account for a proportion of subjects,  $p_w$ , withdrawn from a study, a conservative adjustment to the total sample size suggested in the literature (Freedman 1982; Machin and Campbell 2005) is applied as follows:

$$n_w = \frac{n}{1 - p_w}$$

Withdrawal is assumed to be independent of administrative censoring and failure (event) times.

Power is estimated using the formula

$$1 - \beta = \Phi \left[ |\beta_1| \sigma \{np_E(1 - R^2)\}^{1/2} - z_{1-\alpha/k} \right]$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function.

The estimate of the regression coefficient for a fixed power,  $1 - \beta$ , and a sample size,  $n$ , is computed as

$$\beta_1^2 = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2}{\sigma^2 np_E(1 - R^2)}$$

Either of the two values  $|\beta_1|$  and  $-|\beta_1|$  satisfy the above equation. By default or if `direction(lower)` is specified, `power cox` reports the negative of the two values, which corresponds to the reduction in a hazard of a failure for a one-unit increase in  $x_1$ . If `direction(upper)` is specified, `power cox` reports the positive of the two values, which corresponds to the increase in a hazard of a failure for a one-unit increase in  $x_1$ .

Similarly, if the `effect(hratio)` option is used, the corresponding value of the hazard ratio less than 1 for `direction(lower)` and greater than 1 for `direction(upper)` is reported to reflect, respectively, the reduction or increase in hazard for a one-unit increase in  $x_1$ .

## References

- Barthel, F. M.-S., A. G. Babiker, P. Royston, and M. K. B. Parmar. 2006. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Statistics in Medicine* 25: 2521–2542.
- Barthel, F. M.-S., P. Royston, and A. G. Babiker. 2005. A menu-driven facility for complex sample size calculation in randomized controlled trials with a survival or a binary outcome: Update. *Stata Journal* 5: 123–129.

- Cox, D. R. 1972. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34: 187–220.
- Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.
- Hosmer, D. W., Jr., S. A. Lemeshow, and S. May. 2008. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. 2nd ed. New York: Wiley.
- Hsieh, F. Y., and P. W. Lavori. 2000. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials* 21: 552–560.
- Klein, J. P., and M. L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer.
- Krall, J. M., V. A. Uthoff, and J. B. Harley. 1975. A step-up procedure for selecting variables associated with survival. *Biometrics* 31: 49–57.
- Lachin, J. M., and M. A. Foulkes. 1986. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 42: 507–519.
- Machin, D., and M. J. Campbell. 2005. *Design of Studies for Medical Research*. Chichester, UK: Wiley.
- Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68: 316–319.
- . 1983. Sample-size formula for the proportional-hazards regression model. *Biometrics* 39: 499–503.
- Væth, M., and E. Skovlund. 2004. A simple approach to power and sample size calculations in logistic regression and Cox regression models. *Statistics in Medicine* 23: 1781–1792.

Also see [\[PSS\] intro](#) for more references.

## Also see

- [\[PSS\] power](#) — Power and sample-size analysis for hypothesis tests
- [\[PSS\] power exponential](#) — Power analysis for the exponential test
- [\[PSS\] power logrank](#) — Power analysis for the log-rank test
- [\[PSS\] power, graph](#) — Graph results from the power command
- [\[PSS\] power, table](#) — Produce table of results from the power command
- [\[PSS\] Glossary](#)
- [\[R\] test](#) — Test linear hypotheses after estimation
- [\[ST\] stcox](#) — Cox proportional hazards model
- [\[ST\] sts test](#) — Test equality of survivor functions

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power exponential` computes sample size or power for survival analysis comparing two exponential survivor functions by using parametric tests for the difference between hazards or, optionally, for the difference between log hazards. It accommodates unequal allocation between the two groups, flexible accrual of subjects into the study, and group-specific losses to follow-up. The accrual distribution may be chosen to be uniform or truncated exponential over a fixed accrual period. Losses to follow-up are assumed to be exponentially distributed. Also the computations may be performed using the conditional or the unconditional approach.

## Quick start

Sample size for a test of exponential hazard rates  $H_0: \lambda_2 = \lambda_1$  versus  $H_a: \lambda_2 \neq \lambda_1$  given control-group hazard rate  $h_1 = 0.4$ , experimental-group hazard rate  $h_2 = 0.2$ , equal group sizes, and no censoring using default power of 0.8 and significance level  $\alpha = 0.05$

```
power exponential .4 .2
```

Same as above, specified using a hazard ratio of 0.5 instead of the experimental-group hazard rate

```
power exponential .4, hratio(.5)
```

As above, but specify hazard ratios of 0.4, 0.45, 0.5, and 0.55

```
power exponential .4, hratio(.4(.05).55)
```

As above, but display results in a graph

```
power exponential .4, hratio(.4(.05).55) graph
```

Total and per group sample sizes given twice as many observations in the experimental group as in the control group

```
power exponential .4 .2, nratio(2)
```

Sample size with survival probabilities  $s_1 = 0.65$  and  $s_2 = 0.8$  and reference survival time 2

```
power exponential .65 .8, time(2)
```

As above, specified using survival probability  $s_1$  and a hazard ratio

```
power exponential .65, time(2) hratio(.52)
```

As above, for a one-sided test with power of 0.95

```
power exponential .65, time(2) hratio(.52) onesided power(.95)
```

Sample size for a test of a log hazard-ratio given  $h_1 = 0.4$  and hazard ratio of 0.5

```
power exponential .4, hratio(.5) loghazard
```

As above, specifying corresponding survival probabilities  $s_1$  and  $s_2$  at reference time 2

```
power exponential .45 .67, time(2) loghazard
```

Sample size for a design with a 10-year follow-up period and a 1-year accrual period

```
power exponential .4 .2, fperiod(10) aperiod(1)
```

Power for a test of  $H_0: \lambda_2 = \lambda_1$ , with  $h_1 = 0.4$ ,  $h_2 = 0.2$ , a sample size of 80, and default  $\alpha = 0.05$

```
power exponential .4 .2, n(80)
```

Power for a test of the log hazard-ratio with  $\alpha = 0.01$

```
power exponential .4, hratio(.5) loghazard n(200) alpha(.01)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

*Specify hazard rates*

```
power exponential [ $h_1$  [ $h_2$ ]] [, power(numlist) options]
```

*Specify survival probabilities*

```
power exponential  $s_1$  [ $s_2$ ], time(#) [ power(numlist) options]
```

*Compute power*

*Specify hazard rates*

```
power exponential [ $h_1$  [ $h_2$ ]], n(numlist) [options]
```

*Specify survival probabilities*

```
power exponential  $s_1$  [ $s_2$ ], time(#) n(numlist) [options]
```

where

$h_1$  is the hazard rate in the control group;

$h_2$  is the hazard rate in the experimental group;

$s_1$  is the survival probability in the control group at reference (base) time  $t$ ; and

$s_2$  is the survival probability in the experimental group at reference (base) time  $t$ .

$h_1$ ,  $h_2$  and  $s_1$ ,  $s_2$  may each be specified either as one number or as a list of values in parentheses; see [U] [11.1.8 numlist](#).



<i>options</i>	Description
Main	
* <u>t</u> ime( <i>numlist</i> )	reference time $t$ for survival probabilities $s_1$ and $s_2$
* <u>a</u> lpha( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>p</u> ower( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>b</u> eta( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power
* <u>n</u> 1( <i>numlist</i> )	sample size of the control group
* <u>n</u> 2( <i>numlist</i> )	sample size of the experimental group
* <u>n</u> ratio( <i>numlist</i> )	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<u>n</u> fractional	allow fractional sample sizes
* <u>h</u> ratio( <i>numlist</i> )	hazard ratio (effect size) of the experimental to the control group; default is <code>hratio(0.5)</code> ; may not be combined with <code>lnhratio()</code> or <code>hdifference()</code>
* <u>l</u> nhratio( <i>numlist</i> )	log hazard-ratio (effect size) of the experimental to the control group; may not be combined with <code>hratio()</code> or <code>hdifference()</code>
* <u>h</u> difference( <i>numlist</i> )	difference between the experimental-group and control-group hazard rates (effect size); may not be combined with <code>hratio()</code> or <code>lnhratio()</code>
* <u>d</u> ifference( <i>numlist</i> )	synonym for <code>hdifference()</code>
<u>l</u> oghazard	power or sample-size computation for the test of the difference between log hazards; default is the test of the difference between hazards
<u>u</u> nconditional	power or sample-size computation using the unconditional approach
<u>e</u> ffect( <i>effect</i> )	specify the type of effect to display; default is method specific
<u>o</u> nesided	one-sided test; default is two sided
<u>p</u> arallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)

Accrual/Follow-up

- \* studytime(*numlist*) duration of the study; if not specified, the study is assumed to continue until all subjects experience an event (fail)
- \* fperiod(*numlist*) length of the follow-up period; if not specified, the study is assumed to continue until all subjects experience an event (fail)
- \* aperiod(*numlist*) length of the accrual period; default is `aperiod(0)`, meaning no accrual
- \* aprob(*numlist*) proportion of subjects accrued by time  $t_a$  under truncated exponential accrual; default is `aprob(0.5)`
- \* aptime(*numlist*) proportion of the accrual period,  $t_a/\text{aperiod}()$ , by which proportion of subjects in `aprob()` is accrued; default is `aptime(0.5)`
- \* atime(*numlist*) reference accrual time  $t_a$  by which the proportion of subjects in `aprob()` is accrued; default value is  $0.5 \times \text{aperiod}()$
- \* ashape(*numlist*) shape of the truncated exponential accrual distribution; default is `ashape(0)`, meaning uniform accrual
- \* lossprob(*numlist*) proportion of subjects lost to follow-up by time `losstime()` in the control and the experimental groups; default is `lossprob(0)`, meaning no losses to follow-up
- \* lossprob1(*numlist*) proportion of subjects lost to follow-up by time `losstime()` in the control group; default is `lossprob1(0)`, meaning no losses to follow-up in the control group
- \* lossprob2(*numlist*) proportion of subjects lost to follow-up by time `losstime()` in the experimental group; default is `lossprob2(0)`, meaning no losses to follow-up in the experimental group
- \* losstime(*numlist*) reference time  $t_L$  by which the proportion of subjects specified in `lossprob()`, `lossprob1()`, or `lossprob2()` is lost to follow-up; default is `losstime(1)`
- \* losshaz(*numlist*) loss hazard rates in the control and the experimental groups; default is `losshaz(0)`, meaning no losses to follow-up
- \* losshaz1(*numlist*) loss hazard rates in the control group; default is `losshaz1(0)`, meaning no losses to follow-up in the control group
- \* losshaz2(*numlist*) loss hazard rates in the experimental group; default is `losshaz2(0)`, meaning no losses to follow-up in the experimental group

Table	
<code>[no]table [(tablespec)]</code>	suppress table or display results as a table; see [PSS] <b>power, table</b>
<code>saving(filename [, replace])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph [(graphopts)]</code>	graph results; see [PSS] <b>power, graph</b>
Reporting	
<code>show</code>	display group-specific numbers of events and, in the presence of loss to follow-up, numbers of losses
<code>show(showspec)</code>	display group-specific numbers of events, numbers of losses, and event probabilities
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 **numlist**.

`notitle` does not appear in the dialog box.

<i>effect</i>	Description
<code>hratio</code>	hazard ratio
<code>lnhratio</code>	log hazard-ratio
<code>hdifference</code>	difference between hazard rates
<code>lnhdifference</code>	difference between log hazard-rates (equivalent to log hazard-ratio)
<code>difference</code>	synonym for <code>hdifference</code>

<i>showspec</i>	Description
<code>events</code>	numbers of events
<code>losses</code>	numbers of losses
<code>eventprobs</code>	event probabilities
<code>all</code>	all the above

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
alpha	significance level	$\alpha$
power	power	$1 - \beta$
beta	type II error probability	$\beta$
N	total number of subjects	$N$
N1	number of subjects in the control group	$N_1$
N2	number of subjects in the experimental group	$N_2$
nratio	ratio of sample sizes, experimental to control	$N_2/N_1$
delta	effect size	$\delta$
s1	survival probability in the control group	$S_1(t)$
s2	survival probability in the experimental group	$S_2(t)$
time	reference survival time	$t$
h1	hazard rate in the control group	$\lambda_1$
h2	hazard rate in the experimental group	$\lambda_2$
hdiff	difference between hazard rates	$\lambda_2 - \lambda_1$
hratio	hazard ratio	$\Delta$
lnhratio	log hazard-ratio	$\ln(\Delta)$
studytime	duration of a study	$T$
fperiod	follow-up period	$f$
aperiod	accrual period	$r$
aprob	proportion of subjects accrued by time <i>atime</i> (or by <i>aptime</i> × 100% of accrual period)	$p_a$
aptime	proportion of an accrual period by which <i>aprob</i> × 100% of subjects are accrued	$t_a/r$
atime	reference accrual time	$t_a$
ashape	shape of the accrual distribution	$\gamma$
E0	total number of events under $H_0$	$E_0$
E01	number of events in the control group under $H_0$	$E_{01}$
E02	number of events in the experimental group under $H_0$	$E_{02}$
Ea	total number of events under $H_a$	$E_a$
Ea1	number of events in the control group under $H_a$	$E_{a1}$
Ea2	number of events in the experimental group under $H_a$	$E_{a2}$
Pr_E01	control-group probability of an event under $H_0$	$\text{Pr}_{E_{01}}$
Pr_E02	experimental-group probability of an event under $H_0$	$\text{Pr}_{E_{02}}$
Pr_Ea1	control-group probability of an event under $H_a$	$\text{Pr}_{E_{a1}}$
Pr_Ea2	experimental-group probability of an event under $H_a$	$\text{Pr}_{E_{a2}}$
lossprob	proportion of subjects lost to follow-up in the control and experimental groups	$L(t_L)$
lossprob1	proportion of subjects lost to follow-up in the control group	$L_1(t_L)$
lossprob2	proportion of subjects lost to follow-up in the experimental group	$L_2(t_L)$

<code>losstime</code>	reference loss-to-follow-up time	$t_L$
<code>losshaz</code>	loss hazard rate in the control and experimental groups	$\eta$
<code>losshaz1</code>	loss hazard rate in the control group	$\eta_1$
<code>losshaz2</code>	loss hazard rate in the experimental group	$\eta_2$
<code>L0</code>	total number of losses under $H_0$	$L_0$
<code>L01</code>	number of losses in the control group under $H_0$	$L_{01}$
<code>L02</code>	number of losses in the experimental group under $H_0$	$L_{02}$
<code>La</code>	total number of losses under $H_a$	$L_a$
<code>La1</code>	number of losses in the control group under $H_a$	$L_{a1}$
<code>La2</code>	number of losses in the experimental group under $H_a$	$L_{a2}$
<code>target</code>	target parameter; synonym for <code>h2</code> or <code>hratio</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if option `beta()` is specified.

Column `hratio` is shown in the default table if option `hratio()` is specified or implied by the command.

Columns `nratio` and `lnhratio` are shown in the default table if the corresponding options are specified.

Columns `h1`, `h2`, `s1`, and `s2` are available and are shown in the default table when the corresponding command arguments are specified.

Columns `time`, `studytime`, `fperiod`, `aperiod`, `aprob`, `aptime`, `atime`, `ashape`, `losshaz`, `losshaz1`, `losshaz2`, `lossprob`, `lossprob1`, `lossprob2`, and `losstime` are available and are shown in the default table when the corresponding options are specified.

Columns containing numbers of events, numbers of losses, and probabilities of an event are displayed if specified or if respective options `show(events)`, `show(losses)`, or `show(eventprobs)` are specified. If `show` is specified, numbers of events and losses are displayed. If `show(all)` is specified, numbers of events, numbers of losses, and probabilities are displayed.

## Options

### Main

`time(#)` specifies a fixed time  $t$  (reference survival time) such that the proportions of subjects in the control and experimental groups still alive past this time point are as specified in  $s_1$  and  $s_2$ . If this option is specified, the input parameters  $s_1$  and  $s_2$  are the survival probabilities  $S_1(t)$  and  $S_2(t)$ . Otherwise, the input parameters are assumed to be hazard rates  $\lambda_1$  and  $\lambda_2$  given as  $h_1$  and  $h_2$ , respectively.

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `nfractional`; see [PSS] [power](#).

`hratio(numlist)` specifies the hazard ratio (effect size) of the experimental group to the control group. The default is `hratio(0.5)`. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by a test with a certain power. If  $h_1$  and  $h_2$  (or  $s_1$  and  $s_2$ ) are given, `hratio()` is not allowed and the hazard ratio is computed as  $h_2/h_1$  [or  $\ln(s_2)/\ln(s_1)$ ]. Also see [Alternative ways of specifying effect](#) for various specifications of an effect size.

This option is not allowed with the effect-size determination and may not be combined with `lnhratio()` or `hdifference()`.

`lnhratio(numlist)` specifies the log hazard-ratio (effect size) of the experimental group to the control group. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by a test with a certain power. If  $h_1$  and  $h_2$  (or  $s_1$  and  $s_2$ ) are given, `lnhratio()` is not allowed and the log hazard-ratio is computed as  $\ln(h_2/h_1)$  [or  $\ln\{\ln(s_2)/\ln(s_1)\}$ ]. Also see [Alternative ways of specifying effect](#) for various specifications of an effect size.

This option is not allowed with the effect-size determination and may not be combined with `hratio()` or `hdifference()`.

`hdifference(numlist)` specifies the difference between the experimental-group hazard rate and the control-group hazard rate. It requires that the control-group hazard rate, the command argument  $h_1$ , is specified. `hdifference()` provides a way of specifying an effect size; see *Alternative ways of specifying effect* for details.

This option is not allowed with the effect-size determination and may not be combined with `hratio()` or `lnhratio()`.

`loghazard` requests sample-size or power computation for the test of the difference between log hazards (or the log hazard-ratio test). This option implies uniform accrual. By default, the test of the difference between hazards is assumed.

`unconditional` requests that the unconditional approach be used for sample-size or power computation; see *The conditional versus unconditional approaches* and *Methods and formulas for details*.

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. *effect* is one of `hratio`, `lnhratio`, `hdifference`, or `lnhdifference`. By default, the effect size `delta` is a hazard ratio, `effect(hratio)`, for a hazard-ratio test and a log hazard-ratio, `effect(lnhratio)`, for a log hazard-ratio test (when `schoenfeld` is specified).

`onesided`, `parallel`; see [PSS] **power**.

---

Accrual/Follow-up

---

`studytime(numlist)` specifies the duration of the study,  $T$ . By default, it is assumed that subjects are followed up until the last subject experiences an event (fails). The (minimal) follow-up period is defined as the length of the period after the recruitment of the last subject to the study until the end of the study. If  $r$  is the length of an accrual period and  $f$  is the length of the follow-up period, then  $T = r + f$ . You can specify only two of the three options `studytime()`, `fperiod()`, and `aperiod()`.

`fperiod(numlist)` specifies the follow-up period of the study,  $f$ . By default, it is assumed that subjects are followed up until the last subject experiences an event (fails). The (minimal) follow-up period is defined as the length of the period after the recruitment of the last subject to the study until the end of the study. If  $T$  is the duration of a study and  $r$  is the length of an accrual period, then the follow-up period is  $f = T - r$ . You can specify only two of the three options `studytime()`, `fperiod()`, and `aperiod()`.

`aperiod(numlist)` specifies the accrual period,  $r$ , during which subjects are to be recruited into the study. The default is `aperiod(0)`, meaning no accrual. You can specify only two of the three options `studytime()`, `fperiod()`, and `aperiod()`.

`aprob(numlist)` specifies the proportion of subjects expected to be accrued by time  $t^*$  according to the truncated exponential distribution. The default is `aprob(0.5)`. This option is useful when the shape parameter is unknown but the proportion of accrued subjects at a certain time is known. `aprob()` is often used in conjunction with `aptime()` or `atime()`. This option may not be specified with `ashape()` or `loghazard` and requires specifying a nonzero accrual period in `aperiod()`.

`aptime(numlist)` specifies the proportion of the accrual period,  $t^*/r$ , by which the proportion of subjects specified in `aprob()` is expected to be accrued according to the truncated exponential distribution. The default is `aptime(0.5)`. This option may not be combined with `atime()`, `ashape()`, or `loghazard` and requires specifying a nonzero accrual period in `aperiod()`.

- `atime(numlist)` specifies the time point  $t^*$ , reference accrual time, by which the proportion of subjects specified in `aprob()` is expected to be accrued according to the truncated exponential distribution. The default value is  $0.5 \times r$ . This option may not be combined with `aptime()`, `ashape()`, or `loghazard` and requires specifying a nonzero accrual period in `aperiod()`. The value in `atime()` may not exceed the value in `aperiod()`.
- `ashape(numlist)` specifies the shape,  $\gamma$ , of the truncated exponential accrual distribution. The default is `ashape(0)`, meaning uniform accrual. This option is not allowed in conjunction with `loghazard` and requires specifying a nonzero accrual period in `aperiod()`.
- `lossprob(numlist)` specifies the proportion of subjects lost to follow-up by time `losstime()` in the control and the experimental groups. The default is `lossprob(0)`, meaning no losses to follow-up. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob1()`, `lossprob2()`, `losshaz()`, `losshaz1()`, or `losshaz2()`.
- `lossprob1(numlist)` specifies the proportion of subjects lost to follow-up by time `losstime()` in the control group. The default is `lossprob1(0)`, meaning no losses to follow-up in the control group. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob()`, `losshaz()`, `losshaz1()`, or `losshaz2()`.
- `lossprob2(numlist)` specifies the proportion of subjects lost to follow-up by time `losstime()` in the experimental group. The default is `lossprob2(0)`, meaning no losses to follow-up in the experimental group. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob()`, `losshaz()`, `losshaz1()`, or `losshaz2()`.
- `losstime(numlist)` specifies the time at which the proportion of subjects specified in `lossprob()` or `lossprob1()` and `lossprob2()` is lost to follow-up, also referred to as the reference loss to follow-up time. The default is `losstime(1)`. This option requires specifying `lossprob()`, `lossprob1()`, or `lossprob2()`.
- `losshaz(numlist)` specifies an exponential hazard rate of losses to follow-up common to both the control and the experimental groups. The default is `losshaz(0)`, meaning no losses to follow-up. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob()`, `lossprob1()`, `lossprob2()`, `losshaz1()`, or `losshaz2()`.
- `losshaz1(numlist)` specifies an exponential hazard rate of losses to follow-up,  $\eta_1$ , in the control group. The default is `losshaz1(0)`, meaning no losses to follow-up in the control group. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob()`, `lossprob1()`, `lossprob2()`, or `losshaz()`.
- `losshaz2(numlist)` specifies an exponential hazard rate of losses to follow-up,  $\eta_2$ , in the experimental group. The default is `losshaz2(0)`, meaning no losses to follow-up in the experimental group. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob()`, `lossprob1()`, `lossprob2()`, or `losshaz()`.

## Table

`table`, `table()`, `notable`; see [PSS] [power, table](#).

`saving()`; see [PSS] [power](#).

## Graph

`graph`, `graph()`; see [PSS] [power, graph](#). Also see the [column](#) table for a list of symbols used by the graphs.

`show` and `show(showspec)` specify to display additional output containing the numbers of events, losses to follow-up, and event probabilities. If `show` is specified, group-specific numbers of events and, in the presence of losses to follow-up, group-specific numbers of losses to follow-up are displayed for the null and alternative hypotheses. With the table output, the numbers are displayed as additional columns.

`showspec` may contain any combination of `events`, `losses`, `eventprobs`, and `all.events` displays the group-specific numbers of events under the null and alternative hypotheses. `losses`, if present, displays group-specific numbers of losses under the null and alternative hypotheses. `eventprobs` displays group-specific event probabilities under the null and alternative hypotheses. `all` displays all the above.

The following option is available with `power exponential` but is not shown in the dialog box: `notitle`; see [PSS] `power`.

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power exponential*

*Alternative ways of specifying effect*

*Computing sample size*

*Computing sample size in the absence of censoring*

*Computing sample size in the presence of censoring*

*Nonuniform accrual*

*Exponential losses to follow-up*

*The conditional versus unconditional approaches*

*Link to the sample-size and power computation for the log-rank test*

*Computing power*

*Testing hypotheses about two exponential survivor functions*

This entry describes the `power exponential` command and the methodology for power and sample-size analysis for a two-sample comparison of exponential survivor functions. See [PSS] `intro` for a general introduction to power and sample-size analysis and [PSS] `power` for a general introduction to the `power` command using hypothesis tests. See *Survival data* in [PSS] `intro` for an introduction to power and sample-size analysis for survival data.

## Introduction

Let  $S_1(t)$  and  $S_2(t)$  be the exponential survivor functions with hazard rates  $\lambda_1$  and  $\lambda_2$  in the control and experimental groups, respectively. Define  $\delta$  to be the treatment effect that can be expressed as a difference,  $\psi = \lambda_2 - \lambda_1$ , between hazard rates or as the log of the hazard ratio (a difference between log hazard-rates),  $\ln(\Delta) = \ln(\lambda_2/\lambda_1) = \ln(\lambda_2) - \ln(\lambda_1)$ . Negative values of the treatment effect  $\delta$  imply the superiority of the experimental treatment over the standard (control) treatment. Denote  $r$  and  $T$  to be the length of the accrual period and the total duration of the study, respectively. Then, the follow-up period  $f$  is  $f = T - r$ .

Consider a study designed to compare the exponential survivor functions,  $S_1(t) = e^{-\lambda_1 t}$  and  $S_2(t) = e^{-\lambda_2 t}$ , of the two treatment groups. The disparity in survivor functions may be tested using the hazards  $\lambda_1$  and  $\lambda_2$  for the exponential model. Depending on the definition of the treatment effect  $\delta$ , two test statistics based on the difference and on the log ratio of the hazards may be used to conduct tests of the difference between survivor functions using respective null hypotheses,  $H_0: \psi = 0$  and  $H_0: \ln(\Delta) = 0$ .



The basic formula for the sample-size and power calculations for the test of  $H_0: \psi = 0$  is proposed by Lachin (1981). He also derives the equation relating the sample size and power allowing for uniform accrual of subjects into the study over the period from 0 to  $r$ . Lachin and Foulkes (1986) extend this formula to truncated exponential accrual over the interval 0 to  $r$  and exponential losses to follow-up over the interval 0 to  $T$ .

The simplest method for the sample-size and power calculations for the test of  $H_0: \ln(\Delta) = 0$  is presented by George and Desu (1974). Rubinstein, Gail, and Santner (1981) extend their method to account for uniform accrual and exponential losses to follow-up and apply it to planning the duration of a survival study. The formula that relates the sample size and power for this test and takes into account the uniform accrual and exponential losses to follow-up is formulated by Lakatos and Lan (1992), based on the derivations of Rubinstein, Gail, and Santner (1981).

You can use `power exponential` to

- compute required number of events and sample size when you know power and effect size; or
- compute power when you know sample size (number of events) and effect size.

You can also supply effect size as hazard rates, survival probabilities, hazard ratio, or log hazard-ratio; adjust results for censoring; adjust results for uniform or exponential accrual; adjust results for group-specific exponentially distributed losses to follow-up; and compute results using the conditional or unconditional approach.

## Using `power exponential`

`power exponential` computes sample size or power for a test comparing two exponential survivor functions. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced- or equal-allocation design; see [PSS] [unbalanced designs](#) for a description of how to specify an unbalanced design.

To compute a total sample size, you specify an effect size and, optionally, the power of the test in the `power()` option. The default power is set to 0.8. By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see [Fractional sample sizes](#) in [PSS] [unbalanced designs](#) for an example. The `nfractional` option is allowed only for sample-size determination.

To compute power, you must specify the total sample size in the `n()` option and an effect size.

An effect size may be specified as a hazard ratio in option `hratio()`, as a log hazard-ratio in option `lnhratio()`, or as a difference between hazard rates in option `hdifference()`. By default, a hazard ratio of 0.5 is assumed. For a fixed-duration study, the control-group hazard rate  $h_1$  or the control-group survival probability  $s_1$  must also be specified. See [Alternative ways of specifying effect](#) below for details.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()` or specify one of the group sizes and `nratio()` when computing power or effect size. See [Two samples](#) in [PSS] [unbalanced designs](#) for more details.

If the `time()` option is specified, the command's input parameters are the values of survival probabilities in the control (or the less favorable) group,  $S_1(t)$ , and in the experimental group,  $S_2(t)$ , at a fixed time,  $t$  (reference survival time), specified in `time()`, given as  $s_1$  and  $s_2$ , respectively. Otherwise, the input parameters are assumed to be the values of the hazard rates in the control group,  $\lambda_1$ , and in the experimental group,  $\lambda_2$ , given as  $h_1$  and  $h_2$ , respectively. If survival probabilities are

specified, they are converted to hazard rates by using the formula for the exponential survivor function and the value of time  $t$  in `t()`.

By default, the estimates of sample sizes or power for the test of the difference between hazards are reported. This may be changed to the test versus the difference between log hazards by using the `loghazard` option. The default conditional approach may be replaced with the unconditional approach by using `unconditional`; see *The conditional versus unconditional approaches*.

If the duration of a study ( $T$ ) in option `studytime()`, the length of a follow-up period ( $f$ ) in option `fperiod()`, or the length of an accrual period ( $r$ ) in option `aperiod()` is not specified, then the study is assumed to continue until all subjects experience an event (failure), regardless of how much time is required. If only `studytime()` is specified or only `fperiod()` is specified, the length of the accrual period is assumed to be zero and the follow-up period equals the duration of the study. If only `aperiod()` is specified, the length of the follow-up is assumed to be zero and the duration of the study equals the length of the accrual period (continuous accrual until the end of the study). If either `aperiod()` or `fperiod()` is specified with `studytime()`, the other one is computed using the relationship  $T = r + f$ . If both `aperiod()` or `fperiod()` are specified, a fixed-duration study of length  $T = r + f$  is assumed.

If an accrual period of length  $r$  is specified in the `aperiod()` option, uniform accrual over the period  $[0, r]$  is assumed. The accrual distribution may be changed to truncated exponential when the shape parameter is specified in `ashape()`. The combination of the `aprob()` and `aptime()` (or `atime()`) options may be used in place of the `ashape()` option to request the desired shape of the truncated exponential accrual. For examples, see *Nonuniform accrual*.

To take into account exponential losses to follow-up, the `losshaz()` or `lossprob()` and `losstime()` options may be used. Instead of specifying losses common to both groups, you can use options `losshaz1()` and `losshaz2()` or `lossprob1()` and `lossprob2()` to specify group-specific losses to follow-up. For examples, see *Exponential losses to follow-up*.

## Alternative ways of specifying effect

`power exponential` provides several ways to specify the disparity between the control-group and experimental-group survivor functions for sample-size and power determinations. You can specify group hazard rates or group survival probabilities at a fixed time  $t$  directly. If survival probabilities are specified, they are converted to hazard rates by using the formula for the exponential survivor function and the value of time  $t$ . Alternatively, you can specify the control-group hazard rate or the control-group survival probability and an effect size expressed as a hazard ratio, a log hazard-ratio, or a difference between the two hazard rates. The corresponding experimental-group hazard rate will then be computed using the specified values of the control-group hazard rate and effect size.

By default, `power exponential` performs computation assuming a hazard ratio of 0.5. You can use the `hratio()` option to specify a different value for the hazard ratio or you can use the `lnhratio()` option to specify an effect size as a log hazard-ratio. If a control-group hazard rate or survival probability is specified, you can also specify an effect size as a difference between the experimental-group and control-group hazard rates in option `hdifference()`.

For a fixed-duration study when not all subjects experience an event by the end of the study, a control-group hazard rate or a control-group survival probability at time  $t$  must be specified in addition to an effect size.

You specify the control-group hazard rate  $h_1$  following the command name. You can use any of the three options mentioned above to specify an effect size. The experimental-group hazard rate is then computed using the specified values of the control-group hazard rate and effect size.

```
power exponential  $h_1$  [ , hratio() | lnhratio() | hdifference() ... ]
```

Alternatively, you can specify the experimental-group hazard rate  $h_2$  directly.

```
power exponential  $h_1$   $h_2$  [ , ... ]
```

Instead of the control-group hazard rate, you can specify the control-group survival probability  $s_1$  at time  $t$ ; the reference time  $t$  must be specified in option `time()`.

```
power exponential  $s_1$  , time(#) [hratio() | lnhratio() | hdifference() ... ]
```

Similarly to hazard rates, you can specify the experimental-group survival probability at time  $t$  instead of an effect size.

```
power exponential  $s_1$   $s_2$  , time(#) [ ... ]
```

The displayed effect size `delta` corresponds to the difference between hazard rates (or the hazard ratio if the control-group hazard is not specified) for the hazard-difference test and to the log hazard-ratio for the log hazard-ratio (or log hazard-difference) test when the `loghazard` option is specified. You can change this by specifying the `effect()` option: `effect(hratio)` reports the hazard ratio, `effect(lnhratio)` reports the log hazard-ratio, and `effect(hdifference)` reports the difference between the experimental-group and control-group hazard rates.

In the following sections, we describe the use of `power exponential` accompanied by examples for computing sample size and power.

## Computing sample size

To compute sample size and number of events, you must specify an effect size and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified. A hazard ratio of 0.5 is assumed if an effect size is not specified. See [Alternative ways of specifying effect](#) for various ways of specifying an effect size.

Consider the following two types of survival studies: the first type, a type I study, is when investigators have enough resources to monitor the subjects until all of them experience an event (failure) and the second type, a type II study, is when the study terminates after a fixed period of time, regardless of whether all subjects experienced an event by that time.

## Computing sample size in the absence of censoring

In this subsection we explore sample-size estimates using different approximations for a type I study. Examples of sample-size determination for a type II study are presented in the next subsection.

In survival studies, the requirement for the sample size is based on the requirement to observe a certain number of events (failures) to ensure a prespecified power of a test to detect a difference in survivor functions. For a type I study, the number of subjects required for the study is the same as the number of events required to be observed in the study because all subjects experience an event by the end of the study.

### ▷ Example 1: Sample size using the Lachin method

Consider an example from [Lachin \(1981, 107\)](#). A clinical trial is to be conducted to compare the survivor functions in the control and the experimental groups with a one-sided exponential test, based on the difference between hazards, of the superiority of a new treatment ( $H_a: \psi < 0$ ) for a disease with moderate levels of mortality. Subjects in the control group receive a standard treatment and subjects in the experimental group receive a new treatment. From previous studies the yearly hazard rate for the standard treatment was found to be  $\lambda_1 = 0.3$ , corresponding to 50% survival after 2.3 years. The investigators would like to know how many subjects are required to detect a reduction in hazard to  $\lambda_2 = 0.2$  ( $H_a: \psi = -0.1$ ), which corresponds to an increase in survival to 63% at 2.3 years, with 90% power, equal-sized groups, and a significance level,  $\alpha$ , of 0.05.

To obtain the estimate of the sample size for the above study, we supply hazard rates 0.3 and 0.2 as arguments and specify the `power(0.9)` option for 90% power and the `onesided` option for a one-sided test.

```
. power exponential 0.3 0.2, power(0.9) onesided
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.1000 (hazard difference)

Survival information:
    h1 =    0.3000
    h2 =    0.2000

Estimated sample sizes:
    N =        218
    N per group =    109
```

From the output, a total of 218 events (subjects) must be observed (recruited) in a study to ensure a power of 90% of a one-sided exponential test to detect a 13% increase in survival probability of subjects in the experimental group with  $\alpha = 0.05$ . Our estimate of 218 of the total number of subjects (109 per group) required for the study is the same as the one reported in [Lachin \(1981, 107\)](#). ◀

### ▷ Example 2: Sample size using the George–Desu method

[Example 1](#) reports the sample size obtained using the approximation of [Lachin \(1981\)](#) for the test based on the hazard difference. To obtain the sample size using the approximation of [George and Desu \(1974\)](#), for the equivalent alternative  $H_a: \ln(\Delta) = -0.4055$  (a test based on the log of the hazard ratio), we need to specify the `loghazard` option.

```

. power exponential 0.3 0.2, power(0.9) onesided loghazard
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log hazard-ratio, conditional
Ho: ln(HR) = 0 versus Ha: ln(HR) < 0

Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =   -0.4055 (log hazard-ratio)

Survival information:
      h1 =    0.3000
      h2 =    0.2000

Estimated sample sizes:
      N =      210
      N per group = 105

```

The George–Desu method yields a slightly smaller estimate (210) of the total number of events (subjects). [George and Desu \(1974\)](#) studied the accuracy of the two approximations based on  $\psi$  and  $\ln(\Delta)$  and concluded that the former is slightly conservative; that is, it gives slightly larger sample-size estimates. The latter was found to be accurate to one or two units of the exact solution for equal-sized groups.

◀

## □ Technical note

The approach from [example 2](#) may also be used to obtain an approximation to the sample size or power for the exact  $F$  test of equality of two exponential mean analysis (life) times (using the relation between a mean and a hazard rate of the exponential distribution,  $\mu = 1/\lambda$ ).

For example, the sample size of 210 obtained above may be used as an approximation to the number of subjects required in a study of which the goal is to detect an increase in a mean analysis (life) time of the experimental group from  $3.33 = 1/0.3$  to  $5 = 1/0.2$  by using the one-sided 5%-level  $F$  test with 90% power.

The test statistic of the  $F$  test is a ratio of two sample means from two exponential distributions that has an exact  $F$  distribution. The George–Desu method is based on the normal approximation of the distribution of the log of this test statistic. [George and Desu \(1974\)](#) studied this approximation for equal-sized groups and some common values of significance levels, powers, and hazard ratios and found it to be accurate to one or two units of the exact solution.

□

## ▷ Example 3: Alternative ways of specifying effect

In [Alternative ways of specifying effect](#), we described various ways in which the survival information of the groups can be supplied to `power exponential`. Here we demonstrate several examples.

In [example 1](#), we specified the survival information by supplying the control-group and experimental-group hazard rates.

Instead of the experimental-group hazard rate, we can specify the difference between hazards in the `hdifference()` option and obtain identical results.

```
. power exponential 0.3, power(0.9) onesided hdifference(-0.1)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =   -0.1000 (hazard difference)

Survival information:
      h1 =    0.3000
      h2 =    0.2000
      h2 - h1 =   -0.1000

Estimated sample sizes:
      N =        218
      N per group =    109
```

We can redisplay the effect size delta as a hazard ratio instead of the hazard difference:

```
. power exponential 0.3, power(0.9) onesided hdifference(-0.1) effect(hratio)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =    0.6667 (hazard ratio)

Survival information:
      h1 =    0.3000
      h2 =    0.2000
      h2 - h1 =   -0.1000

Estimated sample sizes:
      N =        218
      N per group =    109
```

We can specify the hazard ratio of  $0.2/0.3 = 0.66667$  in the `hratio()` option instead of `hdifference(-0.1)`.

```
. power exponential 0.3, power(0.9) onesided hratio(0.6667)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =   -0.1000 (hazard difference)

Survival information:
      h1 =    0.3000
      h2 =    0.2000
      hratio =    0.6667

Estimated sample sizes:
      N =        218
      N per group =    109
```

We can obtain the same results from `power exponential` if we specify the control-group survival probability of 0.5 at time  $t = 2.3$ .

```
. power exponential 0.5, time(2.3) power(0.9) onesided hratio(0.6667)
note: input parameters are survival probabilities
Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.1004 (hazard difference)
Survival information:
    h1 =    0.3014          s1 =    0.5000
    h2 =    0.2009          s2 =    0.6299
    hratio = 0.6667        t =    2.3000
Estimated sample sizes:
    N =        218
    N per group =    109
```

We can also specify the experimental-group survival probability of 0.63 at time  $t = 2.3$  directly instead of specifying the hazard ratio.

```
. power exponential 0.5 0.63, time(2.3) power(0.9) onesided
note: input parameters are survival probabilities
Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.1005 (hazard difference)
Survival information:
    h1 =    0.3014          s1 =    0.5000
    h2 =    0.2009          s2 =    0.6300
    t =    2.3000
Estimated sample sizes:
    N =        218
    N per group =    109
```

◀

## ► Example 4: Unbalanced design

By default, `power exponential` computes sample size for a balanced- or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size for an unbalanced design by specifying the `nratio()` option.

In [example 1](#), we assumed the same numbers of subjects in the two groups. Suppose that we anticipate to recruit twice as many subjects in the experimental group, that is,  $n_2/n_1 = 2$ . We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```

. power exponential 0.3 0.2, power(0.9) onesided nratio(2)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.1000 (hazard difference)
    N2/N1 =    2.0000

Survival information:
    h1 =    0.3000
    h2 =    0.2000

Estimated sample sizes:
    N =      242
    N1 =     81
    N2 =    161
    N2/N1 =   1.9877

```

We need a total of 242 subjects—81 in the control group and 161 in the experimental group.

When different from the specified allocation rate, `power exponential` also displays the actual allocation rate corresponding to the reported rounded group sample sizes. If you wish, you can specify the `nfractional` option to see sample sizes without rounding; see [Fractional sample sizes](#) in [PSS] [unbalanced designs](#) for more information.

Also see [Two samples](#) in [PSS] [unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.

◀

## Computing sample size in the presence of censoring

Often in practice, investigators may not have enough resources to continue a study until all subjects experience an event and, therefore, plan to terminate the study after a fixed period,  $T$ . Some subjects may not experience an event by the end of the study, in which case the (administrative) censoring of subjects occurs. In the presence of censoring, the number of subjects required in a study will be larger than the number of events required to be observed in the study.

We investigate how terminating the study after some fixed period,  $T$ , before all subjects experience an event affects the requirements for the sample size. The duration of a study is divided into two phases: an accrual phase of a length  $r$ , during which subjects are recruited to the study, and a follow-up phase of a length  $f$ , during which subjects are followed up until the end of the study and no new subjects enter the study. The duration of a study,  $T$ , is the sum of the lengths of the two phases.

Consider the following study designs. In the first study design, *A*, each subject is followed up for a length of time  $T$ . Here the minimum follow-up time  $f$  is equal to  $T$ , and, consequently,  $r = 0$ . In practice, however, subjects will often enter the study at random times and will be followed up until the end of a study at time  $T$ , in which case the subjects observed later will have a shorter follow-up than subjects who entered the study at the beginning. Therefore, the minimum follow-up time  $f$  will be less than  $T$ , and  $r$  will be equal to  $T - f$ . In this case the length of the accrual period,  $r$ , must be taken into account in the computations. In the presence of an accrual period, subjects may be recruited continuously during a period of length  $T$  ( $r = T$ ,  $f = 0$ ) for the second study design, *B*. Or subjects may be recruited for a fixed period,  $r$ , and then followed up for a period of time,  $f$ ,



during which no new subjects enter the trial, so that the total duration of study is  $T = r + f$  (the third design, C).

### ▷ Example 5: Sample size in the presence of accrual and follow-up periods

Continuing with [example 1](#), assume that the investigators have resources to continue the study for only 5 years,  $T = 5$ . We specify the duration of the study in the `studytime()` option, and we tabulate sample-size values for different lengths of an accrual period specified as a list (see [\[U\] 11.1.8 numlist](#)) in `aperiod()`. For simplicity, we use the `table()` option to obtain a table containing only columns `power`, `N`, `aperiod`, `fperiod`, `h1`, `h2`, and `alpha`.

```
. power exponential 0.3 0.2, power(0.9) onesided aperiod(0(1)5) studytime(5)
> table(power N aperiod fperiod h1 h2 alpha)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1
```

power	N	aperiod	fperiod	h1	h2	alpha
.9	304	0	5	.3	.2	.05
.9	322	1	4	.3	.2	.05
.9	344	2	3	.3	.2	.05
.9	378	3	2	.3	.2	.05
.9	426	4	1	.3	.2	.05
.9	502	5	0	.3	.2	.05

Note: Uniform accrual; 50% accrued by 50% of accrual period.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS\] power, table](#). If you wish to produce a power plot, see [\[PSS\] power, graph](#).

The first and the last entries of the above table correspond to the extreme cases of no accrual (design *A*) and no follow-up (design *B*), respectively. When `aperiod()` is specified, a uniform accrual is assumed which implies, for example, that 50% of the subjects will be recruited once 50% of the accrual period has elapsed.

For design *A*, the estimate of the sample size, 304, is larger than the earlier estimate of 218 from [example 1](#). That is, if the study in example 1 terminates after 5 years, the requirement for the sample size increases by 39% to ensure that the same number of 218 events is observed.

By trying different values of the follow-up period, we may find that a 30-year follow-up is required if the investigators can recruit no more than 218 subjects: 30 years is required to observe an event for all subjects in this study.

```

. power exponential 0.3 0.2, power(0.9) onesided fperiod(30)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.1000 (hazard difference)

Accrual and follow-up information:
    duration =   30.0000
    follow-up =   30.0000

Survival information:
    h1 =    0.3000
    h2 =    0.2000

Estimated sample sizes:
    N =      218
    N per group =    109

```

Returning to our [table](#), for design *B*, instead of being monitored for 5 years, subjects are continuously recruited throughout those 5 years; the total sample size increases from 304 to 502. The reason for such an increase is that the average analysis time (the time when a subject is at risk of a failure) decreases from 5 to 2.5 and, therefore, reduces the probability of a subject failing by the end of the study.

In general, the estimates of the total sample size steadily increase as the length of the follow-up decreases. That is, the presence of a follow-up period reduces the requirement for the number of subjects in the study. For example, a clinical trial with a 3-year uniform accrual and a 2-year follow-up needs a total of 378 subjects (189 per group) compared with the total of 502 subjects required for a study with no follow-up and a 5-year accrual.

◀

## ▷ Example 6: Uniform accrual

In [example 5](#), we investigated the effect of the length of accrual on sample size for a type II study when not all subjects experience an event by the end of the study. We specified the length of the accrual period in option `aperiod()` and the duration of the study in option `studytime()`. When `aperiod()` is specified, the accrual distribution is assumed to be uniform, that is, 10% of the subjects are expected to be recruited once 10% of the accrual period has elapsed, 25% of subjects are expected to be recruited once 25% of the accrual period has elapsed, 50% of subjects are expected to be recruited once 50% of the accrual period has elapsed, and so on. Let's compute the sample size for a study with a 3-year uniform accrual and a 2-year follow-up. We use options `aperiod()` and `fperiod()` to specify the accrual and follow-up periods, respectively.

```

. power exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.1000 (hazard difference)

Accrual and follow-up information:
    duration =    5.0000
    follow-up =    2.0000
    accrual =    3.0000 (uniform)

Survival information:
    h1 =    0.3000
    h2 =    0.2000

Estimated sample sizes:
    N =        378
    N per group =    189

```

The required total sample size is 378 with 189 subjects per group. This is the same sample size we obtained in the table from [example 5](#) with the corresponding values of the accrual and follow-up periods.

◀

## Nonuniform accrual

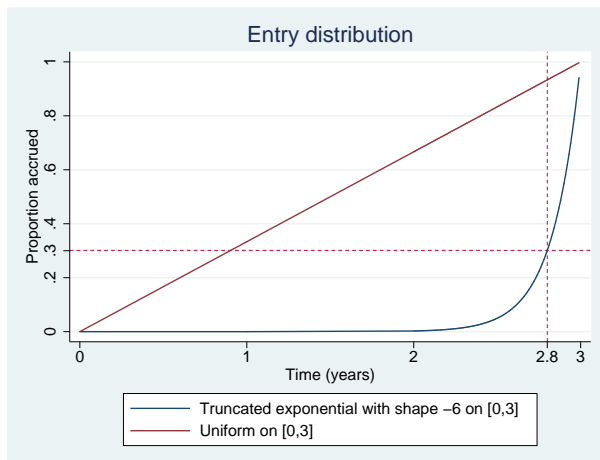
In the presence of an accrual period, `power exponential` performs computations assuming uniform accrual over the period of time  $r$ , specified in `aperiod()`. The assumption of uniform accrual may be relaxed by requesting a truncated exponential accrual over the interval 0 to  $r$  with shape  $\gamma$  as specified in `ashape(#)`. If an estimate of  $\gamma$  is unavailable, the proportion of subjects expected to be recruited,  $G(t^*)$ , may be specified in `aprob()` along with either the fixed time by which the subjects were recruited,  $t^*$ , in option `atime()` or the elapsed proportion of the accrual period,  $t^*/r$ , in option `aptime()`. This information is used to find the corresponding  $\gamma$  by using

$$G(t^*) = \{1 - \exp(-\gamma t^*)\} / \{1 - \exp(-\gamma r)\}$$

Also see [Cleves et al. \(2010, sec. 16.2\)](#) for more information, and see [Methods and formulas](#) for technical details.

### ► Example 7: Truncated exponential entry distribution

Continuing with [example 6](#), we investigate the influence of nonuniform accrual on the estimate of the sample size for a study with a 3-year accrual and a 2-year follow-up. Suppose that the recruitment of subjects to the study is slow for most of the accrual period and increases rapidly toward the end of the recruitment. Consider an extreme case of such an accrual corresponding to shape parameter  $-6$ . The graph of a uniform entry distribution and an exponential entry distribution with shape  $-6$  truncated over  $[0, 3]$  is given below.



From the above graph, the accrual of subjects is extremely slow during most of the recruitment period, with 70% of subjects being recruited within the last few months of a 3-year accrual period. Stated another way, according to the graph, only 30% of subjects are expected to be recruited during the first 2.8 years.

To obtain the estimate of the sample size for this study, we type

```
. power exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2) ashape(-6)
note: input parameters are hazard rates
```

Estimated sample sizes for two-sample comparison of survivor functions

Exponential test, hazard difference, conditional

Ho:  $h_2 = h_1$  versus Ha:  $h_2 < h_1$

Study parameters:

```
alpha = 0.0500
power = 0.9000
delta = -0.1000 (hazard difference)
```

Accrual and follow-up information:

```
duration = 5.0000
follow-up = 2.0000
accrual = 3.0000 (exponential)
accrual(%) = 50.00 (by time t*)
t* = 2.8845 (96.15% of accrual)
```

Survival information:

```
h1 = 0.3000
h2 = 0.2000
```

Estimated sample sizes:

```
N = 516
N per group = 258
```

and conclude that 516 subjects have to be recruited to this study. This sample size ensures 90% power of a one-sided, 5%-level test to detect a reduction in hazard from 0.3 to 0.2 when the accrual of subjects follows the considered truncated exponential distribution. For this extreme case of a negative truncated exponential entry distribution (the concave entry distribution), the estimate of the sample, 516, increases substantially compared with an estimate of 378 from [example 6](#), which assumes a uniform entry distribution. On the other hand, a truncated exponential distribution with positive values of the shape parameter (convex entry distribution) will reduce the requirement for the sample size when compared with uniform accrual.

Suppose that we do not know (or do not wish to guess) the value of the shape parameter. The only information available to us from the above graph is that 30% of the subjects are expected to be recruited in the first 2.8 years. We submit this information in the `aprob()` and `atime()` options, as shown below, and obtain the same estimate of 516 for sample size.

```
. power exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> aprob(0.3) atime(2.8)
note: input parameters are hazard rates
Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1
Study parameters:
      alpha = 0.0500
      power = 0.9000
      delta = -0.1000 (hazard difference)
Accrual and follow-up information:
      duration = 5.0000
      follow-up = 2.0000
      accrual = 3.0000 (exponential)
      accrual(%) = 30.00 (by time t*)
      t* = 2.8000 (93.33% of accrual)
Survival information:
      h1 = 0.3000
      h2 = 0.2000
Estimated sample sizes:
      N = 516
      N per group = 258
```

Another way we can supply the information about accrual is by specifying a percentage of subjects expected to be recruited by a certain percentage of the accrual period. For example, and equivalent to the above specification, 30% of subjects are expected to be recruited after 93.33% of the accrual period has elapsed. We submit this information in the `aprob()` and `aptime()` options, and we again obtain the same estimate of 516 for sample size.

```
. power exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> aprob(0.3) aptime(0.9333)
note: input parameters are hazard rates
Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1
Study parameters:
      alpha = 0.0500
      power = 0.9000
      delta = -0.1000 (hazard difference)
Accrual and follow-up information:
      duration = 5.0000
      follow-up = 2.0000
      accrual = 3.0000 (exponential)
      accrual(%) = 30.00 (by time t*)
      t* = 2.7999 (93.33% of accrual)
Survival information:
      h1 = 0.3000
      h2 = 0.2000
Estimated sample sizes:
      N = 516
      N per group = 258
```

## Exponential losses to follow-up

Apart from administrative censoring, subjects may not experience an event by the end of the study because of being lost to follow-up for various reasons. See [Survival data](#) in [PSS] [intro](#) and [PSS] [Glossary](#) for a more detailed description. [Rubinstein, Gail, and Santner \(1981\)](#) and [Lachin and Foulkes \(1986\)](#) extend sample-size and power computations to take into account exponentially distributed losses to follow-up. In addition to being exponentially distributed, losses to follow-up are assumed to be independent of the survival times.

### ► Example 8: Exponential losses to follow-up

Suppose that in [example 6](#), in the study with a 3-year uniform accrual and a 2-year follow-up, yearly loss hazards in the control and the experimental groups are 0.2. A loss hazard rate common to both groups can be specified in option `losshaz()`.

```
. power exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> lossshaz(0.2) show
note: input parameters are hazard rates

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2 = h1 versus Ha: h2 < h1

Study parameters:
      alpha =    0.0500
      power =    0.9000
      delta =   -0.1000 (hazard difference)

Accrual and follow-up information:
      duration =    5.0000
      follow-up =    2.0000
      accrual =    3.0000 (uniform)

Survival information:
      h1 =    0.3000
      h2 =    0.2000

Loss-to-follow-up information:
      lh1 =    0.2000
      lh2 =    0.2000

Estimated expected number of events:
      E|Ha =    213      E|Ho =    216
      E1|Ha =    121      E1|Ho =    108
      E2|Ha =    92       E2|Ho =    108

Estimated expected number of losses to follow-up:
      L|Ha =    173      L|Ho =    172
      L1|Ha =    81      L1|Ho =    86
      L2|Ha =    92      L2|Ho =    86

Estimated sample sizes:
      N =    500
      N per group =    250
```

The sample size required for a one-sided, 5%-level test to detect a reduction in hazard from 0.3 to 0.2 with 90% power increases from 378 (see [example 6](#)) to 500. We observe that for the extreme case of losses to follow-up, sample size increases significantly. A conservative adjustment commonly applied in practice is  $n(1 + p_L)$ , where  $p_L$  is the expected proportion of losses to follow-up in both groups combined. For this example,  $p_L$  may be computed as  $0.5(0.369 + 0.324) \approx 0.35$  from table 2 of [Lachin and Foulkes \(1986\)](#). Then the conservative estimate of the sample size is  $378(1 + 0.35) = 510$ , which is slightly greater than 500, the actual required sample size.

We also requested that additional information about the expected number of events and losses to follow-up under the null and under the alternative hypothesis be displayed by using the `show` option. From the above output, a total of 173 subjects (81 from the control group and 92 from the experimental group) are expected to be lost in the study with exponentially distributed losses with yearly rates of 0.2 in each group under the alternative hypothesis.

If the proportion of subjects lost to follow-up by a fixed period in each group is available, it can be supplied by using the `lossprob()` and `losstime()` options rather than loss to follow-up rates. For example, in the above study approximately 33%,  $1 - \exp(-0.2 \times 2) \approx 0.33$ , of subjects in each group are lost at time 2 (years). We can obtain the same estimates of sample sizes by typing

```
. power exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> lossprob(0.33) losstime(2)
(output omitted)
```

◄

## The conditional versus unconditional approaches

Denote  $\delta$  to be the effect size, and denote  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$  to be the maximum likelihood estimates of the respective hazard-rate parameters. Consider the two effect-size estimators based on the difference between the hazard rates,  $\hat{\lambda}_2 - \hat{\lambda}_1$ , and based on the log of the hazard ratio,  $\ln(\hat{\lambda}_2/\hat{\lambda}_1)$ . Both estimators are asymptotically normal under the null and under the alternative hypothesis.

We adopt [Chow, Shao, and Wang \(2008, 173\)](#) terminology when referring to the conditional and unconditional tests. The conditional test is the test that uses the constraint  $\lambda_2 = \lambda_1$  (conditional on  $H_0$ ) when computing the variance of the effect-size estimator under the null. The unconditional test is the test that does not use the above constraint when computing the variance of the effect-size estimator under the null. The score and the Wald tests are each one of the examples of conditional and unconditional tests, respectively. [Chow, Shao, and Wang \(2008\)](#) note that neither of the two tests (conditional or unconditional) is always more powerful than the other under the alternative hypothesis. Therefore, there is no definite recommendation of which one is preferable in practice.

The conditional approach relies on the following relationship between sample size and power (given in [Lachin 1981](#)) to compute estimates of required sample size or power,

$$|\delta| = z_{1-\alpha} \{\text{Var}(\delta, H_0)\}^{1/2} + z_{1-\beta} \{\text{Var}(\delta, H_a)\}^{1/2}$$

where  $z_{1-\alpha}$  and  $z_{1-\beta}$  are the  $(1-\alpha)$ th and the  $(1-\beta)$ th quantiles of the standard normal distribution, and  $\text{Var}(\delta, H_0)$  and  $\text{Var}(\delta, H_a)$  are the asymptotic variances under the null and under the alternative, respectively, of the effect-size estimator,  $\hat{\delta}$ . This approach uses the variance of the estimator conditional on the hypothesis type.

The unconditional approach replaces  $\text{Var}(\delta, H_0)$  with  $\text{Var}(\delta, H_a)$  in the above and uses the variance under the alternative to compute the estimates of sample size and power:

$$|\delta| = (z_{1-\alpha} + z_{1-\beta}) \{\text{Var}(\delta, H_a)\}^{1/2}$$

Therefore, the resulting formulas based on the two approaches are different.

[Lakatos and Lan \(1992\)](#) formulate the sample-size formula for the log hazard-ratio test based on the method of [Rubinstein, Gail, and Santner \(1981\)](#). This formula is based on the unconditional approach. [Lachin and Foulkes \(1986\)](#) provide the sample-size formula for the test of the log of the hazard ratio that uses the conditional approach. They also present both conditional and unconditional

versions of formulas for the test based on the difference between hazards. As noted by [Lachin and Foulkes \(1986\)](#), sample sizes estimated based on the unconditional approach will be larger than the estimates based on the conditional approach for equal-sized groups.

Both approaches are available with `power exponential`; the conditional is the default and the unconditional may be requested by specifying the `unconditional` option. Refer to [Methods and formulas](#) for the formulas underlying these approaches.

### ► Example 9: Sample size using the Rubinstein–Gail–Santner method

Consider the following scenario in [Lakatos and Lan \(1992, table I\)](#). A 10-year survival study with a 1-year accrual period and a 9-year follow-up is conducted to compare the survivor functions of the two groups by using a two-sided, 0.05 exponential test based on the log of the hazard ratio. The probability of surviving to the end of a study for subjects in the control group is 0.8 [ $S_1(t) = 0.8, t = 10$ ]. Subjects are recruited uniformly over the interval  $[0, 1]$ . [Lakatos and Lan \(1992\)](#) report an estimate of 664 for the sample size required to detect a change in the hazard of the experimental group corresponding to the hazard ratio  $\Delta = 0.5$  with 90% power by using the Rubinstein–Gail–Santner (1981) method. To obtain the estimates according to this method, we need to specify both `loghazard` and `unconditional`.

```
. power exponential 0.8, t(10) power(0.9) aperiod(1) fperiod(9) loghazard
> unconditional
note: input parameters are survival probabilities
Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log hazard-ratio, unconditional
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
      alpha = 0.0500
      power = 0.9000
      delta = -0.6931 (log hazard-ratio)
Accrual and follow-up information:
      duration = 10.0000
      follow-up = 9.0000
      accrual = 1.0000 (uniform)
Survival information:
      h1 = 0.0223          s1 = 0.8000
      h2 = 0.0112          s2 = 0.8944
      hratio = 0.5000      t = 10.0000
Estimated sample sizes:
      N = 664
      N per group = 332
```

Because the default value of the hazard ratio is 0.5, we omit the `hratio(0.5)` option in the above. From the output, we obtain the same estimate of 664 of the sample size as reported in [Lakatos and Lan \(1992\)](#).

◀

In the absence of censoring, the estimates of the sample size or power based on the test of log of the hazard ratio are the same for the conditional and the unconditional approaches. For example, both

```
. power exponential 0.8, t(10) power(0.9) loghazard
(output omitted)
```

and

```
. power exponential 0.8, t(10) power(0.9) loghazard unconditional
(output omitted)
```



produce the same estimate of the sample size (88). The asymptotic variance of maximum likelihood estimates of the log of the hazard ratio does not depend on hazard rates when there is no censoring and, therefore, does not depend on the type of hypothesis,  $\text{Var}(\widehat{\delta}, H_0) = \text{Var}(\widehat{\delta}, H_a) = 2/N$ .

## Link to the sample-size and power computation for the log-rank test

### ► Example 10: Sample size using the Freedman and the Schoenfeld methods

Continuing with examples 1 and 2, Lachin (1981, 106) gives another approximation to obtain the estimate of the sample size under the equal-group allocation. This approximation coincides with the formula derived by Freedman (1982) for the number of events in the context of the log-rank test. We can obtain such an estimate by using power logrank and by specifying the hazard ratio of 0.66667 computed earlier.

```
. power logrank, hratio(0.66667) power(0.9) onesided
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
      alpha = 0.0500
      power = 0.9000
      delta = 0.6667 (hazard ratio)
      hratio = 0.6667
Censoring:
      Pr_E = 1.0000
Estimated number of events and sample sizes:
      E = 216
      N = 216
      N per group = 108
```

The estimate, 216, of the sample size is the same as given in Lachin (1981, 107) and is slightly smaller than the estimate, 218, obtained in example 1 and larger than the estimate, 210, obtained using the George–Desu method in example 2.

The approximation due to George and Desu (1974) is the same as the approximation to the number of events derived by Schoenfeld (1981) in application to the log-rank test. We can confirm that by typing

```
. power logrank, hratio(0.66667) power(0.9) onesided schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) < 0
Study parameters:
      alpha = 0.0500
      power = 0.9000
      delta = -0.4055 (log hazard-ratio)
      hratio = 0.6667
Censoring:
      Pr_E = 1.0000
Estimated number of events and sample sizes:
      E = 210
      N = 210
      N per group = 105
```

We obtain the same estimate of 210 as when using `power exponential` with the `loghazard` option in [example 2](#).



## Computing power

Sometimes the number of subjects available for enrollment into the study is limited. In such cases, the researchers may want to investigate with what power they can detect a desired treatment effect for a given sample size.

To compute power, you must specify the sample size in the `n()` option and an effect size. A hazard ratio of 0.5 is assumed if an effect size is not specified. Also see [Alternative ways of specifying effect](#) for various ways of specifying an effect size.

### ► Example 11: Power determination

We verify the power computation for the study from [example 9](#). We expect the power estimate to be close to 0.9.

The only thing we change in the `power exponential` command from [example 9](#) is replacing the `power(0.9)` option with the `n(664)` option.

```
. power exponential 0.8, t(10) n(664) aperiod(1) fperiod(9)
> loghazard unconditional
note: input parameters are survival probabilities

Estimated power for two-sample comparison of survivor functions
Exponential test, log hazard-ratio, unconditional
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0

Study parameters:
      alpha =    0.0500
         N =     664
  N per group =   332
      delta =  -0.6931 (log hazard-ratio)

Accrual and follow-up information:
      duration = 10.0000
      follow-up =  9.0000
      accrual =  1.0000 (uniform)

Survival information:
      h1 =    0.0223          s1 =    0.8000
      h2 =    0.0112          s2 =    0.8944
      hratio =  0.5000          t =   10.0000

Estimated power:
      power =    0.9000
```

We obtain the estimate of power 0.9.



## Testing hypotheses about two exponential survivor functions

### ► Example 12: Using streg to perform the log hazard-ratio test

In this example, we demonstrate the importance of sample-size computations to ensure a high power of a test to detect a difference between exponential survivor functions. We consider an asymptotic Wald (or normal  $z$ ) test to test whether the log of the hazard ratio is zero.

Continuing with [example 11](#), suppose that the investigators have only 100 subjects available for the study. As we see below, the power to detect a 50% risk reduction in a hazard of the experimental group (the hazard ratio of 0.5) decreases from 90% to 24%:

```
. power exponential 0.8, t(10) n(100) aperiod(1) fperiod(9)
> loghazard unconditional
note: input parameters are survival probabilities

Estimated power for two-sample comparison of survivor functions
Exponential test, log hazard-ratio, unconditional
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0

Study parameters:
      alpha =    0.0500
        N =     100
  N per group =    50
      delta =   -0.6931 (log hazard-ratio)

Accrual and follow-up information:
      duration = 10.0000
      follow-up =  9.0000
      accrual =  1.0000 (uniform)

Survival information:
      h1 =    0.0223      s1 =    0.8000
      h2 =    0.0112      s2 =    0.8944
      hratio = 0.5000      t =   10.0000

Estimated power:
      power =    0.2414
```

To demonstrate the implication of this reduction, consider the following example. We generate the data according to the study from [example 11](#) with the following code:

```
program simdata
  args n h1 h2 r
  set obs `n'
  generate double entry = `r'*runiform()
  generate double u = runiform()
  /* random allocation to two groups of equal sizes */
  generate double u1 = runiform()
  generate double u2 = runiform()
  sort u1 u2, stable
  generate byte drug = (_n<=`n'/2)
  /* exponential failure times with rates h1 and h2 */
  generate double failtime = entry - ln(1-u)/`h1' if drug==0
  replace failtime = entry - ln(1-u)/`h2' if drug==1
end

. clear
. set seed 234
. quietly simdata 100 0.0223 0.0112 1
```

The entry times of subjects are generated from a uniform  $[0, 1)$  distribution and stored in variable `entry`. The subjects are randomized to two groups of equal size of 50 subjects each. The survival times are generated from exponential distribution with the hazard rate of  $-\ln(0.8)/10 = 0.0223$  in the control group, `drug = 0`, and the hazard rate of  $0.5 \times 0.0223 = 0.0112$  in the experimental group, `drug = 1`, conditional on subjects' entry times in `entry`.

Before analyzing these survival data, we need to set up the data properly using `stset`. The failure-time variable is `failtime`. The study terminates at  $t = 10$ , so we use `exit(time 10)` with `stset` to specify that all failure times past 10 are to be treated as censored. Because subjects enter the study at random times (`entry`) and become at risk of a failure upon entering the study, we also specify the `origin(entry)` option to ensure that the analysis time is adjusted for the entry times. For more details, see [ST] [stset](#).

```
. stset failtime, exit(time 10) origin(entry)
      failure event: (assumed to fail at time=failtime)
obs. time interval: (origin, failtime]
exit on or before: time 10
      t for analysis: (time-origin)
              origin: time entry
```

---

```
100 total observations
  0 exclusions
```

---

```
100 observations remaining, representing
  7 failures in single-record/single-failure data
921.825 total analysis time at risk and under observation
              at risk from t =          0
earliest observed entry t =          0
              last observed exit t = 9.990494
```

To perform the log hazard-ratio test, we fit an exponential regression model on `drug` by using `streg` (see [ST] [streg](#)). We can express the log of the hazard ratio in terms of regression coefficients as follows:  $\ln(\Delta) = \ln(\lambda_2/\lambda_1) = \ln\{\exp(\beta_0 + \beta_1)/\exp(\beta_0)\} = \beta_1$ , where  $\beta_0$  and  $\beta_1$  are the estimated coefficients for the constant and `drug` in the regression model. Then the test of  $H_0: \ln(\Delta) = 0$  may be rewritten in terms of a coefficient on `drug` as  $H_0: \beta_1 = 0$ . This test is part of the standard output after `streg`.

```
. streg drug, distribution(exponential) nohr
      failure _d: 1 (meaning all fail)
      analysis time _t: (failtime-origin)
              origin: time entry
      exit on or before: time 10
Iteration 0: log likelihood = -29.718762
Iteration 1: log likelihood = -29.049311
Iteration 2: log likelihood = -29.014323
Iteration 3: log likelihood = -29.014222
Iteration 4: log likelihood = -29.014222
Exponential regression -- log relative-hazard form
No. of subjects =          100          Number of obs =          100
No. of failures =           7
Time at risk = 921.824925
Log likelihood = -29.014222
LR chi2(1) = 1.41
Prob > chi2 = 0.2352
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	.9428118	.83666	1.13	0.260	-.6970117	2.582635
_cons	-5.453234	.7071068	-7.71	0.000	-6.839137	-4.06733

From the output table above, the  $p$ -value for a two-sided test of the coefficient for drug, 0.260, is greater than 0.05. On that basis, we do not have evidence to reject the null hypothesis of no difference between the two exponential survivor functions. Therefore, we make an incorrect decision because we simulated the data with different group hazard rates. If we were to repeat this, say, 100 times, using different datasets simulated according to the alternative  $H_a: \ln(\Delta) = \ln(0.5) = -0.6931$  (see [R] [simulate](#)), for roughly 76 of them we would have failed to reject the null hypothesis of no difference (a type II error). Therefore, more subjects are required to be able to detect the log of the hazard ratio of  $-0.4055$  in this study. ◀

### ▶ Example 13: Using results from `streg` to perform the Wald test of hazard difference

We obtain the power of the test based on the difference between hazards for the study in [example 12](#) (omit the `loghazard` option from the syntax of `power exponential`).

```
. power exponential 0.8, t(10) n(100) aperiod(1) fperiod(9) unconditional
note: input parameters are survival probabilities
Estimated power for two-sample comparison of survivor functions
Exponential test, hazard difference, unconditional
Ho: h2 = h1 versus Ha: h2 != h1
Study parameters:
      alpha =    0.0500
          N =     100
  N per group =    50
      delta =   -0.0112 (hazard difference)
Accrual and follow-up information:
      duration =   10.0000
      follow-up =    9.0000
      accrual =    1.0000 (uniform)
Survival information:
      h1 =    0.0223          s1 =    0.8000
      h2 =    0.0112          s2 =    0.8944
      hratio = 0.5000          t =   10.0000
Estimated power:
      power =    0.2458
```

We obtain a power estimate of 0.2458, which is close to 0.2414 from [example 12](#).

To test the difference between hazard rates by using the Wald test, we express this difference in terms of coefficients,  $\lambda_2 - \lambda_1 = \exp(\beta_0)\{\exp(\beta_1) - 1\}$ , and we use `testnl` ([R] [testnl](#)) after `streg` to perform the nonlinear hypothesis test of  $H_0: \exp(\beta_0)\{\exp(\beta_1) - 1\} = 0$ .

```
. testnl exp(_b[_cons])*(exp(_b[drug])-1) = 0
      (1) exp(_b[_cons])*(exp(_b[drug])-1) = 0
           chi2(1) =          1.35
           Prob > chi2 =        0.2451
```

We obtain the same conclusions from the Wald test based on the difference between hazards as in [example 12](#). That is, based on the  $p$ -value of 0.2451, we fail to reject the null hypothesis of no difference between hazards of two groups (or miss the alternative  $H_a: \psi = -0.0112$  corresponding to reduction in hazard from roughly 0.02 to 0.01) for the data from [example 12](#). ◀

Often in practice, to test the disparity in two exponential survivor functions, the log-rank test is used instead of the hazard-difference test. Also the Wald (or the score) test from the Cox model is used instead of the exponential log hazard-ratio test. Refer to [ST] [sts test](#) and [ST] [stcox](#) for examples on how to perform these tests (also see [PSS] [power logrank](#) and [PSS] [power cox](#)).

Sometimes the estimates of sample size and power obtained under the assumption of the exponential model are used as an approximation to the results used in a more general context of the log-rank test or the Cox proportional hazards model. Refer to [Lachin \(2011, 483–484\)](#) for the rationale behind this. Also see [Lakatos and Lan \(1992\)](#) for a discussion of the circumstances under which sample-size estimates obtained assuming the exponential model may be inaccurate when used with more general proportional hazards models.

## Stored results

`power exponential` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(hratio)</code>	hazard ratio
<code>r(lnhratio)</code>	log hazard-ratio
<code>r(hdiff)</code>	difference between hazard rates
<code>r(h1)</code>	hazard in the control group (if specified)
<code>r(h2)</code>	hazard in the experimental group
<code>r(s1)</code>	survival probability in the control group (if specified)
<code>r(s2)</code>	survival probability in the experimental group (if specified)
<code>r(time)</code>	reference survival time (if <code>time()</code> is specified)
<code>r(aperiod)</code>	length of the accrual period (if specified)
<code>r(fperiod)</code>	length of the follow-up period (if specified)
<code>r(studytime)</code>	duration of the study (if specified)
<code>r(ashape)</code>	shape parameter (if <code>aperiod()</code> is specified)
<code>r(aprob)</code>	shape parameter (if <code>aprob()</code> is specified)
<code>r(aptime)</code>	proportion of accrual period (if <code>aptime()</code> is specified)
<code>r(ptime)</code>	reference accrual time (if <code>ptime()</code> is specified)
<code>r(losshaz)</code>	loss hazard rate in both groups (if specified)
<code>r(losshaz1)</code>	loss hazard in the control group (if specified)
<code>r(losshaz2)</code>	loss hazard in the experimental group (if specified)
<code>r(lossprob)</code>	proportions of subjects lost to follow-up in both groups (if <code>lossprob()</code> is specified)
<code>r(losstime)</code>	reference loss to follow-up time (if <code>losstime()</code> is specified)
<code>r(unconditional)</code>	1 if <code>unconditional</code> is specified; 0 otherwise
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise

### Macros

<code>r(type)</code>	test
<code>r(method)</code>	<code>exponential</code>
<code>r(test)</code>	hazard difference or log-hazard difference
<code>r(accrual)</code>	uniform or exponential
<code>r(effect)</code>	<code>hratio</code> , <code>lnhratio</code> , <code>hdifference</code> , or <code>lnhdifference</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

Matrix

<code>r(pss_table)</code>	table of results
<code>r(Pr_vec)</code>	1×4 matrix of probabilities of an event (when computed)
<code>r(Ea_vec)</code>	1×3 matrix of expected number of events under the alternative (when computed)
<code>r(E0_vec)</code>	1×3 matrix of expected number of events under the null (when computed)
<code>r(La_vec)</code>	1×3 matrix of expected number of losses under the alternative (when computed)
<code>r(L0_vec)</code>	1×3 matrix of expected number of losses under the null (when computed)

## Methods and formulas

By default, `power exponential` computes the sample size required to achieve a specified power to detect a difference between hazard rates,  $\psi_a = \lambda_{2a} - \lambda_{1a}$ , using the method of [Lachin \(1981\)](#). If `loghazard` is specified, the sample size required to detect a log of the hazard ratio  $\ln(\Delta_a) = \ln(\lambda_{2a}/\lambda_{1a})$  with specified power is reported using the formula derived by [George and Desu \(1974\)](#). In the presence of an accrual period, the methods of [Lachin and Foulkes \(1986\)](#) or (for uniform accrual only) [Rubinstein, Gail, and Santner \(1981\)](#) (if `loghazard` and `unconditional` are specified) are used.

In addition to the notation given in [Introduction](#), denote  $n$ ,  $n_1$ , and  $n_2$  to be the total number of subjects required for the study, the number of subjects in the control group, and the number of subjects in the experimental group, respectively. Let  $R = n_2/n_1$  denote the ratio of sample sizes of the experimental group to the control group. Let  $p_1 = n_1/n = 1/(1 + R)$  and  $p_2 = n_2/n = 1 - p_1 = R/(1 + R)$  be the proportions of subjects allocated to the control and the experimental groups;  $\gamma$  be the shape parameter of the truncated exponential distribution with p.d.f.  $g(z) = \gamma \exp(-\gamma z) / \{1 - \exp(-\gamma r)\}$ ,  $0 \leq z \leq r$ ,  $\gamma \neq 0$ ;  $\eta_1$  and  $\eta_2$  be the loss hazards in the control and the experimental groups; and  $z_{(1-\alpha/k)}$  and  $z_{(1-\beta)}$  be the  $(1 - \alpha/k)$ th and the  $(1 - \beta)$ th quantiles of the standard normal distribution, with  $k = 1$  for the one-sided test and  $k = 2$  for the two-sided test. Denote  $\bar{\lambda} = p_1 \lambda_1 + p_2 \lambda_2$ . Recall that the difference between hazards is denoted by  $\psi = \lambda_2 - \lambda_1$  and the hazard ratio is denoted by  $\Delta = \lambda_2/\lambda_1$ .

If survival probabilities  $S_1(t)$  and  $S_2(t)$  at a fixed time  $t$  are specified rather than hazard rates, the hazard rates are computed as  $\lambda_i = -\ln\{S_i(t)\}/t$ ,  $i = 1, 2$ . If loss to follow-up probabilities  $L_1(t_L)$  and  $L_2(t_L)$  at a fixed time  $t_L$  are given instead of loss to follow-up hazard rates, the loss hazard rates are computed as  $\eta_i = -\ln\{1 - L_i(t_L)\}/t_L$ ,  $i = 1, 2$ .

All formulas below are derived under the assumption of exponential survival distributions with hazard rates in the control and the experimental groups  $\lambda_1$  and  $\lambda_2$ , respectively, and rely on large-sample properties of the maximum likelihood estimates of  $\lambda_1$  and  $\lambda_2$ .

Denote  $\xi_o = \zeta(\bar{\lambda}, \gamma, \eta_1)p_1^{-1} + \zeta(\bar{\lambda}, \gamma, \eta_2)p_2^{-1}$  and  $\xi_a = \zeta(\lambda_1, \gamma, \eta_1)p_1^{-1} + \zeta(\lambda_2, \gamma, \eta_2)p_2^{-1}$ .

The formula for the sample-size calculation using the conditional approach is

$$n = \frac{\left(z_{1-\alpha/k}\xi_o^{1/2} + z_{1-\beta}\xi_a^{1/2}\right)^2}{\delta^2}$$

and using the unconditional approach is

$$n = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2 \xi_a}{\delta^2}$$

where  $\zeta(\lambda, \gamma, \eta) = \lambda^2/p_E$  if  $\delta = \psi$ ,  $\zeta(\lambda, \gamma, \eta) = 1/p_E$  if  $\delta = \ln(\Delta)$ , and  $p_E$  is to be defined later.  $\lambda$  and  $\eta$  denote a failure hazard rate and a loss to follow-up hazard rate.

In the absence of censoring, the overall probability of an event (failure),  $p_E$ , is set to 1. Here the resulting formula for the sample size for the log hazard-ratio test depends only on the ratio of hazards and not on the individual group hazard rates. The resulting sample size formula for the test of the difference may also be rewritten as a function of the ratio of hazards only. Therefore, under no censoring, for a fixed value of the hazard ratio  $\Delta = \lambda_2/\lambda_1$ , the estimates of the sample size (or power) will be constant with respect to varying hazard rates  $\lambda_1$  and  $\lambda_2$ .

In the presence of censoring, when each subject is followed up for a fixed period  $f = T$ ,

$$p_E = p_E(\lambda, \eta) = \frac{\lambda}{\lambda + \eta} [1 - \exp\{-(\lambda + \eta)T\}]$$

In the presence of an accrual period, the probability of an event is defined as

$$p_E = p_E(\lambda, \eta) = \frac{\lambda}{(\lambda + \eta)} \left[ 1 - \frac{\exp\{-(\lambda + \eta)(T - r)\} - \exp\{-(\lambda + \eta)T\}}{(\lambda + \eta)r} \right]$$

or

$$p_E = p_E(\lambda, \gamma, \eta) = \frac{\lambda}{(\lambda + \eta)} \left( 1 + \frac{\gamma \exp\{-(\lambda + \eta)T\} [1 - \exp\{(\lambda + \eta - \gamma)r\}]}{(\lambda + \eta - \gamma) \{1 - \exp(-\gamma r)\}} \right)$$

under uniform or truncated exponential accrual with shape  $\gamma$  over  $[0, r]$ , respectively. Uniform accrual is assumed for  $|\gamma| < 10^{-6}$ .

The formulas are obtained from Lachin (1981), Lachin and Foulkes (1986), and Lakatos and Lan (1992). To avoid division by 0 in the case  $\lambda + \eta = \gamma$ , the probability of an event is taken to be the limit of the above expression,  $p_E = \lim_{\lambda + \eta \rightarrow \gamma} p_E(\lambda, \gamma, \eta)$ .

The number of subjects required to be recruited in each group is obtained as  $n_1 = n/(1 + R)$  and  $n_2 = nR/(1 + R)$ . If `nfractional` is not specified, sample sizes are rounded to integer values; see *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

The expected number of events and losses to follow-up are computed as suggested by Lachin and Foulkes (1986). Under the null hypothesis,

$$\begin{aligned} E_{H_0} &= n_1 p_E(\bar{\lambda}, \gamma, \eta_1) + n_2 p_E(\bar{\lambda}, \gamma, \eta_2) \\ L_{H_0} &= n_1 (\eta_1 / \bar{\lambda}) p_E(\bar{\lambda}, \gamma, \eta_1) + n_2 (\eta_2 / \bar{\lambda}) p_E(\bar{\lambda}, \gamma, \eta_2) \end{aligned}$$

and under the alternative hypothesis,

$$\begin{aligned} E_{H_a} &= n_1 p_E(\lambda_1, \gamma, \eta_1) + n_2 p_E(\lambda_2, \gamma, \eta_2) \\ L_{H_a} &= n_1 (\eta_1 / \lambda_1) p_E(\lambda_1, \gamma, \eta_1) + n_2 (\eta_2 / \lambda_2) p_E(\lambda_2, \gamma, \eta_2) \end{aligned}$$

For unconditional tests, the expected number of events and losses to follow-up under the null is computed by setting  $\bar{\lambda} = \lambda_1$ . The estimates of the expected number of events and losses to follow-up in each group are rounded to the nearest integer.

To obtain the estimate of the power,  $1 - \beta$ , the formulas for the sample size are solved for  $z_{(1-\beta)}$  and the normal cumulative distribution function is used to obtain the corresponding probability  $1 - \beta$ .



To obtain the unknown shape parameter,  $\gamma$ , of a truncated exponential entry distribution, an iterative procedure is used to solve the equation

$$p_a = G(t_a) = \frac{1 - \exp(-\gamma t_a)}{1 - \exp(-\gamma r)}$$

for a given proportion of subjects  $p_a$  recruited at a given time,  $t_a$ , for  $t_a \in [0, r]$ .

## References

- Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.
- Cleves, M. A., W. W. Gould, R. G. Gutierrez, and Y. V. Marchenko. 2010. *An Introduction to Survival Analysis Using Stata*. 3rd ed. College Station, TX: Stata Press.
- Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.
- George, S. L., and M. M. Desu. 1974. Planning the size and duration of a clinical trial studying the time to some critical event. *Journal of Chronic Diseases* 27: 15–24.
- Lachin, J. M. 1981. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 2: 93–113.
- . 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.
- Lachin, J. M., and M. A. Foulkes. 1986. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 42: 507–519.
- Lakatos, E., and K. K. G. Lan. 1992. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine* 11: 179–191.
- Rubinstein, L. V., M. H. Gail, and T. J. Santner. 1981. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *Journal of Chronic Diseases* 34: 469–479.
- Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68: 316–319.

Also see [\[PSS\] intro](#) for more references.

## Also see

- [\[PSS\] power](#) — Power and sample-size analysis for hypothesis tests
- [\[PSS\] power cox](#) — Power analysis for the Cox proportional hazards model
- [\[PSS\] power logrank](#) — Power analysis for the log-rank test
- [\[PSS\] power, graph](#) — Graph results from the power command
- [\[PSS\] power, table](#) — Produce table of results from the power command
- [\[PSS\] Glossary](#)
- [\[R\] test](#) — Test linear hypotheses after estimation
- [\[ST\] streg](#) — Parametric survival models

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

## Description

`power logrank` computes sample size, power, or effect size for survival analysis comparing survivor functions in two groups by using the log-rank test. The results can be obtained using the Freedman or Schoenfeld approaches. Effect size can be expressed as a hazard ratio or as a log hazard-ratio. The command supports unbalanced designs, and provides options to account for administrative censoring, uniform accrual, and withdrawal of subjects from the study.

## Quick start

Sample size for a log-rank test of  $H_0: \Delta = 0$  versus  $H_a: \Delta \neq 0$  using the Freedman method for alternative hazard ratio  $\Delta_a = 0.76$  without censoring and with default power of 0.8 and significance level  $\alpha = 0.05$

```
power logrank, hratio(.76)
```

As above, but use Schoenfeld's method

```
power logrank, hratio(.76) schoenfeld
```

Sample size for censored design with survival probabilities  $surv_1 = 0.3$  and  $surv_2 = 0.4$

```
power logrank .3 .4
```

Same as above, specified as  $surv_1 = 0.3$  and hazard ratio of 0.76

```
power logrank .3, hratio(.76)
```

As above, but for hazard ratios of 0.65, 0.7, 0.75, and 0.8

```
power logrank .3, hratio(.65(.05).8)
```

As above, but show results in a graph of hazard ratio versus sample size

```
power logrank .3, hratio(.65(.05).8) graph
```

Sample size for a one-sided test with  $\alpha = 0.01$

```
power logrank .3, hratio(.76) onesided alpha(.01)
```

Sample size adjusted for 10% withdrawal from the study

```
power logrank .3, hratio(.76) wdprob(.1)
```

Power for a design with censoring and a sample size of 300

```
power logrank .3 .4, n(300)
```

As above, but specify twice as many observations in the experimental group

```
power logrank .3 .4, n(300) nratio(2)
```

Effect size for a design without censoring, sample size of 300, power of 0.8, and default  $\alpha = 0.05$

```
power logrank, n(300) power(.8)
```

As above, but for a censored design with control-group survival probability of 0.3

```
power logrank .3, n(300) power(.8)
```

## Menu

Statistics > Power and sample size

## Syntax

*Compute sample size*

```
power logrank [surv1 [surv2]] [, power(numlist) options]
```

*Compute power*

```
power logrank [surv1 [surv2]], n(numlist) [options]
```

*Compute effect size*

```
power logrank [surv1], n(numlist) power(numlist) [options]
```

where

*surv*<sub>1</sub> is the survival probability in the control (reference) group at the end of the study  $t^*$ , and *surv*<sub>2</sub> is the survival probability in the experimental (comparison) group at the end of the study  $t^*$ . *surv*<sub>1</sub> and *surv*<sub>2</sub> may each be specified either as one number or as a list of values in parentheses; see [U] 11.1.8 [numlist](#).

<i>options</i>	Description
Main	
* <u>alpha</u> ( <i>numlist</i> )	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> ( <i>numlist</i> )	power; default is <code>power(0.8)</code>
* <u>beta</u> ( <i>numlist</i> )	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
* <u>n1</u> ( <i>numlist</i> )	sample size of the control group
* <u>n2</u> ( <i>numlist</i> )	sample size of the experimental group
* <u>nratio</u> ( <i>numlist</i> )	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<u>nfractional</u>	allow fractional sample sizes
* <u>hratio</u> ( <i>numlist</i> )	hazard ratio (effect size) of the experimental to the control group; default is <code>hratio(0.5)</code> ; may not be combined with <code>lnhratio()</code>
* <u>lnhratio</u> ( <i>numlist</i> )	log hazard-ratio (effect size) of the experimental to the control group; may not be combined with <code>hratio()</code>
<u>schoenfeld</u>	use the formula based on the log hazard-ratio in calculations; default is to use the formula based on the hazard ratio
<u>effect</u> ( <i>effect</i> )	specify the type of effect to display; default is method-specific
<u>direction</u> ( <u>lower</u>   <u>upper</u> )	direction of the effect for effect-size determination; default is <code>direction(lower)</code> , which means that the postulated value of the parameter is smaller than the hypothesized value
<u>onesided</u>	one-sided test; default is two sided
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Censoring	
<u>simpson</u> (# # #   <i>matname</i> )	survival probabilities in the control group at three specific time points to compute the probability of an event (failure), using Simpson's rule under uniform accrual
<u>st1</u> ( <i>varname<sub>s</sub></i> <i>varname<sub>t</sub></i> )	variables <i>varname<sub>s</sub></i> , containing survival probabilities in the control group, and <i>varname<sub>t</sub></i> , containing respective time points, to compute the probability of an event (failure), using numerical integration under uniform accrual
* <u>wdprob</u> ( <i>numlist</i> )	proportion of subjects anticipated to withdraw from the study; default is <code>wdprob(0)</code>
Table	
[ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]	suppress table or display results as a table; see [PSS] <b>power, table</b>
<u>saving</u> ( <i>filename</i> [, <u>replace</u> ])	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [ ( <i>graphopts</i> ) ]	graph results; see [PSS] <b>power, graph</b>

iteration	
<code>init(#)</code>	initial value for effect size
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	suppress the title

\*Starred options may be specified either as one number or as a list of values; see [U] 11.1.8 [numlist](#).

`notitle` does not appear in the dialog box.

<i>effect</i>	Description
<code>hratio</code>	hazard ratio
<code>lnhratio</code>	log hazard-ratio

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	$\alpha$
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	$\beta$
<code>N</code>	total number of subjects	$N$
<code>N1</code>	number of subjects in the control group	$N_1$
<code>N2</code>	number of subjects in the experimental group	$N_2$
<code>nratio</code>	ratio of sample sizes, experimental to control	$N_2/N_1$
<code>delta</code>	effect size	$\delta$
<code>E</code>	total number of events (failures)	$E$
<code>hratio</code>	hazard ratio	$\Delta$
<code>lnhratio</code>	log hazard-ratio	$\ln(\Delta)$
<code>s1</code>	survival probability in the control group	$S_1(T)$
<code>s2</code>	survival probability in the experimental group	$S_2(T)$
<code>Pr_E</code>	overall probability of an event (failure)	$p_E$
<code>Pr_w</code>	probability of withdrawals	$p_w$
<code>target</code>	target parameter; <code>hratio</code> or <code>lnhratio</code>	
<code>_all</code>	display all supported columns	

Column `beta` is shown in the default table in place of column `power` if option `beta()` is specified.

Column `lnhratio` is shown in the default table in place of column `hratio` if specified.

Columns `s1` and `s2` are available only when specified.

Columns `nratio` and `Pr_w` are shown in the default table if specified.

## Options

## Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `nfractional`; see [PSS] [power](#).

`hratio(numlist)` specifies the hazard ratio (effect size) of the experimental group to the control group.

The default is `hratio(0.5)`. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by the log-rank test with a certain power.

You can specify an effect size either as a hazard ratio in `hratio()` or as a log hazard-ratio in `lnhratio()`. The default is `hratio(0.5)`. If both arguments `surv1` and `surv2` are specified, `hratio()` is not allowed and the hazard ratio is instead computed as  $\ln(\text{surv}_2)/\ln(\text{surv}_1)$ .

This option is not allowed with the effect-size determination and may not be combined with `lnhratio()`.

`lnhratio(numlist)` specifies the log hazard-ratio (effect size) of the experimental group to the control group. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by the log-rank test with a certain power.

You can specify an effect size either as a hazard ratio in `hratio()` or as a log hazard-ratio in `lnhratio()`. The default is `hratio(0.5)`. If both arguments `surv1` and `surv2` are specified, `lnhratio()` is not allowed and the log hazard-ratio is computed as  $\ln\{\ln(\text{surv}_2)/\ln(\text{surv}_1)\}$ .

This option is not allowed with the effect-size determination and may not be combined with `hratio()`.

`schoenfeld` requests calculations using the formula based on the log hazard-ratio, according to [Schoenfeld \(1981\)](#). The default is to use the formula based on the hazard ratio, according to [Freedman \(1982\)](#).

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. *effect* is one of `hratio` or `lnhratio`. By default, the effect size `delta` is a hazard ratio, `effect(hratio)`, for a hazard-ratio test and a log hazard-ratio, `effect(lnhratio)`, for a log hazard-ratio test (when `schoenfeld` is specified).

`direction()`, `onesided`, `parallel`; see [PSS] [power](#). `direction(lower)` is the default.

## Censoring

`simpson(# # # | matname)` specifies survival probabilities in the control group at three specific time points to compute the probability of an event (failure) using Simpson's rule under the assumption of uniform accrual. Either the actual values or a  $1 \times 3$  matrix, *matname*, containing these values can be specified. By default, the probability of an event is approximated as an average of the failure probabilities  $1-s_1$  and  $1-s_2$ ; see [Methods and formulas](#). `simpson()` may not be combined with `st1()` and may not be used if command argument `surv1` or `surv2` is specified. This option is not allowed with effect-size computation.

`st1(varnames, varnamet)` specifies variables *varname<sub>s</sub>*, containing survival probabilities in the control group, and *varname<sub>t</sub>*, containing respective time points, to compute the probability of an event (failure) using numerical integration under the assumption of uniform accrual; see [R] [dydx](#). The minimum and the maximum values of *varname<sub>t</sub>* must be the length of the follow-up period and the duration of the study, respectively. By default, the probability of an event is approximated as an average of the failure probabilities  $1-s_1$  and  $1-s_2$ ; see [Methods and formulas](#). `st1()` may not be combined with `simpson()` and may not be used if command argument `surv1` or `surv2` is specified. This option is not allowed with effect-size computation.

`wdprob(numlist)` specifies the proportion of subjects anticipated to withdraw from the study. The default is `wdprob(0)`. `wdprob()` is allowed only with sample-size computation.

#### Table

`table`, `table()`, `notable`; see [PSS] [power](#), [table](#).

`saving()`; see [PSS] [power](#).

#### Graph

`graph`, `graph()`; see [PSS] [power](#), [graph](#). Also see the *column* table for a list of symbols used by the graphs.

#### Iteration

`init(#)` specifies an initial value for the estimated hazard ratio or, if `schoenfeld` is specified, for the estimated log hazard-ratio during the effect-size determination.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS] [power](#).

The following option is available with `power logrank` but is not shown in the dialog box:

`notitle`; see [PSS] [power](#).

## Remarks and examples

Remarks are presented under the following headings:

*Introduction*

*Using power logrank*

*Computing sample size*

*Computing sample size in the absence of censoring*

*Computing sample size in the presence of censoring*

*Withdrawal of subjects from the study*

*Including information about subject accrual*

*Computing power*

*Computing effect size*

*Testing a hypothesis about two survivor functions using a log-rank test*

This entry describes the `power logrank` command and the methodology for power and sample-size analysis for a two-sample comparison of survivor functions using a log-rank test. See [PSS] [intro](#) for a general introduction to power and sample-size analysis and [PSS] [power](#) for a general introduction to the `power` command using hypothesis tests. See *Survival data* in [PSS] [intro](#) for an introduction to power and sample-size analysis for survival data.

## Introduction

Consider a survival study comparing the survivor functions in two groups using the log-rank test. Let  $S_1(t)$  and  $S_2(t)$  denote the survivor functions of the control and the experimental groups, respectively. The key assumption of the log-rank test is that the hazard functions are proportional. That is,  $h_2(t) = \Delta h_1(t)$  for any  $t$  or, equivalently,  $S_2(t) = \{S_1(t)\}^\Delta$ , where  $\Delta$  is the hazard ratio. If  $\Delta < 1$ , the survival in the experimental group is higher relative to the survival in the control group; the new treatment is superior to the standard treatment. If  $\Delta > 1$ , then the standard treatment is superior to the new treatment. Under the proportional-hazards assumption, the test of the equality of the two survivor functions  $H_0: S_1(t) = S_2(t)$  versus  $H_a: S_1(t) \neq S_2(t)$  is equivalent to the test  $H_0: \Delta = 1$  versus  $H_a: \Delta \neq 1$  or  $H_0: \ln(\Delta) = 0$  versus  $H_a: \ln(\Delta) \neq 0$ .

The methods implemented in `power logrank` for power and sample-size analysis relate the power of the log-rank test directly to the number of events observed in the study. Depending on whether censoring occurs in a study, the required number of subjects is either equal to the number of events or is computed using the estimates of the number of events and the combined probability of an event (failure). Thus, in the presence of censoring, in addition to the number of events, the probability of a subject not being censored (failing) needs to be estimated to obtain the final estimate of the required number of subjects in the study.

To determine the required number of events, the investigator must specify the size or significance level,  $\alpha$ , and the clinically significant difference between the two treatments (effect size) to be detected by the log-rank test,  $H_a: \Delta = \Delta_a$ , with prespecified power  $1 - \beta$ . The effect size, a difference between the two treatments, is usually expressed as a hazard ratio,  $\Delta_a$ , using the `hratio()` option. Alternatively, you may specify an effect size as a log hazard-ratio,  $\ln(\Delta_a)$ , in the `lnhratio()` option.

When all subjects fail by the end of the study (no censoring), a **type I study**, the information above is sufficient to obtain the number of subjects required in the study. Often, in practice, not all subjects fail by the end of the study, in which case censoring of subjects occurs (a **type II study**). Here the estimates of the survival probabilities in the control and experimental groups are necessary to estimate an overall probability of an event and, then, the required sample size.

`power logrank` supports two methods, those of [Freedman \(1982\)](#) and [Schoenfeld \(1981\)](#), to obtain the estimates of the number of events or power (see also [Marubini and Valsecchi \[1997, 127, 134\]](#) and [Collett \[2003b, 301, 306\]](#)). The final estimates of the sample size are based on the approximation of the probability of an event due to [Freedman \(1982\)](#), the default, or in the presence of uniform accrual, due to [Schoenfeld \(1983\)](#) (see also [Collett 2003b](#)).

You can use `power logrank` to

- compute required number of events and sample size when you know power and effect size (expressed as a hazard ratio or log hazard-ratio);
- compute power when you know sample size (number of events) and effect size (expressed as a hazard ratio or log hazard-ratio); or
- compute effect size (hazard ratio or log hazard-ratio) and experimental-group survival when you know sample size (number of events) and power.

You can also choose between the Freedman or Schoenfeld computational approaches, adjust results for administrative censoring, adjust results for uniform accrual of subjects to the study, and adjust results for withdrawal of subjects from the study.

## Using `power logrank`

`power logrank` computes sample size, power, or effect size for the log-rank test comparing the survivor functions in two groups. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced- or equal-allocation design; see [\[PSS\] unbalanced designs](#) for a description of how to specify an unbalanced design.

To compute a total sample size, you specify an effect size and optionally power of the test in the `power()` option. The default power is set to 0.8. By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see [Fractional sample sizes in \[PSS\] unbalanced designs](#) for an example. The `nfractional` option is allowed only for sample-size determination.



To compute power, you must specify the total sample size in the `n()` option and an effect size.

An effect size may be specified either as a hazard ratio supplied in the `hratio()` option or as a log hazard-ratio supplied in the `lnhratio()` option. If neither is specified, a hazard ratio of 0.5 is assumed.

To compute an effect size, which may be expressed either as a hazard ratio or as a log hazard-ratio, you must specify the total sample size in the `n()` option; the power in the `power()` option; and, optionally, the direction of the effect. The direction is lower by default, `direction(lower)`, which means that the target hazard ratio is assumed to be less than one or that target log hazard-ratio is negative. In other words, the experimental treatment is presumed to be an improvement over the control treatment. If you want to change the direction to upper, corresponding to the target hazard ratio being greater than one, use `direction(upper)`.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see *Two samples* in [PSS] **unbalanced designs** for more details.

As we mentioned earlier, the effect size for power logrank may be expressed as a hazard ratio or as a log hazard-ratio. By default, the effect size, which is labeled as `delta` in the output, corresponds to the hazard ratio for the Freedman method and to the log hazard-ratio for the Schoenfeld method. You can change this by specifying the `effect()` option: `effect(hratio)` (the default) reports the hazard ratio and `effect(lnhratio)` reports the log hazard-ratio.

By default, all computations assume no censoring. In the presence of **administrative censoring**, you must specify a survival probability at the end of the study in the control group as the first command argument. You can also specify a survival probability at the end of the study in the experimental group as the second command argument. Otherwise, it will be computed using the specified hazard ratio or log hazard-ratio and the control-group survival probability. To accommodate an **accrual period** under the assumption of uniform accrual, survival information may instead be supplied in option `simpson()` or in option `st1()`; see *Including information about subject accrual* for details.

When computing sample size, you can adjust for withdrawal of subjects from the study by specifying the anticipated proportion of withdrawals in the `wdprob()` option.

By default, power logrank performs computations for a hazard-ratio test. Use the `schoenfeld` option to request computations for a log-hazard-ratio test.

In the presence of censoring, effect-size determination requires iteration. The default initial value of the estimated hazard ratio or, if `schoenfeld` is specified, of log hazard-ratio is obtained based on the formula assuming no censoring. This value may be changed by specifying the `init()` option. See [PSS] **power** for the descriptions of other options that control the iteration procedure.

In the following sections, we describe the use of power logrank accompanied by examples for computing sample size, power, and effect size.

## Computing sample size

To compute sample size and number of events, you must specify an effect size (a hazard ratio or a log hazard-ratio) and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified. A hazard ratio of 0.5 is assumed if an effect size is not specified.

**Computing sample size in the absence of censoring**

We demonstrate several examples of how to use `power logrank` to obtain the estimates of sample size and number of events using [Freedman \(1982\)](#) and [Schoenfeld \(1981\)](#) methods for uncensored data (a type I study when no censoring of subjects occurs).

▷ **Example 1: Number of events (failures) using Freedman method**

Consider a survival study to be conducted to compare the survivor function of subjects receiving a treatment (the experimental group) to the survivor function of those receiving a placebo or no treatment (the control group) using the log-rank test. Suppose that the study continues until all subjects fail (no censoring). The investigator wants to know how many events need to be observed in the study to achieve a power of 80% of a two-sided log-rank test with  $\alpha = 0.05$  to detect a 50% reduction in the hazard of the experimental group ( $\Delta_a = 0.5$ ). Because the default settings of `power logrank` are `power(0.8)`, `alpha(0.05)`, and `hratio(0.5)`, to obtain the estimate of the required number of events for the above study using the Freedman method (the default), we simply type

```
. power logrank
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.5000 (hazard ratio)
    hratio =    0.5000

Censoring:
    Pr_E =    1.0000

Estimated number of events and sample sizes:
    E =        72
    N =        72
    N per group =    36
```

From the output, a total of 72 events (failures) must be observed to achieve the required power of 80%. Because all subjects experience an event by the end of the study (`Pr_E=1.0000`), the number of subjects required to be recruited to the study is equal to the number of events. That is, the investigator needs to recruit a total of 72 subjects (36 per group) to the study.

## ▷ Example 2: Number of events (failures) using Schoenfeld method

Following [example 1](#), we can request the Schoenfeld method by specifying the `schoenfeld` option.

```
. power logrank, schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.6931 (log hazard-ratio)
      hratio =    0.5000
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample sizes:
      E =        66
      N =        66
      N per group =    33
```

We obtain a slightly smaller estimate (66) of the total number of events and subjects.

◀

## □ Technical note

[Freedman \(1982\)](#) and [Schoenfeld \(1981\)](#) derive the formulas for the number of events based on the asymptotic distribution of the log-rank test statistic. [Freedman \(1982\)](#) uses the asymptotic mean and variance of the log-rank test statistic expressed as a function of the true hazard ratio,  $\Delta$ , whereas [Schoenfeld \(1981\)](#) (see also [Collett \[2003b, 302\]](#)) bases the derivation on the asymptotic mean of the log-rank test statistic as a function of the true log hazard-ratio,  $\ln(\Delta)$ . We label the corresponding approaches as “Freedman method” and “Schoenfeld method” in the output.

For values of the hazard ratio close to one, the two methods tend to give similar results. In general, the Freedman method gives higher estimates than the Schoenfeld method. The performance of the Freedman method was studied by [Lakatos and Lan \(1992\)](#) and was found to slightly overestimate the sample size under the assumption of proportional hazards. [Hsieh \(1992\)](#) investigated the performance of the two methods under unequal allocation and concluded that Freedman’s formula predicts the highest power for the log-rank test when the sample-size ratio of the two groups equals the reciprocal of the hazard ratio. Schoenfeld’s formula predicts highest powers when sample sizes in the two groups are equal.

□

### ▷ Example 3: Unbalanced design

By default, `power logrank` computes sample size for a balanced- or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size and number of events for an unbalanced design by specifying the `nratio()` option.

Continuing with [example 1](#), we anticipate being able to recruit twice as many subjects in the experimental group; that is,  $n_2/n_1 = 2$ . We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```
. power logrank, nratio(2)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.5000 (hazard ratio)
      hratio =    0.5000
      N2/N1 =    2.0000
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample sizes:
      E =        63
      N =        63
      N1 =       21
      N2 =       42
```

We need a total of 63 subjects—21 in the control group and 42 in the experimental group.

Also see [Two samples in \[PSS\] unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.



### Computing sample size in the presence of censoring

Because of constraints on costs and time, it is often infeasible to continue the study until all subjects experience an event. Instead, the study terminates at some prespecified point in time. As a result, some subjects may not experience an event by the end of the study; that is, administrative censoring of subjects occurs. This increases the requirement on the number of subjects in the study to ensure that a certain number of events is observed.

In the presence of censoring (for a type II study), [Freedman \(1982\)](#) assumes the following. The analysis occurs at a fixed time  $t^*$  after the last patient was accrued, and all information about subject follow-up beyond time  $t^*$  is excluded. To minimize an overestimation of the sample size because of neglecting this information, the author suggests choosing  $t^*$  as the minimum follow-up time,  $f$ , beyond which the frequency of occurrence of events is low (the time at which, say, 85% of the total events expected are observed). Under this assumption, the number of required subjects does not depend on the rates of accrual and occurrence of events but only on the proportions of patients in the two treatment groups,  $s_1$  and  $s_2$ , surviving after  $f$ . See [Including information about subject accrual](#) about how to compute the sample size in the presence of a long accrual.

If censoring of subjects occurs, the probability of a subject not being censored needs to be estimated to obtain an accurate estimate of the required sample size. The assumption above justifies a simple procedure, suggested by [Freedman \(1982\)](#) and used by default by `power logrank`, to compute

this probability using the estimates of survival probabilities at the end of the study in the control and the experimental groups. Therefore, for a type II study (under administrative censoring), these probabilities must be supplied to `power logrank`.

#### ► Example 4: Sample size in the presence of censoring using Freedman method

Consider an example from [Machin et al. \(2009, 91\)](#) of a study of patients with resectable colon cancer. The goal of the study was to compare the efficacy of the drug levamisole against a placebo with respect to relapse-free survival, using a one-sided log-rank test with a significance level of 5%. The investigators anticipated a 10% increase (from 50% to 60%, with a respective hazard ratio of 0.737) in the survival of the experimental group with respect to the survival of the control (placebo) group at the end of the study. They wanted to detect this increase with a power of 80%. To obtain the required sample size, we enter the survival probabilities 0.5 and 0.6 as arguments and specify the `onesided` option to request a one-sided test.

```
. power logrank 0.5 0.6, onesided
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.7370 (hazard ratio)
      hratio = 0.7370
Censoring:
      s1 = 0.5000
      s2 = 0.6000
      Pr_E = 0.4500
Estimated number of events and sample sizes:
      E = 270
      N = 600
      N per group = 300
```

From the above output, the investigators would have to observe a total of 270 events (relapses) to detect a 26% decrease in the hazard ( $\Delta_a = 0.737$ ) of the experimental group relative to the hazard of the control group with a power of 80% using a one-sided log-rank test with  $\alpha = 0.05$ . They would have to recruit a total of 600 patients (300 per group) to observe that many events.

In contrast, in the absence of censoring, only 270 subjects would have been required to detect a decrease in hazard corresponding to  $\Delta_a = 0.737$ :

```

. power logrank, hratio(0.737) onesided
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    hratio =    0.7370
Censoring:
    Pr_E =    1.0000
Estimated number of events and sample sizes:
    E =      270
    N =      270
    N per group =    135

```

◀

### ► Example 5: Sample size in the presence of censoring using Schoenfeld method

If we wanted to use the Schoenfeld method to calculate sample size for the study described in [example 4](#), we could type

```

. power logrank 0.5 0.6, onesided schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) < 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =   -0.3052 (log hazard-ratio)
    hratio =    0.7370
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =      266
    N =      590
    N per group =    295

```

We find that 590 subjects are required in the study to observe a total of 266 events to ensure a power of 80%.

See the technical note in [Computing sample size in the absence of censoring](#) for a brief comparison of the Freedman and Schoenfeld methods.

◀

## ▷ Example 6: Alternative ways of specifying effect

If we wish, we can redefine effect size  $\delta$  in [example 4](#) to be a log hazard-ratio by specifying the `effect()` option.

```
. power logrank 0.5 0.6, onesided effect(lnhratio)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =  -0.3052 (log hazard-ratio)
    hratio =    0.7370
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =      270
    N =      600
    N per group =    300
```

The effect size  $\delta$  now contains the value of the log hazard-ratio.

Continuing with [example 4](#), instead of the estimate of the survival probability in the experimental group, we may have an estimate of the hazard ratio  $\Delta_a$ . For example, the estimate of the hazard ratio in this example is 0.737. We can specify the value of the hazard ratio in the `hratio()` option instead of specifying the experimental-group survival probability 0.6.

```
. power logrank 0.5, onesided hratio(0.737)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    hratio =    0.7370
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =      270
    N =      600
    N per group =    300
```

Alternatively, instead of the hazard ratio we can specify the log hazard-ratio in option `lnhratio()`.

```
. power logrank 0.5, onesided lnhratio(-0.3052)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    ln(hratio) = -0.3052
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =    270
    N =    600
    N per group =    300
```

The results are identical to the prior results. The estimate of the log hazard-ratio is now displayed in the output instead of the hazard ratio.

◀

## Withdrawal of subjects from the study

Under administrative censoring, the subject is known to have experienced either of the two outcomes by the end of the study: survival or failure. Often, in practice, subjects may withdraw from the study before it terminates and therefore may not experience an event by the end of the study (or be censored) for nonadministrative reasons. Withdrawal of subjects from a study may greatly affect the estimate of the sample size and must be accounted for in the computations. Refer to [Survival data](#) in [PSS] [intro](#) and [PSS] [Glossary](#) for a formal definition of [withdrawal](#).

[Freedman \(1982\)](#) suggests a conservative adjustment for the estimate of the sample size in the presence of withdrawal, which is implemented in `power logrank`. Withdrawal is assumed to be independent of failure (event) times and administrative censoring.

The proportion of subjects anticipated to withdraw from a study may be specified by using `wdprob()`.



## ▷ Example 7: Withdrawal of subjects from the study

Continuing with [example 4](#), suppose that a withdrawal rate of 10% is expected in the study of colon cancer patients. To account for this, we also specify `wdprob(0.1)`.

```
. power logrank 0.5 0.6, onesided wdprob(0.1)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    hratio =    0.7370
Censoring and withdrawal:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
    Pr_w =    0.1000
Estimated number of events and sample sizes:
    E =    270
    N =    666
    N per group =    333
```

The estimate of the total sample size using the Freedman method increases from 600 to 666 when the withdrawal rate is assumed to be 10%. The adjustment of the estimate of the sample size for the withdrawal of subjects is conservative. It assumes equal withdrawals from each group; that is, 10% of subjects are lost by the end of the study in each group. This adjustment affects only the estimates of the sample sizes but not the number of events. This is because withdrawal is assumed to be independent of event times, and the ratio of subjects surviving until the end of the study in the two groups does not change under equal withdrawals.

◀

## Including information about subject accrual

Many clinical studies have an accrual period of  $r$ , during which the subjects are recruited to the study, and a follow-up period of  $f = T - r$ , during which the subjects are followed until the end of the study,  $T$ , and no new subjects enter the study. The information about the duration of the accrual and follow-up periods affects the probability of a subject experiencing an event or failing during the study.

[Freedman \(1982\)](#) suggests approximating the combined event-free probability as an average of the survival probabilities in the control and the experimental groups at the minimum follow-up time,  $t^* = f$  (the default approach used in `power logrank`). However, for a long accrual of subjects, this approach may overestimate the required number of subjects, often seriously, because it does not take into account the information about subject follow-up beyond time  $f$ . Here [Freedman \(1982\)](#) proposes to use the survival probabilities at the average follow-up time, defined as  $t^* = (f + T)/2 = f + 0.5r$ , instead of the minimum follow-up time,  $f$ .

Alternatively, [Schoenfeld \(1983\)](#) (see also [Collett \[2003b, 306\]](#)) presents a formula for the required number of subjects allowing for uniform accrual (entry, recruitment) over  $[0, r]$  and a follow-up period,  $f$ . This information is incorporated into the formula for the probability of an event (or failure). The formula involves the integrals of the survivor functions of the control and the experimental groups. [Schoenfeld \(1983\)](#) suggests approximating the integral by using Simpson's rule, which requires the

estimates of the survivor function at three specific time points:  $f$ ,  $0.5r + f$ , and  $T = r + f$ . It is sufficient to provide the estimates of these three survival probabilities,  $S_1(f)$ ,  $S_1(0.5r + f)$ , and  $S_1(T)$ , for the control group only. The corresponding survival probabilities of the experimental group are automatically computed using the value of the hazard ratio in `hratio()` (or `log hazard-ratio` in `lnhratio()`) and the proportional-hazards assumption.

The three estimates of the survival probabilities of the control group may be supplied by using the `simpson()` option to adjust the estimates of the sample size or power for uniform entry and a follow-up period. If the estimate of the survivor function over an array of values in the range  $[f, T]$  is available from a previous study, it can be supplied using the `st1()` option to form a more accurate approximation of the probability of an event using numerical integration (see [R] [dydx](#)). Here the value of the length of the accrual period is needed for the computation. It is computed as the difference between the maximum and the minimum values of the time variable  $varname_t$ , supplied using `st1()`, that is,  $r = T - f = \max(varname_t) - \min(varname_t)$ .

For more information, see [Cleves et al. \(2010, sec. 16.2\)](#).

### ► Example 8: Sample size in the presence of accrual and follow-up periods

Consider an example described in [Collett \(2003b, 309\)](#) of a survival study of chronic active hepatitis. A new treatment is to be compared with a standard treatment with respect to the survival times of the patients with this disease. The investigators want to detect a change in a hazard ratio of 0.57 with 90% power and a 5% two-sided significance level. Also subjects are to be entered into the study uniformly over a period of 18 months and then followed for 24 months. From the Kaplan–Meier estimate of the survivor function available for the control group, the survival probabilities at  $f = 24$ ,  $0.5r + f = 33$ , and  $T = 42$  months are 0.70, 0.57, and 0.45, respectively.

```
. power logrank, hratio(0.57) power(0.9) schoenfeld simpson(0.7 0.57 0.45)
note: probability of an event is computed using Simpson's rule with
      S1(f) = 0.70, S1(f+r/2) = 0.57, S1(T) = 0.45
      S2(f) = 0.82, S2(f+r/2) = 0.73, S2(T) = 0.63
```

Estimated sample sizes for two-sample comparison of survivor functions

Log-rank test, Schoenfeld method

Ho:  $\ln(\text{HR}) = 0$  versus Ha:  $\ln(\text{HR}) \neq 0$

Study parameters:

```
alpha = 0.0500
power = 0.9000
delta = -0.5621 (log hazard-ratio)
hratio = 0.5700
```

Censoring:

```
Pr_E = 0.3514
```

Estimated number of events and sample sizes:

```
E = 134
N = 380
N per group = 190
```

[Collett \(2003b, 305\)](#) reports the required number of events to be 133, which, apart from rounding, agrees with our estimate of 134. In a later example, [Collett \(2003, 309\)](#) uses the number of events, rounded to 140, to compute the required sample size as  $140/0.35 = 400$ , where 0.35 is the estimate of the combined probability of an event. By hand, without rounding the number of events, we compute the required sample size as  $133/0.35 = 380$  and obtain the same estimate of the total sample size as in the output.

Using the average follow-up time suggested by [Freedman \(1982\)](#), we obtain the following:

```
. power logrank 0.57, hratio(0.57) power(0.9) schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.5621 (log hazard-ratio)
    hratio =    0.5700
Censoring:
    s1 =    0.5700
    s2 =    0.7259
    Pr_E =    0.3521
Estimated number of events and sample sizes:
    E =    134
    N =    378
    N per group =    189
```

We specify the survival probability in the control group at  $t^* = 0.5r + f = 0.5 \times 18 + 24 = 33$  as  $S_1(33) = 0.57$  and the hazard ratio of 0.57 (coincidentally). The survival probability in the experimental group is  $S_2(33) = S_1(33)^\Delta = 0.57^{0.57} = 0.726$ . Here we obtain the estimate of the sample size, 378, which is close to the estimate of 380 computed using the more complicated approximation. In this example, the two approximations produce similar results, but this may not always be the case.

The approximation suggested by [Schoenfeld \(1983\)](#) and [Collett \(2003b\)](#) is considered to be more accurate because it takes into account information about the patient survival beyond the average follow-up time. In general, the [Freedman \(1982\)](#) and [Schoenfeld \(1983\)](#) approximations tend to give similar results when  $\{\tilde{S}(f) + \tilde{S}(T)\}/2 \approx \tilde{S}(0.5r + f)$ ; see [Methods and formulas](#) for a formal definition of  $\tilde{S}(\cdot)$ .

If we use the survival probability in the control group,  $S_1(24) = 0.7$ , at a follow-up time  $t^* = f = 24$  instead of the average follow-up time  $t^* = 33$  in the presence of an accrual period,

```
. power logrank 0.7, hratio(0.57) power(0.9) schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =   -0.5621 (log hazard-ratio)
    hratio =    0.5700
Censoring:
    s1 =    0.7000
    s2 =    0.8160
    Pr_E =    0.2420
Estimated number of events and sample sizes:
    E =    134
    N =    550
    N per group =    275
```

we obtain the estimate of the total sample size of 550, which is substantially greater than the previously estimated sample sizes of 380 and 378.

## Computing power

Sometimes the number of subjects available for the enrollment into the study is limited. In such cases, the researchers may want to investigate with what power they can detect a desired treatment effect for a given sample size.

To compute power, you must specify the sample size in the `n()` option and an effect size (a hazard ratio or a log hazard-ratio). A hazard ratio of 0.5 is assumed if an effect size is not specified.

### ▷ Example 9: Power determination

Recall the colon cancer study described in [example 4](#). Suppose that only 100 subjects are available to be recruited to the study. We find out how this affects the power to detect a hazard ratio of 0.737.

```
. power logrank 0.5, hratio(0.737) onesided n(100)
Estimated power for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
      alpha =    0.0500
        N =     100
  N per group =    50
      delta =    0.7370 (hazard ratio)
      hratio =    0.7370
Number of events and censoring:
      E =      46
      s1 =    0.5000
      s2 =    0.6000
      Pr_E =    0.4500
Estimated power:
      power =    0.2646
```

The power to detect an alternative  $H_a: \Delta = 0.737$  decreased from 0.8 to 0.2646 when the sample size decreased from 600 to 100 (the number of events decreased from 270 to 46).

◀

### ▷ Example 10: Multiple values of study parameters

Continuing with [example 9](#), suppose we want to consider a range of sample sizes. We can specify a list (see [\[U\] 11.1.8 numlist](#)) of sample sizes in the `n()` option. For simplicity, we display only power, sample size, and number of events in the table.

```
. power logrank 0.5, hratio(0.737) onesided n(100(100)600) table(power N E)
Estimated power for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
```

power	N	E
.2646	100	46
.4174	200	91
.5455	300	136
.6505	400	181
.7344	500	226
.8004	600	271

As the sample size increases, the power increases. The decrease in sample size reduces the number of events observed in the study and therefore changes the estimates of the power. If the number of events were fixed, power would have been independent of the sample size, provided that all other parameters were held constant, because the formulas relate power directly to the number of events and not the number of subjects.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] [power, table](#). If you wish to produce a power plot, see [PSS] [power, graph](#).

◀

## Computing effect size

Effect size  $\delta$  for a log-rank test comparing two survivor functions is defined as a hazard ratio (or a log hazard-ratio) of the experimental group to the control group. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by the log-rank test with a certain power.

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, both power and sample size must be specified in options `power()` and `n()`, respectively. Additionally, you may also choose the direction of the effect by specifying the `direction()` option. `direction(lower)` is the default, and it assumes  $\Delta_a < 1$  [or  $\ln(\Delta_a) < 0$ ]. You can use `direction(upper)` to compute  $\Delta_a > 1$  [or  $\ln(\Delta_a) > 0$ ].

### ► Example 11: Effect-size determination

Continuing with [example 10](#), we can find that the value of the hazard ratio that can be detected for a fixed sample size of 100 with 80% power is approximately 0.42, corresponding to an increase in survival probability from 0.5 to roughly 0.75.

```
. power logrank 0.5, onesided n(100) power(0.8)
Performing iteration ...
Estimated hazard ratio for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR < 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =     100
  N per group =    50
Number of events and censoring:
      E =      38
     s1 =    0.5000
     s2 =    0.7455
    Pr_E =    0.3772
Estimated effect size and hazard ratio:
      delta =    0.4237 (hazard ratio)
     hratio =    0.4237
```

Under the censoring information, `power logrank` also reports the experimental-group survival rate at the end of the study corresponding to the computed hazard ratio—`s2=0.7455` in our example.

◀

**Testing a hypothesis about two survivor functions using a log-rank test**▷ **Example 12: Using the log-rank test to detect a change in survival in two groups**

Similar to [example 4](#), consider the generated dataset `drug.dta`, consisting of variables `drug` (a drug type) and `failtime` (a time to failure).

```
. use http://www.stata-press.com/data/r14/drug
(Patient Survival in Drug Trial)
. tabulate drug
```

Treatment type	Freq.	Percent	Cum.
Placebo	50	33.33	33.33
Drug A	50	33.33	66.67
Drug B	50	33.33	100.00
Total	150	100.00	

```
. by drug, sort: summarize failtime
```

```
-> drug = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
failtime	50	1.03876	.5535538	.1687701	2.382302

```
-> drug = Drug A
```

Variable	Obs	Mean	Std. Dev.	Min	Max
failtime	50	1.191802	.5927507	.2366922	2.277536

```
-> drug = Drug B
```

Variable	Obs	Mean	Std. Dev.	Min	Max
failtime	50	1.717314	.8350659	.5511715	3.796102

Failure times of the control group (Placebo) were generated from the Weibull distribution with  $\lambda_w = 0.693$  and  $p = 2$  (see [\[ST\] streg](#)); failure times of the two experimental groups, Drug A and Drug B, were generated from Weibull distributions with hazard functions proportional to the hazard of the control group in ratios 0.737 and 0.42, respectively. The Weibull family of survival distributions is chosen arbitrarily, and the Weibull parameter,  $\lambda_w$ , is chosen such that the survival at 1 year,  $t = 1$ , is roughly equal to 0.5. Subjects are randomly allocated to one of the three groups in equal proportions. Subjects with failure times greater than  $t = 1$  will be censored at  $t = 1$ .

Before analyzing these survival data, we need to set up the data using `stset`. After that, we can use `sts test`, `logrank` to test the survivor functions separately for Drug A against Placebo and Drug B against Placebo by using the log-rank test. See [\[ST\] stset](#) and [\[ST\] sts test](#) for more information about these two commands.

```

. stset failtime, exit(time 1)
      failure event: (assumed to fail at time=failtime)
obs. time interval: (0, failtime]
exit on or before: time 1

```

---

```

      150 total observations
       0 exclusions

```

---

```

      150 observations remaining, representing
      59 failures in single-record/single-failure data
128.985 total analysis time at risk and under observation
                                at risk from t =      0
                                earliest observed entry t =      0
                                last observed exit t =      1

```

```

. sts test drug if drug!=2, logrank
      failure _d: 1 (meaning all fail)
analysis time _t: failtime
exit on or before: time 1

```

**Log-rank test for equality of survivor functions**

drug	Events observed	Events expected
Placebo	25	22.17
Drug A	21	23.83
Total	46	46.00
	chi2(1) =	0.70
	Pr>chi2 =	0.4028

```

. sts test drug if drug!=1, logrank
      failure _d: 1 (meaning all fail)
analysis time _t: failtime
exit on or before: time 1

```

**Log-rank test for equality of survivor functions**

drug	Events observed	Events expected
Placebo	25	16.61
Drug B	13	21.39
Total	38	38.00
	chi2(1) =	7.55
	Pr>chi2 =	0.0060

From the results from `sts test` for the Drug A group, we fail to reject the null hypothesis of no difference between the survivor functions in the two groups; given our simulated data, the test made a type II error. On the other hand, for the Drug B group the one-sided  $p$ -value of 0.003, computed as  $0.006/2 = 0.003$ , suggests that the null hypothesis of nonsuperiority of the experimental treatment be rejected at the 0.005 significance level. We correctly conclude that the data provide the evidence that Drug B is superior to the Placebo.

Results from `sts test`, `logrank` for the two experimental groups agree with findings from examples 9 and 11. For the sample size of 100, the power of the log-rank test to detect the hazard ratio of 0.737 (10% increase in survival) is low (26%), whereas this sample size is sufficient for the test to detect a change in a hazard of 0.42 (25% increase in survival) with approximately 80% power.

Here we simulated our data from the alternative hypothesis and therefore can determine whether the correct decision or a type II error was made by the test. In practice, however, there is no way

to determine the accuracy of the decision from the test. All we know is that in a long series of trials, there is a 5% chance that a particular test will incorrectly reject the null hypothesis, a 74%  $[(1 - 0.25) \times 100\%$  given the power of 0.2646 obtained in [example 9](#)] chance that the test will miss the alternative  $H_a$ :  $\Delta = 0.737$ , and a 20%  $[(1 - 0.8) \times 100\%$  given the power of 0.8 in [example 11](#)] chance that the test will miss the alternative  $H_a$ :  $\Delta = 0.42$ .

◀

## Stored results

`power logrank` stores the following in `r()`:

### Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	total sample size
<code>r(N_a)</code>	actual sample size
<code>r(N1)</code>	sample size of the control group
<code>r(N2)</code>	sample size of the experimental group
<code>r(nratio)</code>	ratio of sample sizes, $N_2/N_1$
<code>r(nratio_a)</code>	actual ratio of sample sizes
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified; 0 otherwise
<code>r(onesided)</code>	1 for a one-sided test; 0 otherwise
<code>r(E)</code>	total number of events (failures)
<code>r(hratio)</code>	hazard ratio
<code>r(lnhratio)</code>	log hazard-ratio
<code>r(s1)</code>	survival probability in the control group (if specified)
<code>r(s2)</code>	survival probability in the experimental group (if specified)
<code>r(Pr_E)</code>	probability of an event (failure)
<code>r(Pr_w)</code>	proportion of withdrawals
<code>r(t_min)</code>	minimum time (if <code>st1()</code> is specified)
<code>r(t_max)</code>	maximum time (if <code>st1()</code> is specified)
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table; 0 otherwise
<code>r(init)</code>	initial value for hazard ratio or log hazard-ratio
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged; 0 otherwise

### Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>logrank</code>
<code>r(test)</code>	<code>freedman</code> or <code>schoenfeld</code>
<code>r(effect)</code>	<code>hratio</code> or <code>lnhratio</code>
<code>r(survvar)</code>	name of the variable containing survival probabilities (if <code>st1()</code> is specified)
<code>r(timevar)</code>	name of the variable containing time points (if <code>st1()</code> is specified)
<code>r(direction)</code>	<code>lower</code> or <code>upper</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

### Matrix

<code>r(pss_table)</code>	table of results
<code>r(simpmat)</code>	control-group survival probabilities (if <code>simpson()</code> is specified)



## Methods and formulas

Let  $S_1(t)$  and  $S_2(t)$  denote the survivor functions of the control and the experimental groups and  $\Delta(t) = \ln\{S_2(t)\}/\ln\{S_1(t)\}$  denote the hazard ratio at time  $t$  of the experimental to the control groups. Thus, for a given constant hazard ratio  $\Delta$ , the survivor function of the experimental group at any time  $t > 0$  may be computed as  $S_2(t) = \{S_1(t)\}^\Delta$  under the assumption of proportional hazards. Define  $E$  and  $n$  to be the total number of events and the total number of subjects required for the study, respectively;  $p_w$  to be the proportion of subjects withdrawn from the study (lost to follow-up); and  $z_{(1-\alpha/k)}$  and  $z_{(1-\beta)}$  to be the  $(1 - \alpha/k)$ th and the  $(1 - \beta)$ th quantiles of the standard normal distribution, respectively, with  $k = 1$  for the one-sided test and  $k = 2$  for the two-sided test. Let  $R$  be the allocation ratio to the experimental group with respect to the control group, that is,  $n_2 = Rn_1$ .

The total number of events required to be observed in a study to ensure a power of  $\pi = 1 - \beta$  of the log-rank test to detect the hazard ratio  $\Delta$  with significance level  $\alpha$ , according to [Freedman \(1982\)](#), is

$$E = \frac{1}{R}(z_{1-\alpha/k} + z_{1-\beta})^2 \left( \frac{R\Delta + 1}{\Delta - 1} \right)^2$$

and, according to [Schoenfeld \(1983\)](#) and [Collett \(2003a, 301\)](#), is

$$E = \frac{1}{R}(z_{1-\alpha/k} + z_{1-\beta})^2 \left\{ \frac{1+R}{\ln(\Delta)} \right\}^2$$

Both formulas are approximations and rely on a set of assumptions such as distinct failure times, all subjects completing the course of the study (no withdrawal), and a constant ratio,  $R$ , of subjects at risk in two groups at each failure time.

The total sample size required to observe the total number of events,  $E$ , is given by

$$n = \frac{E}{p_E}$$

The number of subjects required to be recruited in each group is obtained as  $n_1 = n/(1 + R)$  and  $n_2 = nR/(1 + R)$ . If `nfractional` is not specified, sample sizes and the number of events are rounded to integer values; see [Fractional sample sizes in \[PSS\] unbalanced designs](#) for details.

By default, the probability of an event (failure),  $p_E$ , is approximated as suggested by [Freedman \(1982\)](#),

$$p_E = 1 - \frac{S_1(t^*) + RS_2(t^*)}{1 + R}$$

where  $t^*$  is the minimum follow-up time,  $f$ , or, in the presence of an accrual period, the average follow-up time,  $(f + T)/2 = f + 0.5r$ .

If `simpson()` is specified, the probability of an event is approximated using Simpson's rule as suggested by [Schoenfeld \(1983\)](#):

$$p_E = 1 - \frac{1}{6} \left\{ \tilde{S}(f) + 4\tilde{S}(0.5r + f) + \tilde{S}(T) \right\}$$

where  $\tilde{S}(t) = \{S_1(t) + RS_2(t)\}/(1 + R)$  and  $f$ ,  $r$ , and  $T = f + r$  are the follow-up period, the accrual period, and the total duration of the study, respectively.

The methods do not incorporate time explicitly but rather use it to determine values of the survival probabilities  $S_1(t)$  and  $S_2(t)$  used in the computations.

If `st1()` is used, the integral in the expression for the probability of an event

$$p_E = 1 - \frac{1}{r} \int_f^T \tilde{S}(t) dt$$

is computed numerically using cubic splines (see [R] **dydx**). The value of  $r$  is computed as the difference between the maximum and the minimum values of  $varname_t$  in `st1()`,  $r = T - f = \max(varname_t) - \min(varname_t)$ .

To account for the proportion of subjects,  $p_w$ , withdrawn from the study (lost to follow-up), a conservative adjustment to the total sample size is applied as follows:

$$n_w = \frac{n}{1 - p_w}$$

Equal withdrawal rates are assumed in the adjustment of the group sample sizes for the withdrawal of subjects. Equal withdrawals do not affect the estimates of the number of events, provided that withdrawal is independent of event times and the ratio of subjects at risk in two groups remains constant at each failure time.

The power for each method is estimated using the formula

$$\pi = 1 - \beta = \Phi\{|\psi|^{-1}(Rnp_E)^{1/2} - z_{1-\alpha/k}\}$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function;  $\psi = (R\Delta + 1)/(\Delta - 1)$  or  $\psi = (1 + R)/\ln(\Delta)$  if the `schoenfeld` option is specified.

The estimate of the hazard ratio (or log hazard-ratio) for fixed power and sample size is computed (iteratively for censoring) using the formulas for the sample size given above. The value of the hazard ratio (log hazard-ratio) corresponding to the reduction in a hazard of the experimental group relative to the control group is reported by default.

## References

- Cleves, M. A., W. W. Gould, R. G. Gutierrez, and Y. V. Marchenko. 2010. *An Introduction to Survival Analysis Using Stata*. 3rd ed. College Station, TX: Stata Press.
- Collett, D. 2003a. *Modelling Binary Data*. 2nd ed. London: Chapman & Hall/CRC.
- . 2003b. *Modelling Survival Data in Medical Research*. 2nd ed. London: Chapman & Hall/CRC.
- Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.
- Hsieh, F. Y. 1992. Comparing sample size formulae for trials with unbalanced allocation using the logrank test. *Statistics in Medicine* 11: 1091–1098.
- Lakatos, E., and K. K. G. Lan. 1992. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine* 11: 179–191.
- Machin, D., M. J. Campbell, S. B. Tan, and S. H. Tan. 2009. *Sample Size Tables for Clinical Studies*. 3rd ed. Chichester, UK: Wiley–Blackwell.
- Marubini, E., and M. G. Valsecchi. 1997. *Analysing Survival Data from Clinical Trials and Observational Studies*. Chichester, UK: Wiley.

Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68: 316–319.

———. 1983. Sample-size formula for the proportional-hazards regression model. *Biometrics* 39: 499–503.

Also see [PSS] [intro](#) for more references.

## Also see

[PSS] [power](#) — Power and sample-size analysis for hypothesis tests

[PSS] [power cox](#) — Power analysis for the Cox proportional hazards model

[PSS] [power exponential](#) — Power analysis for the exponential test

[PSS] [power, graph](#) — Graph results from the power command

[PSS] [power, table](#) — Produce table of results from the power command

[PSS] [Glossary](#)

[ST] [stcox](#) — Cox proportional hazards model

[ST] [sts test](#) — Test equality of survivor functions

[Description](#)[Syntax](#)[Options](#)[Remarks and examples](#)[Also see](#)

## Description

This entry describes the specifications of unbalanced designs with the `power` command for two-sample hypothesis tests. See [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

## Syntax

*Two samples, compute sample size for unbalanced designs*

Compute total sample size

```
power ..., nratio(numlist) [nfractional] ...
```

Compute one group size given the other

```
power ..., n#(numlist) compute(n1 | n2) [nfractional] ...
```

*Two samples, specify sample size for unbalanced designs*

Specify total sample size and allocation ratio

```
power ..., n(numlist) nratio(numlist) [nfractional] ...
```

Specify one of the group sizes and allocation ratio

```
power ..., n#(numlist) nratio(numlist) [nfractional] ...
```

Specify total sample size and one of the group sizes

```
power ..., n(numlist) n#(numlist) ...
```

Specify group sizes

```
power ..., n1(numlist) n2(numlist) ...
```

*twosampleopts*

Description

* <u>n</u> ( <i>numlist</i> )	total sample size; required to compute power or effect size
* <u>n1</u> ( <i>numlist</i> )	sample size of the control group
* <u>n2</u> ( <i>numlist</i> )	sample size of the experimental group
* <u>nratio</u> ( <i>numlist</i> )	ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes
<code>compute(<i>n1</i>   <i>n2</i>)</code>	solve for $N1$ given $N2$ or for $N2$ given $N1$
<u>nfractional</u>	allow fractional sample sizes

\*Starred options may be specified either as one number or as a list of values; see [\[U\] 11.1.8 numlist](#).

## Options

Main

`n(numlist)` specifies the total number of subjects in the study to be used for power or effect-size determination. If `n()` is specified, the power is computed. If `n()` and `power()` or `beta()` are specified, the minimum effect size that is likely to be detected in a study is computed.

`n1(numlist)` specifies the number of subjects in the control group to be used for power or effect-size determination.

`n2(numlist)` specifies the number of subjects in the experimental group to be used for power or effect-size determination.

`nratio(numlist)` specifies the sample-size ratio of the experimental group relative to the control group,  $N2/N1$ , for power or effect-size determination for two-sample tests. The default is `nratio(1)`, meaning equal allocation between the two groups.

`compute(n1 | n2)` requests that the `power` command compute one of the group sample sizes given the other one instead of the total sample size for two-sample tests. To compute the control-group sample size, you must specify `compute(n1)` and the experimental-group sample size in `n2()`. Alternatively, to compute the experimental-group sample size, you must specify `compute(n2)` and the control-group sample size in `n1()`.

`nfractional` specifies that fractional sample sizes be allowed. When this option is specified, fractional sample sizes are used in the intermediate computations and are also displayed in the output.

## Remarks and examples

Remarks are presented under the following headings:

### *Two samples*

*Specifying total sample size and allocation ratio*

*Specifying group sample sizes*

*Specifying one of the group sample sizes and allocation ratio*

*Specifying total sample size and one of the group sample sizes*

### *Fractional sample sizes*

By default, for a two-sample test, the `power` command assumes a balanced design, but you may request an unbalanced design. A common way of specifying an unbalanced design is by specifying the `nratio()` option. You can also specify group sample sizes directly in the `n1()` and `n2()` options.

All considered options that accept arguments allow you to specify either one value # or a *numlist*, a list of values as described in [U] 11.1.8 *numlist*. For simplicity, we demonstrate these options using only one value.

Below we describe in detail the specifications of unbalanced designs for two-sample methods and the handling of fractional sample sizes.

## Two samples

All two-sample methods, such as `power twomeans` and `power twoproportions`, support the following options for specifying sample sizes: the total sample size `n()`, individual sample sizes `n1()` and `n2()`, and allocation ratio `nratio()`. The `compute()` option is useful if you want to compute one of the group sizes given the other one instead of the total sample size.

We first describe the specifications and then demonstrate their use in real examples.

We start with the sample-size determination—the default computation performed by the `power` command. The “switch” option for sample-size determination is the `power()` option. If you do not specify this option, it is implied with the default value of 0.8 corresponding to 80% power.

By default, group sizes are assumed to be equal; that is, the `nratio(1)` option is implied.

```
. power ..., [nratio(1)] ...
```

You can supply a different allocation ratio  $n_2/n_1$  to `nratio()` to request an unbalanced design.

```
. power ..., nratio(#) ...
```

To compute power or effect size, you must supply information about group sample sizes to `power`. There are several ways for you to do this. The simplest one, perhaps, is to specify the total sample size in the `n()` option.

```
. power ..., n(#) ...
```

The specification above assumes a balanced design in which the two group sizes are the same.

To request an unbalanced design, you can specify the desired allocation ratio between the two groups in the `nratio()` option.

```
. power ..., n(#) nratio(#) ...
```

The `nratio()` options assumes that the supplied values are the ratios of the second (experimental or comparison) group to the first (control or reference) group.

Alternatively, you can specify the two group sizes directly,

```
. power ..., n1(#) n2(#) ...
```

or you can specify one of the group sizes and the allocation ratio:

```
. power ..., n1(#) nratio(#) ...  
. power ..., n2(#) nratio(#) ...
```

Also supported, but perhaps more rarely used, is a combination of the total sample size and one of the group sizes:

```
. power ..., n(#) n1(#) ...  
. power ..., n(#) n2(#) ...
```

Below we demonstrate the described specifications using the `power twomeans` command, which provides PSS analysis for tests of two independent means; see [\[PSS\] power twomeans](#) for details. In all examples, we use a value of 0 for the control-group mean, a value of 1 for the experimental-group mean, and the default values of the other study parameters.

## ► Example 1: Sample-size determination for a balanced design

By default, `power twomeans` computes sample size for a balanced design.

```
. power twomeans 0 1
Performing iteration ...
Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 1.0000
    m1 = 0.0000
    m2 = 1.0000
    sd = 1.0000
Estimated sample sizes:
      N = 34
N per group = 17
```

The required total sample size is 34, with 17 subjects in each group.

The above is equivalent to specifying the `nratio(1)` option:

```
. power twomeans 0 1, nratio(1)
Performing iteration ...
Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 1.0000
    m1 = 0.0000
    m2 = 1.0000
    sd = 1.0000
Estimated sample sizes:
      N = 34
N per group = 17
```

◀

## ► Example 2: Sample-size determination for an unbalanced design

To compute sample size for an unbalanced design, we specify the ratio of the experimental-group size to the control-group size in the `nratio()` option. For example, if we anticipate twice as many subjects in the experimental group as in the control group, we compute the corresponding sample size by specifying `nratio(2)`:

```

. power twomeans 0 1, nratio(2)
Performing iteration ...
Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 1.0000
    m1 = 0.0000
    m2 = 1.0000
    sd = 1.0000
    N2/N1 = 2.0000
Estimated sample sizes:
    N = 39
    N1 = 13
    N2 = 26

```

The required total sample size is 39, with 13 subjects in the control group and 26 subjects in the experimental group. Generally, unbalanced designs require more subjects than the corresponding balanced designs.

◀

### ▶ Example 3: Power determination for a balanced design

To computer power for a balanced design, we specify the total sample size in the `n()` option:

```

. power twomeans 0 1, n(30)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
    N = 30
    N per group = 15
    delta = 1.0000
    m1 = 0.0000
    m2 = 1.0000
    sd = 1.0000
Estimated power:
    power = 0.7529

```



Equivalently, we specify one of the group sizes in the `n1()` or `n2()` option:

```
. power twomeans 0 1, n1(15)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
      N = 30
     N1 = 15
     N2 = 15
    delta = 1.0000
     m1 = 0.0000
     m2 = 1.0000
     sd = 1.0000
Estimated power:
    power = 0.7529
```

Both specifications imply the `nratio(1)` option.

◀

#### ► Example 4: Power determination for an unbalanced design

As we described in *Two samples*, there are a number of ways for you to request an unbalanced design for power determination. Below we provide an example for each specification.

#### Specifying total sample size and allocation ratio

Similarly to [example 2](#) but for power determination, we request an unbalanced design with twice as many subjects in the experimental group as in the control group by specifying the `nratio(2)` option:

```
. power twomeans 0 1, n(30) nratio(2)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
      N = 30
     N1 = 10
     N2 = 20
    N2/N1 = 2.0000
    delta = 1.0000
     m1 = 0.0000
     m2 = 1.0000
     sd = 1.0000
Estimated power:
    power = 0.7029
```

The computed power of 0.7029 is lower than the power of 0.7529 of the corresponding balanced design from [example 3](#).

## Specifying group sample sizes

Instead of the total sample size and the allocation ratio, we can specify the group sample sizes directly in the `n1()` and `n2()` options:

```
. power twomeans 0 1, n1(10) n2(20)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N =      30
     N1 =     10
     N2 =     20
  N2/N1 =    2.0000
  delta =    1.0000
     m1 =    0.0000
     m2 =    1.0000
     sd =    1.0000
Estimated power:
    power =    0.7029
```

## Specifying one of the group sample sizes and allocation ratio

Alternatively, we can specify one of the group sizes and the allocation ratio. Here we specify the control-group size.

```
. power twomeans 0 1, n1(10) nratio(2)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N =      30
     N1 =     10
     N2 =     20
  N2/N1 =    2.0000
  delta =    1.0000
     m1 =    0.0000
     m2 =    1.0000
     sd =    1.0000
Estimated power:
    power =    0.7029
```

We could have specified the experimental-group size instead:

```
. power twomeans 0 1, n2(20) nratio(2)
(output omitted)
```

## Specifying total sample size and one of the group sample sizes

Finally, we can specify a combination of the total sample size and one of the group sizes—the control group:

```
. power twomeans 0 1, n1(10) n(30)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N =      30
     N1 =     10
     N2 =     20
  N2/N1 =    2.0000
   delta =    1.0000
     m1 =    0.0000
     m2 =    1.0000
     sd =    1.0000
Estimated power:
    power =    0.7029
```

or the experimental group:

```
. power twomeans 0 1, n2(20) n(30)
(output omitted)
```

◀

Options `n()`, `n1()`, and `n2()` require integer numbers. When you specify the `n1()` and `n2()` options, your sample sizes are guaranteed to be integers. This is not necessarily true for other specifications for which the resulting sample sizes may be fractional. See [Fractional sample sizes](#) for details about how the `power` command handles fractional sample sizes.

## Fractional sample sizes

Certain sample-size specifications may lead to fractional sample sizes. For example, if you specify an odd value for the total sample size of a two-sample study, the two group sample sizes would have to be fractional to accommodate the specified total sample size. Also, if you specify the `nratio()` option with a two-sample method, the resulting sample sizes may be fractional.

By default, the `power` command rounds sample sizes to integers and uses integer values in the computations. To ensure conservative results, the command rounds down the input sample sizes and rounds up the output sample sizes.

### ▶ Example 5: Output sample sizes

For example, when we compute sample size, the sample size is rounded up to the nearest integer by default:

```

. power onemean 0 1.5
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 1.5000
    m0 = 0.0000
    ma = 1.5000
    sd = 1.0000
Estimated sample size:
    N = 6

```

We computed sample size for a one-sample mean test; see [PSS] [power onemean](#) for details.

We can specify the `nfractional` option to see the corresponding fractional sample size:

```

. power onemean 0 1.5, nfractional
Performing iteration ...
Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 1.5000
    m0 = 0.0000
    ma = 1.5000
    sd = 1.0000
Estimated sample size:
    N = 5.6861

```

The sample size of 6 reported above is the ceiling for the fractional sample size 5.6861.

We can also compute the actual power corresponding to the rounded sample size:

```

. power onemean 0 1.5, n(6)
Estimated power for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0
Study parameters:
    alpha = 0.0500
    N = 6
    delta = 1.5000
    m0 = 0.0000
    ma = 1.5000
    sd = 1.0000
Estimated power:
    power = 0.8325

```

The actual power corresponding to the sample size of 6 is larger than the specified power of 0.8 from the two previous examples because the sample size was rounded up.



On the other hand, the `power` command rounds down the input sample sizes.

## ▷ Example 6: Input sample sizes

For example, let's use `power twomeans` to compute the power of a two-sample means test using a total sample size of 51 and the default settings for other parameters; see [PSS] `power twomeans` for details.

```
. power twomeans 0 1, n(51)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
      alpha =    0.0500
        N =      51
      delta =    1.0000
        m1 =    0.0000
        m2 =    1.0000
        sd =    1.0000
Actual sample sizes:
        N =      50
  N per group =    25
Estimated power:
      power =    0.9337
```

By default, `power twomeans` assumes a balanced design. To accommodate a balanced design, the command rounds down the group sample sizes from 25.5 to 25 for an actual total sample size of 50.

When the specified sample sizes differ from the resulting rounded sample sizes, the actual sample sizes used in the computations are reported. In our example, we requested a total sample size of 51, but the actual sample size used to compute the power was 50.

We can specify the `nfractional` option to request that fractional sample sizes be used in the computations.

```
. power twomeans 0 1, n(51) nfractional
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
      alpha =    0.0500
        N =   51.0000
  N per group = 25.5000
      delta =    1.0000
        m1 =    0.0000
        m2 =    1.0000
        sd =    1.0000
Estimated power:
      power =    0.9382
```

The fractional group sample sizes of 25.5 are now used in the computations.

If we want to preserve the total sample size of 51 and ensure that group sample sizes are integers, we can specify the group sizes directly:

```
. power twomeans 0 1, n1(25) n2(26)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
      N = 51
     N1 = 25
     N2 = 26
  N2/N1 = 1.0400
  delta = 1.0000
     m1 = 0.0000
     m2 = 1.0000
     sd = 1.0000
Estimated power:
    power = 0.9381
```

Alternatively, we can specify one of the group sizes (or the total sample size) and the corresponding allocation ratio  $n_2/n_1 = 26/25 = 1.04$ :

```
. power twomeans 0 1, n1(25) nratio(1.04)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha = 0.0500
      N = 51
     N1 = 25
     N2 = 26
  N2/N1 = 1.0400
  delta = 1.0000
     m1 = 0.0000
     m2 = 1.0000
     sd = 1.0000
Estimated power:
    power = 0.9381
```

We obtain the same power of 0.9381.

In the above, the specified value of a sample-size ratio resulted in integer sample sizes. This may not always be the case. For example, if we specify the sample-size ratio of 1.3,

```
. power twomeans 0 1, n1(25) nratio(1.3)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N1 =      25
  N2/N1 =    1.3000
  delta =    1.0000
     m1 =    0.0000
     m2 =    1.0000
     sd =    1.0000
Actual sample sizes:
      N =      57
     N1 =      25
     N2 =      32
  N2/N1 =    1.2800
Estimated power:
    power =    0.9573
```

the experimental-group size of 32.5 is rounded down to 32. The total sample size used in the computation is 57, and the actual sample-size ratio is 1.28.

As before, we can specify the `nfractional` option to use the fractional experimental-group size of 32.5 in the computations:

```
. power twomeans 0 1, n1(25) nratio(1.3) nfractional
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
Study parameters:
    alpha =    0.0500
      N =   57.5000
     N1 =   25.0000
     N2 =   32.5000
  N2/N1 =    1.3000
  delta =    1.0000
     m1 =    0.0000
     m2 =    1.0000
     sd =    1.0000
Estimated power:
    power =    0.9585
```

4

## Also see

[PSS] [power](#) — Power and sample-size analysis for hypothesis tests

[PSS] [Glossary](#)





# Glossary

**1 : M matched case–control study.** See *matched study*.

**2 × 2 contingency table.** A 2 × 2 contingency table is used to describe the association between a binary independent variable and a binary response variable of interest.

**2 × 2 × K contingency table.** See *stratified 2 × 2 tables*.

**J × 2 contingency table.** A J × 2 contingency table is used to describe the association between an ordinal independent variable with J levels and a binary response variable of interest.

**acceptance region.** In *hypothesis testing*, an acceptance region is a set of sample values for which the *null hypothesis* cannot be rejected or can be accepted. It is the complement of the *rejection region*.

**accrual period or recruitment period or accrual.** The accrual period (or recruitment period) is the period during which subjects are being enrolled (recruited) into a study. Also see *follow-up period*.

**actual alpha, actual significance level.** This is an attained or observed *significance level*.

**administrative censoring.** Administrative censoring is the right-censoring that occurs when the study observation period ends. All subjects complete the course of the study and are known to have experienced one of two outcomes at the end of the study: survival or failure. This type of censoring should not be confused with *withdrawal* and *loss to follow-up*. Also see *censored, censoring, left-censoring, and right-censoring*.

**allocation ratio.** This ratio  $n_2/n_1$  represents the number of subjects in the comparison, *experimental group* relative to the number of subjects in the reference, *control group*. Also see [PSS] *unbalanced designs*.

**alpha.** Alpha,  $\alpha$ , denotes the *significance level*.

**alternative hypothesis.** In *hypothesis testing*, the alternative *hypothesis* represents the counterpoint to which the *null hypothesis* is compared. When the parameter being tested is a scalar, the alternative hypothesis can be either *one sided* or *two sided*.

**alternative value, alternative parameter.** This value of the parameter of interest under the *alternative hypothesis* is fixed by the investigator in a power and sample-size analysis. For example, alternative mean value and alternative mean refer to a value of the mean parameter under the alternative hypothesis.

**analysis of variance, ANOVA.** This is a class of statistical models that studies differences between means from multiple populations by partitioning the variance of the continuous outcome into independent sources of variation due to effects of interest and random variation. The test statistic is then formed as a ratio of the expected variation due to the effects of interest to the expected random variation. Also see *one-way ANOVA, two-way ANOVA, one-way repeated-measures ANOVA, and two-way repeated-measures ANOVA*.

**balanced design.** A balanced design represents an experiment in which the numbers of treated and untreated subjects are equal. For many types of *two-sample hypothesis tests*, the power of the test is maximized with balanced designs.

**beta.** Beta,  $\beta$ , denotes the *probability* of committing a *type II error*, namely, failing to reject the null hypothesis even though it is false.

**between-subjects design.** This is an experiment that has only *between-subjects factors*. See [PSS] *power oneway* and [PSS] *power twoway*.

- between-subjects factor.** This is a [factor](#) for which each subject receives only one of the levels.
- binomial test.** A binomial test is a test for which the exact sampling distribution of the test statistic is binomial; see [\[R\] bitest](#). Also see [\[PSS\] power oneproportion](#).
- bisection method.** This method finds a root  $x$  of a function  $f(x)$  such that  $f(x) = 0$  by repeatedly subdividing an interval on which  $f(x)$  is defined until the change in successive root estimates is within the requested tolerance and function  $f(\cdot)$  evaluated at the current estimate is sufficiently close to zero.
- case-control study.** An [observational study](#) that [retrospectively](#) compares characteristics of subjects with a certain problem (cases) with characteristics of subjects without the problem (controls). For example, to study association between smoking and lung cancer, investigators will sample subjects with and without lung cancer and record their smoking status. Case-control studies are often used to study rare diseases.
- CCT.** See [controlled clinical trial study](#).
- cell means.** These are means of the outcome of interest within cells formed by the cross-classification of the two [factors](#). See [\[PSS\] power twoway](#) and [\[PSS\] power repeated](#).
- cell-means model.** A cell-means model is an [ANOVA](#) model formulated in terms of [cell means](#).
- chi-squared test,  $\chi^2$  test.** This test for which either an asymptotic sampling distribution or a sampling distribution of a test statistic is  $\chi^2$ . See [\[PSS\] power onevariance](#) and [\[PSS\] power twoproportions](#).
- clinical trial.** A clinical trials is an experiment testing a medical treatment or procedure on human subjects.
- clinically meaningful difference, clinically meaningful effect, clinically significant difference.** Clinically meaningful difference represents the magnitude of an effect of interest that is of clinical importance. What is meant by “clinically meaningful” may vary from study to study. In [clinical trials](#), for example, if no prior knowledge is available about the performance of the considered clinical procedure, a standardized [effect size](#) (adjusted for standard deviation) between 0.25 and 0.5 may be considered of clinical importance.
- Cochran-Armitage test.** The Cochran-Armitage test is a test for a linear trend in a probability of response in a  $J \times 2$  [contingency table](#). The test statistic has an asymptotic  $\chi^2$  distribution under the null hypothesis. See [\[PSS\] power trend](#).
- Cochran-Mantel-Haenszel test.** See [Mantel-Haenszel test](#).
- cohort study.** Typically an [observational study](#), a cohort study may also be an [experimental study](#) in which a cohort, a group of subjects who have similar characteristics, is followed over time and evaluated at the end of the study. For example, cohorts of vaccinated and unvaccinated subjects are followed over time to study the effectiveness of influenza vaccines.
- columns in graph.** Think of `power`, `graph()` as graphing the columns of `power`, `table`. One of the columns will be placed on the  $x$  axis, another will be placed on the  $y$  axis, and, if you have more columns with varying values, separate plots will be created for each. Similarly, we use the terms “column symbol”, “column name”, and “column label” to refer to symbols, names, and labels that appear in tables when tabular output is requested.
- common odds ratio.** A measure of association in [stratified  \$2 \times 2\$  tables](#). It can be viewed as a weighted aggregate of stratum-specific [odds ratios](#).
- comparison value.** See [alternative value](#).
- compound symmetry.** A covariance matrix has a compound-symmetry structure if all the variances are equal and all the covariances are equal. This is a special case of the [sphericity](#) assumption.

- concordant pairs.** In a  $2 \times 2$  contingency table, a concordant pair is a pair of observations that are both either successes or failures. Also see *discordant pairs* and *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.
- contrasts.** Contrasts refers to a linear combination of cell means such that the sum of contrast coefficients is zero.
- control group.** A control group comprises subjects that are randomly assigned to a group where they receive no treatment or receives a standard treatment. In *hypothesis testing*, this is usually a reference group. Also see *experimental group*.
- controlled clinical trial study.** This is an *experimental study* in which treatments are assigned to two or more groups of subjects without the randomization.
- critical region.** See *rejection region*.
- critical value.** In *hypothesis testing*, a critical value is a boundary of the *rejection region*.
- cross-sectional study.** This type of *observational study* measures various population characteristics at one point in time or over a short period of time. For example, a study of the prevalence of breast cancer in the population is a cross-sectional study.
- delta.** Delta,  $\delta$ , in the context of power and sample-size calculations, denotes the *effect size*.
- directional test.** See *one-sided test*.
- discordant pairs.** In a  $2 \times 2$  contingency table, discordant pairs are the success–failure or failure–success pairs of observations. Also see *concordant pairs* and *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.
- discordant proportion.** This is a proportion of *discordant pairs* or *discordant sets*. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions** as well as *Introduction* under *Remarks and examples* in [PSS] **power mcc**.
- discordant sets.** In a matched study with multiple controls matched to a case, discordant sets are the sets in which there is any success–failure or failure–success match between the case and any matched control. Also see *Introduction* under *Remarks and examples* in [PSS] **power mcc**.
- dropout.** Dropout is the withdrawal of subjects before the end of a study and leads to incomplete or missing data.
- effect size.** The effect size is the size of the *clinically significant difference* between the treatments being compared, typically expressed as a quantity that is independent of the unit of measure. For example, in a *one-sample mean test*, the effect size is a standardized difference between the mean and its reference value. In other cases, the effect size may be measured as an *odds ratio* or a *risk ratio*. See [PSS] **intro** to learn more about the relationship between effect size and the power of a test.
- effect-size curve.** The effect-size curve is a graph of the estimated *effect size* or *target parameter* as a function of some other study parameter such as the *sample size*. The effect size or target parameter is plotted on the  $y$  axis, and the sample size or other parameter is plotted on the  $x$  axis.
- effect-size determination.** This pertains to the computation of an *effect size* or a *target parameter* given *power*, *sample size*, and other study parameters.
- equal-allocation design.** See *balanced design*.
- exact test.** An exact test is one for which the probability of observing the data under the null hypothesis is calculated directly, often by enumeration. Exact tests do not rely on any asymptotic approximations and are therefore widely used with small datasets. See [PSS] **power oneproportion** and [PSS] **power twoproportions**.

**experimental group.** An experimental group is a group of subjects that receives a treatment or procedure of interest defined in a controlled experiment. In [hypothesis testing](#), this is usually a comparison group. Also see [control group](#).

**experimental study.** In an experimental study, as opposed to an [observational study](#), the assignment of subjects to treatments is controlled by investigators. For example, a study that compares a new treatment with a standard treatment by assigning each treatment to a group of subjects is an experimental study.

**exponential test.** The exponential test is the parametric test comparing the hazard rates,  $\lambda_1$  and  $\lambda_2$ , (or log hazards) from two independent exponential (constant only) regression models with the null hypothesis  $H_0: \lambda_2 - \lambda_1 = 0$  [or  $H_0: \ln(\lambda_2) - \ln(\lambda_1) = \ln(\lambda_2/\lambda_1) = 0$ ].

**exposure odds ratio.** An [odds ratio](#) of exposure in cases relative to controls in a [case-control study](#).

**F test.** An  $F$  test is a test for which a sampling distribution of a test statistic is an  $F$  distribution. See [\[PSS\] power twovariances](#).

**factor, factor variables.** This is a categorical explanatory variable with any number of levels.

**finite population correction.** When sampling is performed without replacement from a finite population, a finite population correction is applied to the standard error of the estimator to reduce sampling variance.

**Fisher–Irwin’s exact test.** See [Fisher’s exact test](#).

**Fisher’s exact test.** Fisher’s exact test is an [exact small-sample test](#) of independence between rows and columns in a  $2 \times 2$  contingency table. Conditional on the marginal totals, the test statistic has a hypergeometric distribution under the null hypothesis. See [\[PSS\] power twoproportions](#) and [\[R\] tabulate twoway](#).

**Fisher’s z test.** This is a  $z$  test comparing one or two correlations. See [\[PSS\] power onecorrelation](#) and [\[PSS\] power twocorrelations](#). Also see [Fisher’s z transformation](#).

**Fisher’s z transformation.** Fisher’s  $z$  transformation applies an inverse hyperbolic tangent transformation to the sample correlation coefficient. This transformation is useful for testing hypothesis concerning [Pearson’s correlation coefficient](#). The exact sampling distribution of the correlation coefficient is complicated, while the transformed statistic is approximately standard normal.

**fixed effects.** Fixed effects represent all levels of the factor that are of interest.

**follow-up period or follow-up.** The (minimum) follow-up period is the period after the last subject entered the study until the end of the study. The follow-up defines the phase of a study during which subjects are under observation and no new subjects enter the study. If  $T$  is the total duration of a study, and  $r$  is the accrual period of the study, then follow-up period  $f$  is equal to  $T - r$ . Also see [accrual period](#).

**follow-up study.** See [cohort study](#).

**Greenhouse–Geisser correction.** See [nonsphericity correction](#).

**hypothesis.** A hypothesis is a statement about a population parameter of interest.

**hypothesis testing, hypothesis test.** This method of inference evaluates the validity of a [hypothesis](#) based on a sample from the population. See [Hypothesis testing](#) under [Remarks and examples](#) in [\[PSS\] intro](#).

**hypothesized value.** See [null value](#).

**interaction effects.** Interaction effects measure the dependence of the effects of one factor on the levels of the other factor. Mathematically, they can be defined as the differences among treatment means that are left after [main effects](#) are removed from these differences.

**Lagrange multiplier test.** See *score test*.

**likelihood-ratio test.** The likelihood-ratio (LR) test is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. Under the null hypothesis, the constrained model fits the data as well as the full model. The LR test requires one to determine the maximal value of the log-likelihood function for both the constrained and the full models. See [PSS] [power twoproportions](#) and [R] [lrtest](#).

**loss to follow-up.** Subjects are lost to follow-up if they do not complete the course of the study for reasons unrelated to the event of interest. For example, loss to follow-up occurs if subjects move to a different area or decide to no longer participate in a study. Loss to follow-up should not be confused with administrative censoring. If subjects are lost to follow-up, the information about the outcome these subjects would have experienced at the end of the study, had they completed the study, is unavailable. Also see *withdrawal*, *administrative censoring*, and *follow-up period or follow-up*.

**lower one-sided test, lower one-tailed test.** A lower one-sided test is a [one-sided test](#) of a scalar parameter in which the [alternative hypothesis](#) is lower one sided, meaning that the alternative hypothesis states that the parameter is less than the value conjectured under the [null hypothesis](#). Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] [intro](#).

**main effects.** These are average, additive effects that are associated with each level of each factor. For example, the main effect of level  $j$  of a factor is the difference between the mean of all observations on the outcome of interest at level  $j$  and the grand mean.

**Mantel–Haenszel test.** The Mantel–Haenszel test evaluates whether the overall degree of association in [stratified  \$2 \times 2\$  tables](#) is significant assuming that the exposure effect is the same across strata. See [PSS] [power cmh](#).

**marginal homogeneity.** Marginal homogeneity refers to the equality of one or more row marginal proportions with the corresponding column proportions. Also see *Introduction* under *Remarks and examples* in [PSS] [power pairedproportions](#).

**marginal proportion.** This represents a ratio of the number of observations in a row or column of a [contingency table](#) relative to the total number of observations. Also see *Introduction* under *Remarks and examples* in [PSS] [power pairedproportions](#).

**matched study.** In a matched study, an observation from one group is matched to one or more observations from another group with respect to one or more characteristics of interest. When multiple matches occur, the study design is  $1 : M$ , where  $M$  is the number of matches. Also see *paired data*, also known as  $1 : 1$  matched data.

**McNemar’s test.** McNemar’s test is a test used to compare two dependent binary populations. The null hypothesis is formulated in the context of a  $2 \times 2$  contingency table as a hypothesis of [marginal homogeneity](#). See [PSS] [power pairedproportions](#) and the `mcc` command in [R] [epitab](#).

**MDES.** See *minimum detectable effect size*.

**mean contrasts.** See *contrasts*.

**minimum detectable effect size.** The minimum detectable [effect size](#) is the smallest effect size that can be detected by hypothesis testing for a given power and sample size.

**minimum detectable value.** The minimum detectable value represents the smallest amount or concentration of a substance that can be reliably measured.

**mixed design.** A mixed design is an experiment that has at least one [between-subjects factor](#) and one [within-subject factor](#). See [PSS] [power repeated](#).

**negative effect size.** In power and sample-size analysis, we obtain a negative [effect size](#) when the postulated value of the parameter under the alternative hypothesis is less than the hypothesized value of the parameter under the null hypothesis. Also see [positive effect size](#).

**nominal alpha, nominal significance level.** This is a desired or requested [significance level](#).

**noncentrality parameter.** In power and sample-size analysis, a noncentrality parameter is the expected value of the test statistic under the alternative hypothesis.

**nondirectional test.** See [two-sided test](#).

**nonsphericity correction.** This is a correction used for the degrees of freedom of a regular  $F$  test in a repeated-measures ANOVA to compensate for the lack of [sphericity](#) of the repeated-measures covariance matrix.

**null hypothesis.** In [hypothesis testing](#), the null [hypothesis](#) typically represents the conjecture that one is attempting to disprove. Often the null hypothesis is that a treatment has no effect or that a statistic is equal across populations.

**null value, null parameter.** This value of the parameter of interest under the [null hypothesis](#) is fixed by the investigator in a power and sample-size analysis. For example, null mean value and null mean refer to the value of the mean parameter under the null hypothesis.

**observational study.** In an observational study, as opposed to an [experimental study](#), the assignment of subjects to treatments happens naturally and is thus beyond the control of investigators. Investigators can only observe subjects and measure their characteristics. For example, a study that evaluates the effect of exposure of children to household pesticides is an observational study.

**observed level of significance.** See [p-value](#).

**odds and odds ratio.** The odds in favor of an event are  $\text{Odds} = p/(1-p)$ , where  $p$  is the probability of the event. Thus if  $p = 0.2$ , the odds are 0.25, and if  $p = 0.8$ , the odds are 4.

The log of the odds is  $\ln(\text{Odds}) = \text{logit}(p) = \ln\{p/(1-p)\}$ , and logistic regression models, for instance, fit  $\ln(\text{Odds})$  as a linear function of the covariates.

The odds ratio is a ratio of two odds:  $\text{Odds}_2/\text{Odds}_1$ . The individual odds that appear in the ratio are usually for an experimental group and a control group or for two different demographic groups.

**one-sample test.** A one-sample test compares a parameter of interest from one sample with a reference value. For example, a one-sample mean test compares a mean of the sample with a reference value.

**one-sided test, one-tailed test.** A one-sided test is a [hypothesis test](#) of a scalar parameter in which the [alternative hypothesis](#) is one sided, meaning that the alternative hypothesis states that the parameter is either less than or greater than the value conjectured under the [null hypothesis](#) but not both. Also see [One-sided test versus two-sided test](#) under [Remarks and examples](#) in [\[PSS\] intro](#).

**one-way ANOVA, one-way analysis of variance.** A one-way ANOVA model has a single [factor](#). Also see [\[PSS\] power oneway](#).

**one-way repeated-measures ANOVA.** A one-way repeated-measures ANOVA model has a single [within-subject factor](#). Also see [\[PSS\] power repeated](#).

**paired data.** Paired data consist of pairs of observations that share some characteristics of interest. For example, measurements on twins, pretest and posttest measurements, before and after measurements, repeated measurements on the same individual. Paired data are correlated and thus must be analyzed by using a [paired test](#).

**paired observations.** See [paired data](#).

**paired test.** A paired test is used to test whether the parameters of interest of two [paired populations](#) are equal. The test takes into account the dependence between measurements. For this reason, paired tests are usually more powerful than their [two-sample](#) counterparts. For example, a paired-means or paired-difference test is used to test whether the means of two paired (correlated) populations are equal.

**Pearson's correlation.** Pearson's correlation  $\rho$ , also known as the product-moment correlation, measures the degree of association between two variables. Pearson's correlation equals the variables' covariance divided by their respective standard deviations, and ranges between  $-1$  and  $1$ . Zero indicates no correlation between the two variables.

**population parameter.** See [target parameter](#).

**positive effect size.** In power and sample-size analysis, we obtain a positive [effect size](#) when the postulated value of the parameter under the alternative hypothesis is greater than the hypothesized value of the parameter under the null hypothesis. Also see [negative effect size](#).

**postulated value.** See [alternative value](#).

**power.** The power of a test is the probability of correctly rejecting the [null hypothesis](#) when it is false. It is often denoted as  $1 - \beta$  in the statistical literature, where  $\beta$  is the [type II error probability](#). Commonly used values for power are 80% and 90%. See [\[PSS\] intro](#) for more details about power.

**power and sample-size analysis.** Power and sample-size analysis investigates the optimal allocation of study resources to increase the likelihood of the successful achievement of a study objective. See [\[PSS\] intro](#).

**power curve.** A power curve is a graph of the estimated [power](#) as a function of some other study parameter such as the sample size. The power is plotted on the  $y$  axis, and the sample size or other parameter is plotted on the  $x$  axis. See [\[PSS\] power, graph](#).

**power determination.** This pertains to the computation of a [power](#) given sample size, effect size, and other study parameters.

**power function.** The power functions is a function of the population parameter  $\theta$ , defined as the probability that the observed sample belongs to the [rejection region](#) of a test for given  $\theta$ . See [Hypothesis testing](#) under *Remarks and examples* in [\[PSS\] intro](#).

**power graph.** See [power curve](#).

**probability of a type I error.** This is the probability of committing a [type I error](#) of incorrectly rejecting the [null hypothesis](#). Also see [significance level](#).

**probability of a type II error.** This is the probability of committing a [type II error](#) of incorrectly accepting the [null hypothesis](#). Common values for the probability of a type II error are 0.1 and 0.2 or, equivalently, 10% and 20%. Also see [beta](#) and [power](#).

**prospective study.** In a prospective study, the population or cohort is classified according to specific [risk factors](#), such that the outcome of interest, typically various manifestations of a disease, can be observed over time and tied in to the initial classification. Also see [retrospective study](#).

**PSS analysis.** See [power and sample-size analysis](#).

**PSS Control Panel.** The PSS Control Panel is a point-and-click graphical user interface for [power and sample-size analysis](#). See [\[PSS\] GUI](#).

**p-value.**  $P$ -value is a probability of obtaining a test statistic as extreme or more extreme as the one observed in a sample assuming the [null hypothesis](#) is true.

**random effects.** Random effects represent a random sample of levels from all possible levels, and the interest lies in all possible levels.

- randomized controlled trial.** In this [experimental study](#), treatments are randomly assigned to two or more groups of subjects.
- RCT.** See *randomized controlled trial*.
- recruitment period.** See *accrual period*.
- reference value.** See *null value*.
- rejection region.** In [hypothesis testing](#), a rejection region is a set of sample values for which the [null hypothesis](#) can be rejected.
- relative risk.** See *risk ratio*.
- retrospective study.** In a retrospective study, a group with a disease of interest is compared with a group without the disease, and information is gathered in a retrospective way about the exposure in each group to various [risk factors](#) that might be associated with the disease. Also see *prospective study*.
- risk difference.** A risk difference is defined as the probability of an event occurring when a risk factor is increased by one unit minus the probability of the event occurring without the increase in the risk factor.
- When the risk factor is binary, the risk difference is the probability of the outcome when the risk factor is present minus the probability when the risk factor is not present.
- When one compares two populations, a risk difference is defined as a difference between the probabilities of an event in the two groups. It is typically a difference between the probability in the comparison group or experimental group and the probability in the reference group or control group.
- risk factor.** A risk factor is a variable that is associated with an increased or decreased probability of an outcome.
- risk ratio.** A risk ratio, also called a relative risk, measures the increase in the likelihood of an event occurring when a risk factor is increased by one unit. It is the ratio of the probability of the event when the risk factor is increased by one unit over the probability without that increase.
- When the risk factor is binary, the risk ratio is the ratio of the probability of the event when the risk factor occurs over the probability when the risk factor does not occur.
- When one compares two populations, a risk ratio is defined as a ratio of the probabilities of an event in the two groups. It is typically a ratio of the probability in the comparison group or experimental group to the probability in the reference group or control group.
- sample size.** This is the number of subjects in a sample. See [\[PSS\] intro](#) to learn more about the relationship between sample size and the power of a test.
- sample-size curve.** A sample-size curve is a graph of the estimated [sample size](#) as a function of some other study parameter such as power. The sample size is plotted on the  $y$  axis, and the power or other parameter is plotted on the  $x$  axis.
- sample-size determination.** This pertains to the computation of a [sample size](#) given power, effect size, and other study parameters.
- Satterthwaite's  $t$  test.** Satterthwaite's  $t$  test is a modification of the [two-sample  \$t\$  test](#) to account for unequal variances in the two populations. See [Methods and formulas](#) in [\[PSS\] power twomeans](#) for details.
- score test.** A score test, also known as a Lagrange multiplier test, is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. The null hypothesis is that the constrained model fits the



data as well as the full model. The score test only requires one to fit the constrained model. See [PSS] [power oneproportion](#) and [R] [prtest](#).

**sensitivity analysis.** Sensitivity analysis investigates the effect of varying study parameters on power, sample size, and other components of a study. The true values of study parameters are usually unknown, and power and sample-size analysis uses best guesses for these values. It is therefore important to evaluate the sensitivity of the computed power or sample size in response to changes in study parameters. See [PSS] [power, table](#) and [PSS] [power, graph](#) for details.

**sign test.** A sign test is used to test the null hypothesis that the median of a distribution is equal to some reference value. A sign test is carried out as a test of binomial proportion with a reference value of 0.5. See [PSS] [power oneproportion](#) and [R] [bitest](#).

**significance level.** In [hypothesis testing](#), the significance level  $\alpha$  is an upper bound for a [probability of a type I error](#). See [PSS] [intro](#) to learn more about the relationship between significance level and the power of a test.

**size of test.** See [significance level](#).

**sphericity assumption.** All differences between levels of the within-subject factor [within-subject](#) factor have the same variances.

**stratified  $2 \times 2$  tables.** Stratified  $2 \times 2$  tables describe the association between a binary independent variable and a binary response variable of interest. The analysis is stratified by a nominal (categorical) variable with  $K$  levels.

**symmetry.** In a  $2 \times 2$  [contingency table](#), symmetry refers to the equality of the off-diagonal elements. For a  $2 \times 2$  table, a test of [marginal homogeneity](#) reduces to a test of symmetry.

**$t$  test.** A  $t$  test is a test for which the sampling distribution of the test statistic is a Student's  $t$  distribution.

A one-sample  $t$  test is used to test whether the mean of a population is equal to a specified value when the variance must also be estimated. The test statistic follows Student's  $t$  distribution with  $N - 1$  degrees of freedom, where  $N$  is the sample size.

A two-sample  $t$  test is used to test whether the means of two populations are equal when the variances of the populations must also be estimated. When the two populations' variances are unequal, a modification to the standard two-sample  $t$  test is used; see [Satterthwaite's  \$t\$  test](#).

**target parameter.** In power and sample-size analysis, the target parameter is the parameter of interest or the parameter in the study about which hypothesis tests are conducted.

**test statistic.** In [hypothesis testing](#), a test statistic is a function of the sample that does not depend on any unknown parameters.

**two-independent-samples test.** See [two-sample test](#).

**two-sample paired test.** See [paired test](#).

**two-sample test.** A two-sample test is used to test whether the parameters of interest of the two independent populations are equal. For example, two-sample means test, two-sample variances, two-sample proportions test, two-sample correlations test.

**two-sided test, two-tailed test.** A two-sided test is a [hypothesis test](#) of a parameter in which the [alternative hypothesis](#) is the complement of the [null hypothesis](#). In the context of a test of a scalar parameter, the alternative hypothesis states that the parameter is less than or greater than the value conjectured under the null hypothesis.

**two-way ANOVA, two-way analysis of variance.** A two-way ANOVA model contains two [factors](#). Also see [PSS] [power twoway](#).

**two-way repeated-measures ANOVA, two-factor ANOVA.** This is a repeated-measures ANOVA model with one within-subject factor and one between-subjects factor. The model can be additive (contain only main effects of the factors) or can contain main effects and an interaction between the two factors. Also see [PSS] [power repeated](#).

**type I error.** The type I error of a test is the error of rejecting the null hypothesis when it is true; see [PSS] [intro](#) for more details.

**type I error probability.** See *probability of a type I error*.

**type I study.** A type I study is a study in which all subjects fail (or experience an event) by the end of the study; that is, no censoring of subjects occurs.

**type II error.** The type II error of a test is the error of not rejecting the null hypothesis when it is false; see [PSS] [intro](#) for more details.

**type II error probability.** See *probability of a type II error*.

**type II study.** A type II study is a study in which there are subjects who do not fail (or do not experience an event) by the end of the study. These subjects are known to be censored.

**unbalanced design.** An unbalanced design indicates an experiment in which the numbers of treated and untreated subjects differ. Also see [PSS] [unbalanced designs](#).

**unequal-allocation design.** See *unbalanced design*.

**upper one-sided test, upper one-tailed test.** An upper one-sided test is a one-sided test of a scalar parameter in which the alternative hypothesis is upper one sided, meaning that the alternative hypothesis states that the parameter is greater than the value conjectured under the null hypothesis. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] [intro](#).

**Wald test.** A Wald test is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. Under the null hypothesis, the constrained model fits the data as well as the full model. The Wald test requires one to fit the full model but does not require one to fit the constrained model. Also see [PSS] [power oneproportion](#) and [R] [test](#).

**withdrawal.** Withdrawal is the process under which subjects withdraw from a study for reasons unrelated to the event of interest. For example, withdrawal occurs if subjects move to a different area or decide to no longer participate in a study. Withdrawal should not be confused with administrative censoring. If subjects withdraw from the study, the information about the outcome those subjects would have experienced at the end of the study, had they completed the study, is unavailable. Also see *loss to follow-up* and *administrative censoring*.

**within-subject design.** This is an experiment that has at least one within-subject factor. See [PSS] [power repeated](#).

**within-subject factor.** This is a factor for which each subject receives several of or all the levels.

**z test.** A  $z$  test is a test for which a potentially asymptotic sampling distribution of the test statistic is a normal distribution. For example, a one-sample  $z$  test of means is used to test whether the mean of a population is equal to a specified value when the variance is assumed to be known. The distribution of its test statistic is normal. See [PSS] [power onemean](#), [PSS] [power twomeans](#), and [PSS] [power pairedmeans](#).

# Subject and author index

See the [combined subject index](#) and the [combined author index](#) in the *Glossary and Index*.