STATA POWER AND SAMPLE-SIZE REFERENCE MANUAL
RELEASE 13

The suggested citation for this software is

StataCorp. 2013. *Stata: Release 13*. Statistical Software. College Station, TX: StataCorp LP.

Contents

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] *Getting Started with Stata for Mac*
 [GSU] *Getting Started with Stata for Unix*
 [GSW] *Getting Started with Stata for Windows*

 [U] *Stata User's Guide*
 [R] *Stata Base Reference Manual*
 [D] *Stata Data Management Reference Manual*
 [G] *Stata Graphics Reference Manual*
 [XT] *Stata Longitudinal-Data/Panel-Data Reference Manual*
 [ME] *Stata Multilevel Mixed-Effects Reference Manual*
 [MI] *Stata Multiple-Imputation Reference Manual*
 [MV] *Stata Multivariate Statistics Reference Manual*
 [PSS] *Stata Power and Sample-Size Reference Manual*
 [P] *Stata Programming Reference Manual*
 [SEM] *Stata Structural Equation Modeling Reference Manual*
 [SVY] *Stata Survey Data Reference Manual*
 [ST] *Stata Survival Analysis and Epidemiological Tables Reference Manual*
 [TS] *Stata Time-Series Reference Manual*
 [TE] *Stata Treatment-Effects Reference Manual:*
   Potential Outcomes/Counterfactual Outcomes
 [I] *Stata Glossary and Index*

 [M] *Mata Reference Manual*

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:

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), and Julious (2010), 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 θ, a target parameter, based on a sample from the population. For simplicity, we consider a simple hypothesis test comparing a population parameter θ 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 θ defined as the probability that the observed sample belongs to the rejection region of a test for a given parameter θ. 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 α, a value between 0 and 1 inclusively. Many tests achieve their significance level—that is, their type-I-error probability equals α, $\Pr(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$—for any parameter in the null space. For other tests, α 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 $H\colon \theta > \theta_0$) or small ($H\colon \theta \leq \theta_0$), whereas a two-sided or nondirectional test asserts that the target parameter is either large or small ($H\colon \theta \neq \theta_0$). 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 α 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, α;
- power, $1 - \beta$;

- a magnitude of an effect of interest or clinically meaningful difference, often expressed as an effect size, δ;
- 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 χ^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 α 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 α 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-α test, for which $\Pr(\text{reject} H_0 | H_0 \text{ is true}) = \alpha$ for any θ in the null space, as opposed to a level-α test, for which $\Pr(\text{reject } H_0 | H_0 \text{ is true}) \leq \alpha$ for any θ 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 μ with 0 from a population with a known standard deviation σ. The null hypothesis is $H_0\colon \mu = 0$, and the alternative hypothesis is $H_a\colon \mu \neq 0$. The test statistic is a sample mean or sample average. It has a normal distribution with mean 0 and standard deviation σ as its null sampling distribution, and it has a normal distribution with mean μ different from 0 and standard deviation σ as its alternative sampling distribution. The overlap between these distributions is determined by the mean difference $\mu - 0 = \mu$ and standard deviation σ. The larger μ 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 μ. The larger the standard deviation σ, the more overlap between the two distributions and the lower the power to detect the difference. Instead of being viewed as a function of μ and σ, 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 δ.

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.

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
```
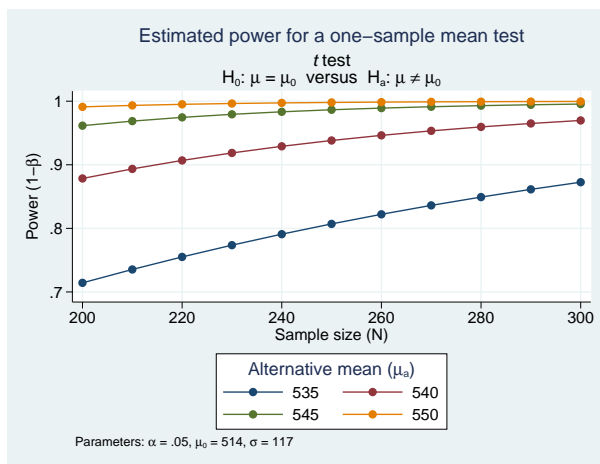

The default graph plots the estimated power on the y axis and the requested sample size on the x axis. A separate curve is plotted for each of the specified alternative means. Power increases as the

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 and binary outcomes; see [PSS] **power** and method-specific entries for more examples. See the `stpower` command in [ST] **stpower** for PSS analysis of survival outcomes. 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.

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., 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., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Marcel 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.

Connor, R. J. 1987. Sample size for testing differences in proportions for the paired-sample design. *Biometrics* 43: 207–211.

Davis, B. R., J. A. Cutler, D. J. Gordon, C. D. Furberg, J. T. Wright, Jr, W. C. Cushman, 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.

Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley.

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 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.

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.

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.

Lenth, R. V. 2001. Some practical guidelines for effective sample size determination. *American Statistician* 55: 187–193.

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.

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., 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.

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.

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.

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.

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**

[ST] **stpower** — Sample size, power, and effect size for survival analysis

Title

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.
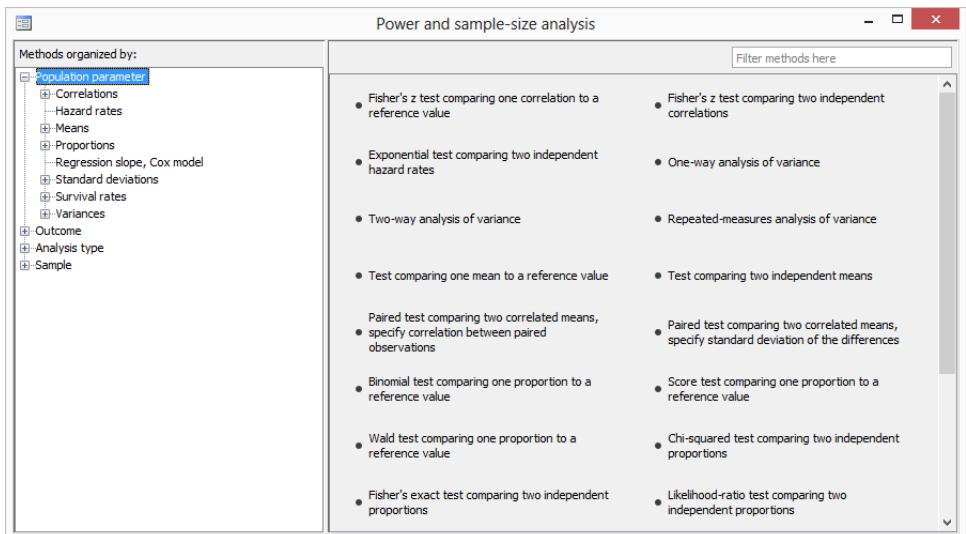
The left pane organizes the methods, and the right pane displays the methods corresponding to the selection in the left pane. On the left, the methods are organized by the type of population parameter, such as mean or proportion; the type of outcome, such as continuous or binary; the type of analysis, such as t test or χ^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.
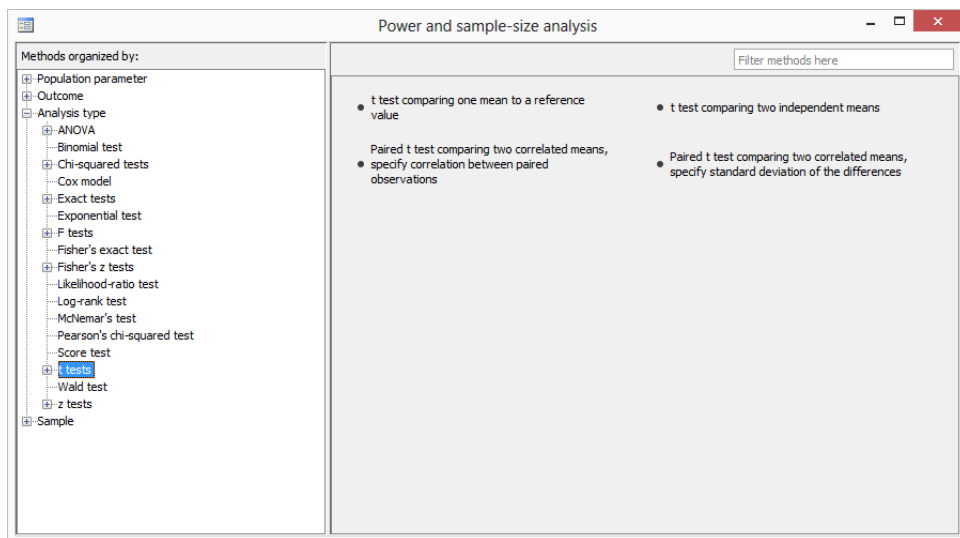
For example, if we are interested in proportions, we can click on **Proportions** within **Population parameter** to see all methods comparing proportions in the right pane.
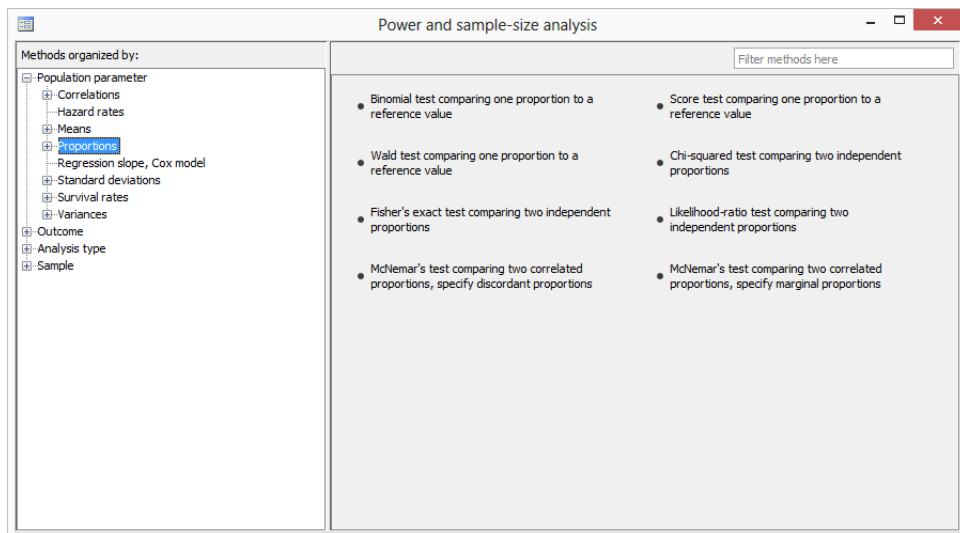

We can expand **Proportions** to further narrow down the choices by clicking on the symbol to the left of **Proportions**.
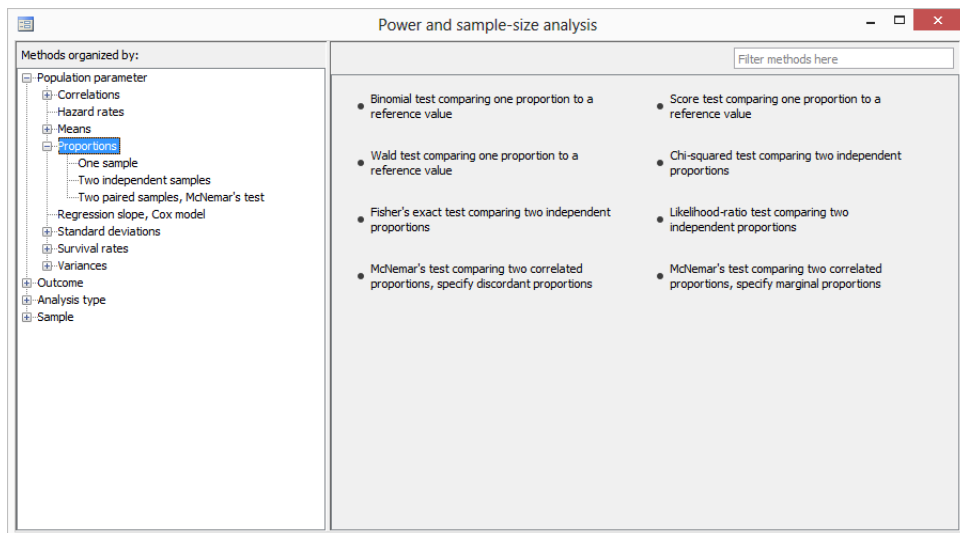
Or we can choose a method by the type of analysis by expanding **Analysis type** and selecting, for example, **t tests**:

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.
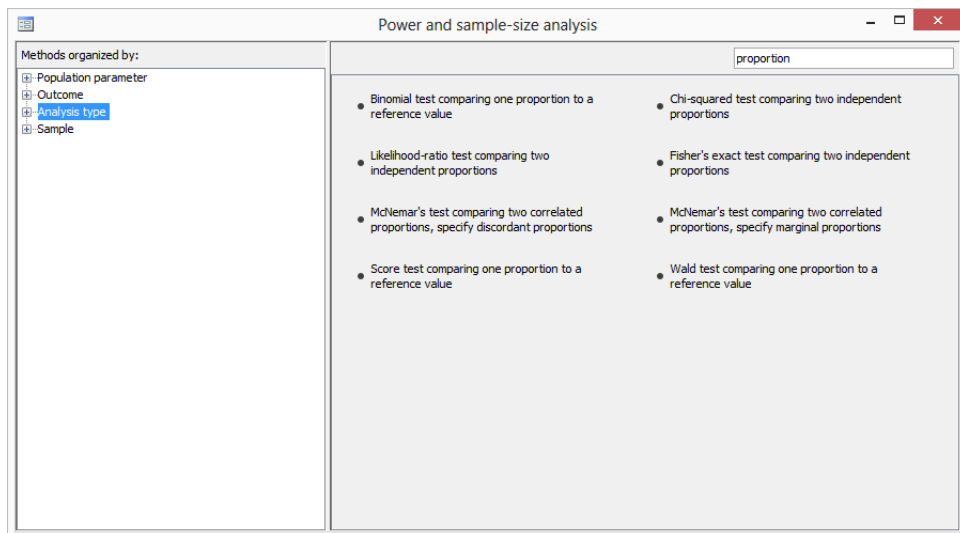
We can specify multiple words in the *Filter* box, and only methods with all the specified words in their titles will appear. For example, if we type "two proportions", only methods with the words "two" and "proportions" in their titles will be shown:
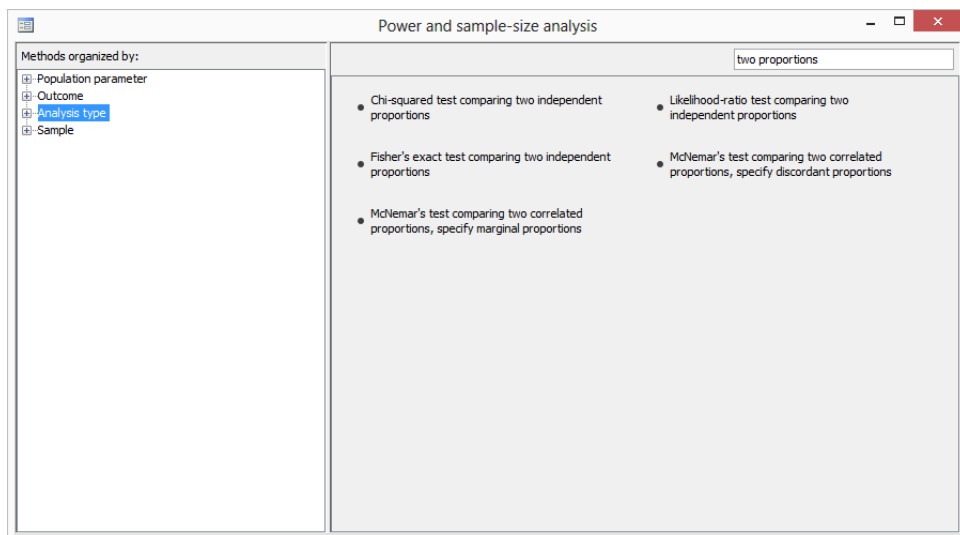
The search is performed within the group of methods selected by the choice in the left pane. In the above example, the search was done within **Analysis type**. When you select one of the top categories in the left pane, the same set of methods appears in the right pane but in the order determined by 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**.
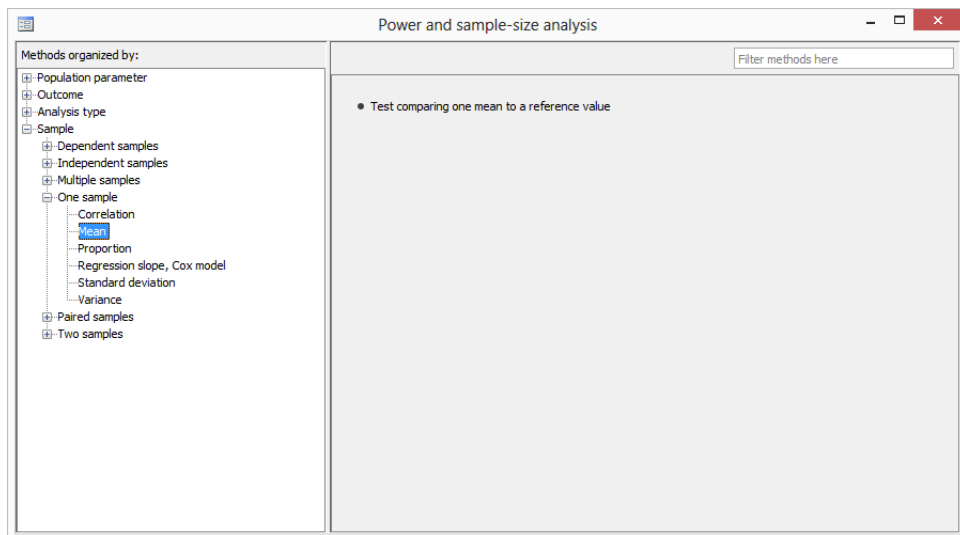
We invoke the dialog box by clicking on the method title in the right pane. The following appears:

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 *numlist*s for sample size and alternative mean in the respective edit boxes:

We also check the *Graph the results* box on the **Graph** tab:
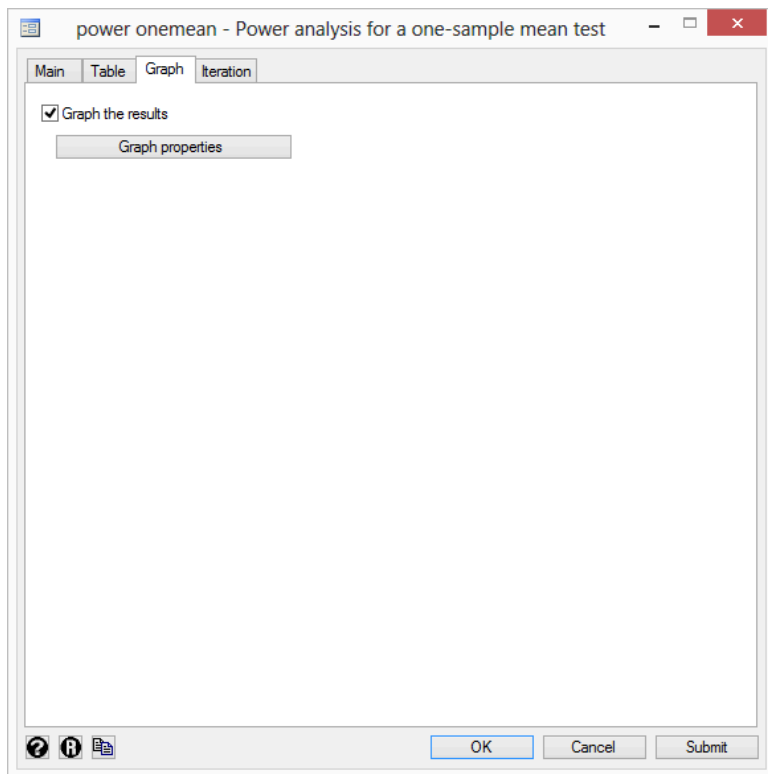
We click on the **Submit** button and obtain the following command and graph:

```
. 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
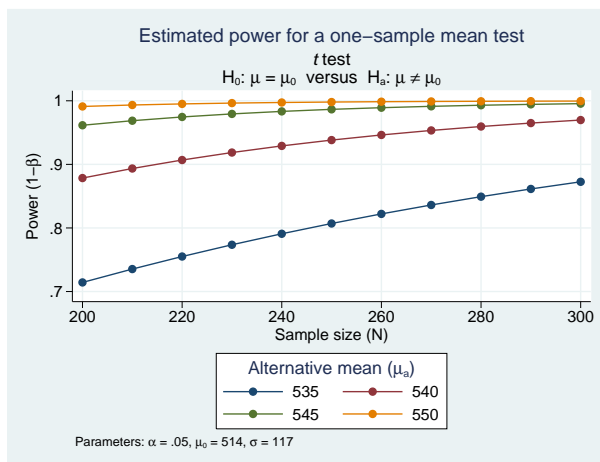

Syntax

Compute sample size

    power *method* ... [, <u>p</u>ower(*numlist*) *power_options* ...]

Compute power

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

Compute effect size and target parameter

    power *method* ..., n(*numlist*) <u>p</u>ower(*numlist*) [*power_options* ...]

method	Description
One sample	
onemean	One-sample mean test (one-sample t test)
oneproportion	One-sample proportion test
onecorrelation	One-sample correlation test
onevariance	One-sample variance test
Two independent samples	
twomeans	Two-sample means test (two-sample t test)
twoproportions	Two-sample proportions test
twocorrelations	Two-sample correlations test
twovariances	Two-sample variances test
Two paired samples	
pairedmeans	Paired-means test (paired t test)
pairedproportions	Paired-proportions test (McNemar's test)
Analysis of variance	
oneway	One-way ANOVA
twoway	Two-way ANOVA
repeated	Repeated-measures ANOVA

power_options	Description
[Main] Main	
* <u>a</u>lpha(*numlist*)	significance level; default is alpha(0.05)
* <u>p</u>ower(*numlist*)	power; default is power(0.8)
* <u>b</u>eta(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	total sample size; required to compute power or effect size
* n1(*numlist*)	sample size of the control group
* n2(*numlist*)	sample size of the experimental group
* <u>nrat</u>io(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
<u>comp</u>ute(n1 | n2)	solve for N_1 given N_2 or for N_2 given N_1
<u>nfrac</u>tional	allow fractional sample sizes
<u>dir</u>ection(<u>u</u>pper|<u>l</u>ower)	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>one</u>sided	one-sided test; default is two sided
<u>par</u>allel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
[Table] Table	
[<u>no</u>] <u>table</u> [(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>sav</u>ing(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph] Graph	
<u>graph</u> [(*graphopts*)]	graph results; see [PSS] **power, graph**
[Iteration] Iteration	
init(#)	initial value of the estimated parameter; default is method specific
<u>iter</u>ate(#)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(#)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(#)	function tolerance; default is ftolerance(1e-12)
[no] log	suppress or display iteration log
[no] dots	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.

Menu

Statistics > Power and sample size

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**).

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 - $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 - $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, N_2/N_1 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. The default is direction(upper), which means that the postulated value of the parameter is larger than the hypothesized value. direction(lower) means that the postulated value is lower 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 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*
> *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.

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 χ^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 χ^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**.

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` $\#_{a1}$, `oratio(#)` ...

or as a relative risk,

. `power twoproportions` $\#_{a1}$, `rrisk()` ...

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 paired-means 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 analysis 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 m_{a1} m_{a2} $\begin{bmatrix} m_{a3} \ \dots \end{bmatrix}$, ...

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` $m_{a1,1}$ $m_{a1,2}$ $\left[\dots\right]$ \ $m_{a2,1}$ $m_{a2,2}$ $\left[\dots\right]$, ...

> `power repeated` $m_{a1,1}$ $m_{a1,2}$ $\left[\dots\right]$ $\left[\backslash m_{a2,1}\ m_{a2,2}\ \left[\dots\right]\right]$, ...

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` m_{a1} m_{a2} m_{a3}, ...

The above command would be equivalent to

> `matrix means = ` (m_{a1}, m_{a2}, m_{a3})

> `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.

◁

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
```
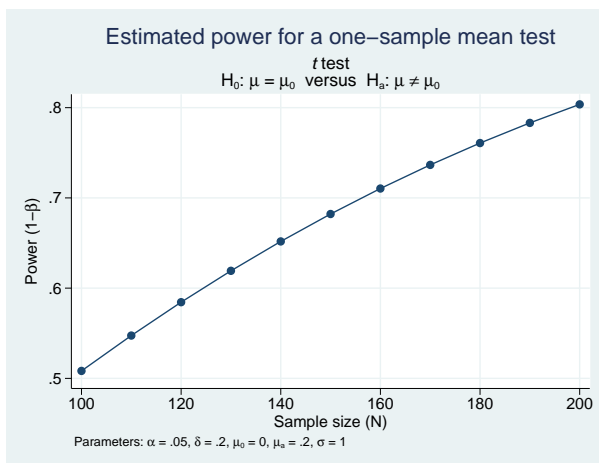

Or we can compute sample size and plot it as a function of the alternative mean when the mean ranges between 0.2 and 1 with a step size of 0.1:

```
. power onemean 0 (0.2(0.1)1), graph
```
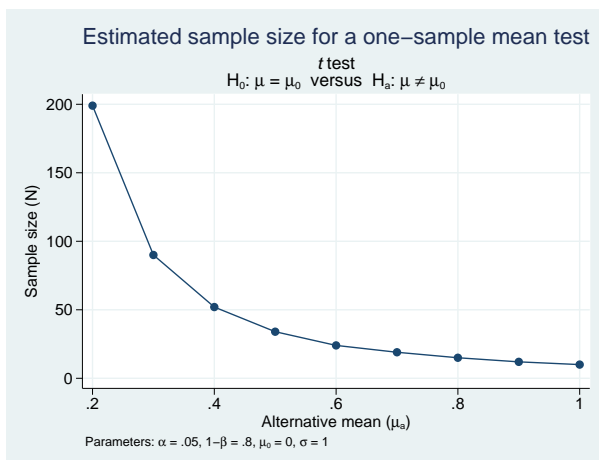
Or we can compute the alternative mean for a given power of 80% and a range of sample-size values between 100 and 200 with a step size of 10, and plot it against the sample size:

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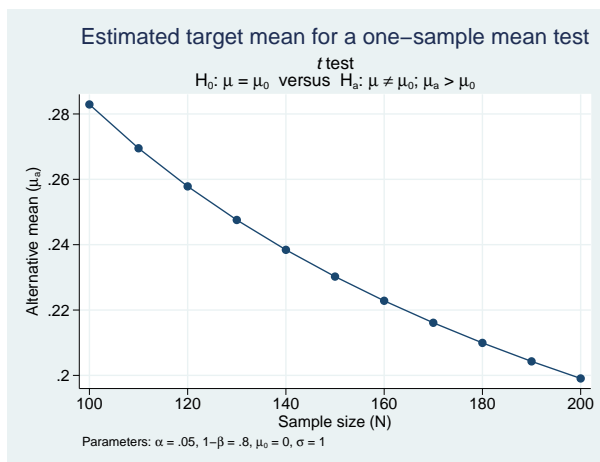

The above graphs are the default graphs produced by `power, graph`. Similarly to tabular output, you can customize graphical output by specifying the `graph()` option.

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("")))
```
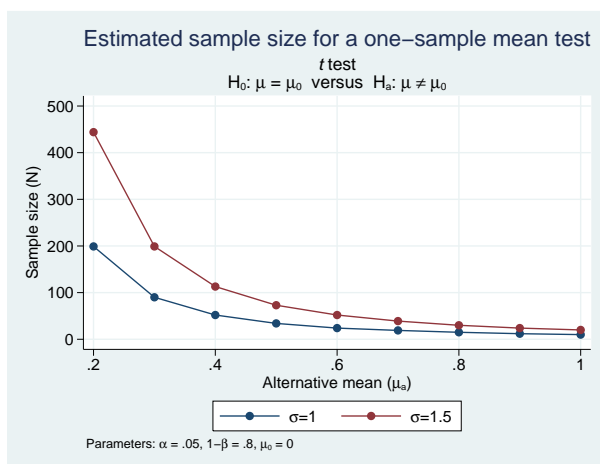

By default, when a graph is produced, the tabular output is suppressed. You can specify the `table` option if you also want to see results in a table.

For more examples, see [PSS] **power, graph**.

Stored results

power stores the following in r():

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

Macros
r(type)	test
r(method)	the name of the specified method
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

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 solvenl() 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.

Reference

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.

Also see

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

[PSS] **Glossary**

[ST] **stpower** — Sample size, power, and effect size for survival analysis

Title

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

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

  $\big[$power$\big]$ ..., graph(*graphopts*) ...

graphopts	Description
[Main]	
y̲dimension(*dimlist* [, *dimopts*])	use *dimlist* to define y axis
x̲dimension(*dimlist* [, *dimopts*])	use *dimlist* to define x axis
plotdimension(*dimlist* [, *dimopts*])	create plots for groups in *dimlist*
bydimension(*dimlist* [, *dimopts*])	create subgraphs for groups in *dimlist*
graphdimension(*dimlist* [, *dimopts*])	create graphs for groups in *dimlist*
horiz̲ontal	swap x and y axes
schemegrid	do not apply default x and y grid lines
name(*name* | *stub* [, replace])	name of graph, or *stub* if multiple graphs
[Labels]	
yregular	place regularly spaced ticks and labels on the y axis
xregular	place regularly spaced ticks and labels on the x axis
yvalues	place ticks and labels on the y axis for each distinct value
xvalues	place ticks and labels on the x axis for each distinct value
collabels(*colspec*)	change default labels for columns
nolabels	label groups with their values, not their labels
allsimplelabels	forgo column label and equal signs in all labels
nosimplelabels	include column label and equal signs in all labels
eqseparator(*string*)	replace equal sign separator with *string*
separator(*string*)	separator for labels when multiple columns are specified in a dimension
noseparator	do not use a separator
format(%*fmt*)	format for converting numeric values to labels
[Plot]	
plotopts(*plot_options*)	affect rendition of all plots
plot#opts(*plot_options*)	affect rendition of #th plot
recast(*plottype*)	plot all plots using *plottype*
[Add plots]	
addplot(*plot*)	add other plots to the generated graph
[Y axis, X axis, Titles, Legend, Overall, By]	
twoway_options	any options documented in [G-3] *twoway_options*
byopts(*byopts*)	how subgraphs are combined, labeled, etc.

dimlist may contain any of the following columns:

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

colspec is

    [*column*](#) "*label*" $\big[$ *column* "*label*" $\big[\ldots \big] \big]$

dimopts	Description
<u>lab</u>els(*lablist*)	list of quoted strings to label each level of the dimension
<u>ela</u>bels(*elablist*)	list of enumerated labels
<u>nol</u>abels	label groups with their values, not their labels
<u>alls</u>implelabels	forgo column name and equal signs in all labels
<u>nos</u>implelabels	include column name and equal signs in all labels
<u>eqsep</u>arator(*string*)	replace equal sign separator with *string* in the dimension
<u>sep</u>arator(*string*)	separator for labels when multiple columns are specified in the dimension
<u>nosep</u>arator	do not use a separator
<u>f</u>ormat(%*fmt*)	format for converting numeric values to labels

where *lablist* is defined as

    "*label*" $\big[$ "*label*" $\big[\ldots \big] \big]$

elablist is defined as

    # "*label*" $\big[$ *#* "*label*" $\big[\ldots \big] \big]$

and the *#*s are the levels of the dimension.

plot_options	Description
[*marker_options*](#)	change look of markers (color, size, etc.)
[*marker_label_options*](#)	add marker labels; change look or position
[*cline_options*](#)	change look of the line

Menu

Statistics > Power and sample size

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.

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 *stub*1, *stub*2, ... 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 graphs' 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 `label()` 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
α	significance level
β	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
δ	effect size
method_symbols	symbols specific to the *method* specified with power

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 0.05-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
```
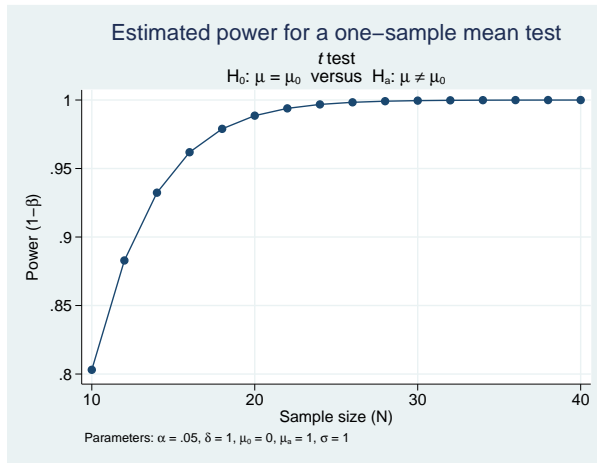

Figure 1.

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 α is 0.05, effect size δ (the standardized mean difference $(\mu_a - \mu_0)/\sigma$ for a one-sample t test) is 1, null mean μ_0 is 0, alternative mean μ_a is 1, and standard deviation σ 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
```
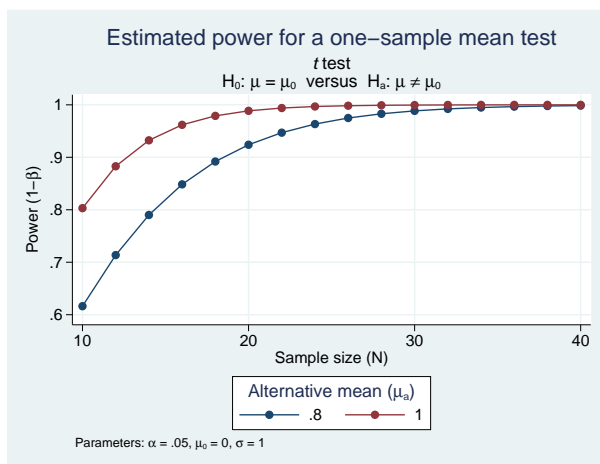

Figure 2.

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 s2gcolor; 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
```
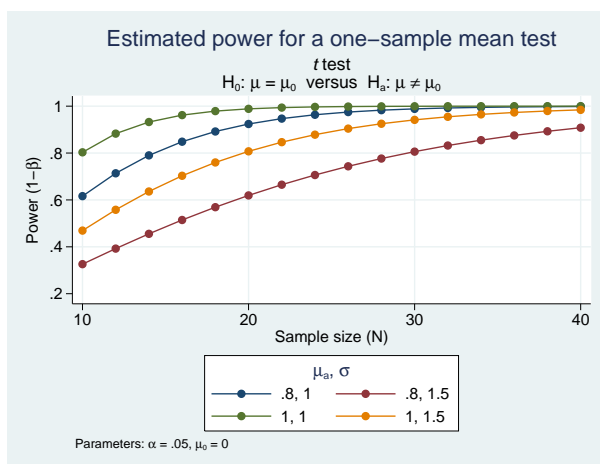

Figure 3.

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
```
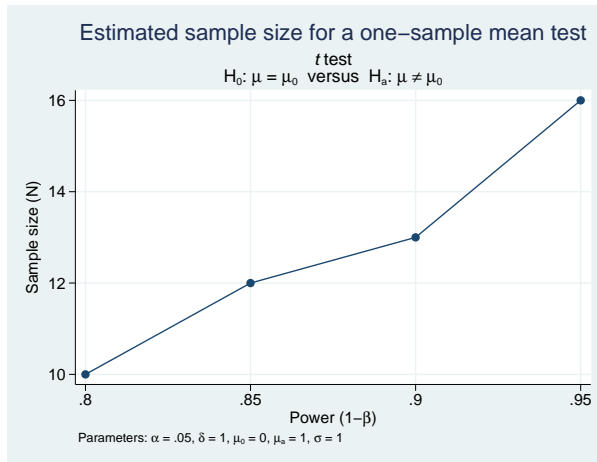

Figure 4.

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
```
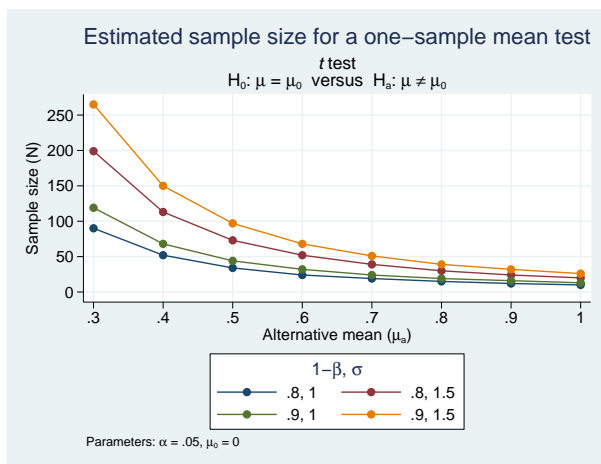

Figure 5.

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.

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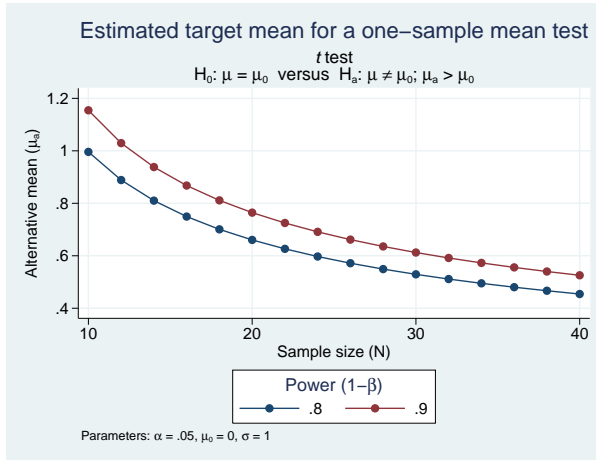
The larger the sample size and the smaller the power, the smaller the values of the alternative mean that can be detected by the test. The default x axis is the sample size whenever the sample size varies.

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, β, 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))
```
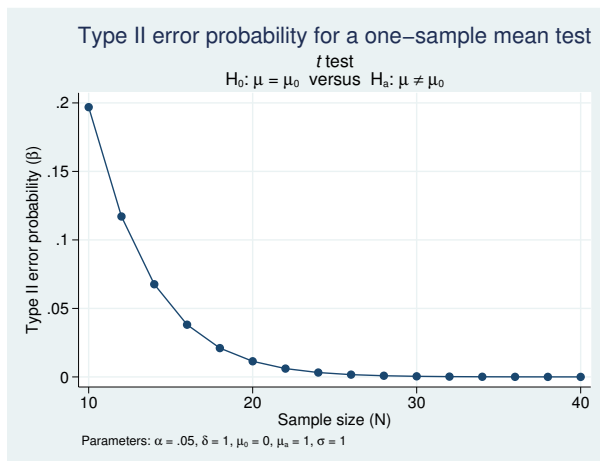

Figure 7.

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

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))
```
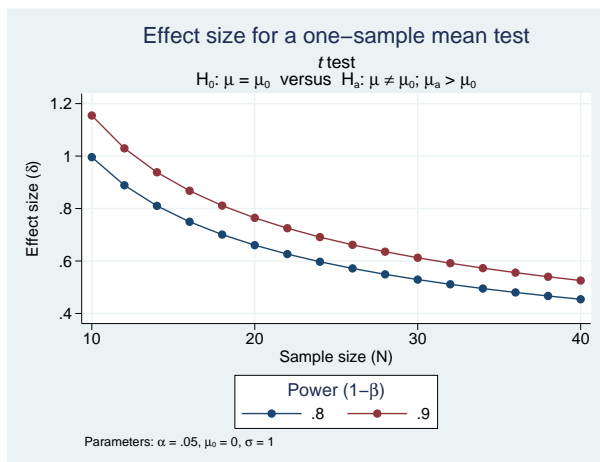

Figure 8.

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))
```
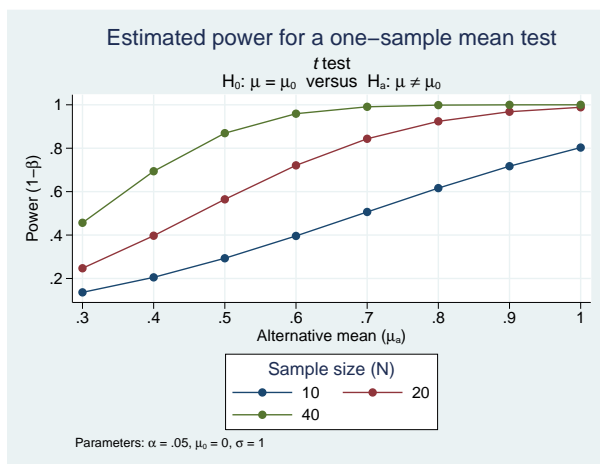

Figure 9.

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))
```
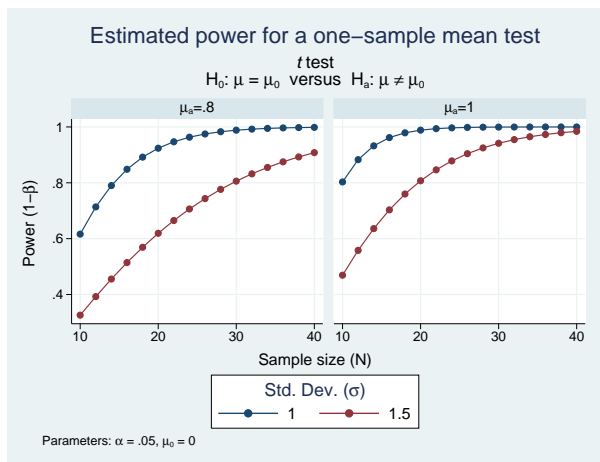

Figure 10.

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 α. 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))
```
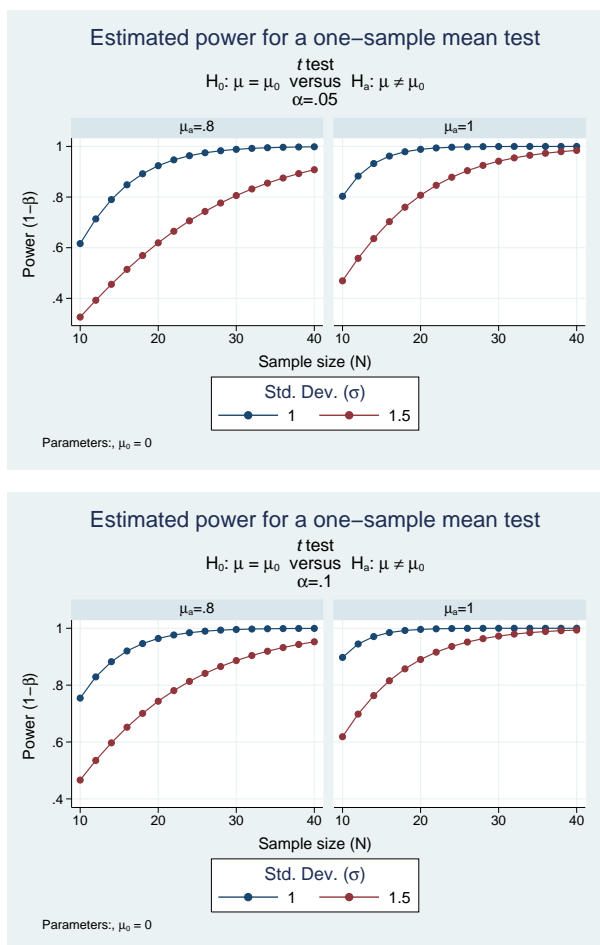

Figure 11.

◁

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)))
```
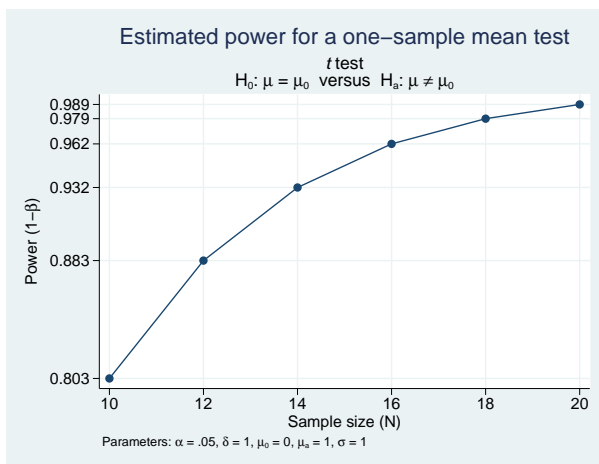

Figure 12.

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))
```
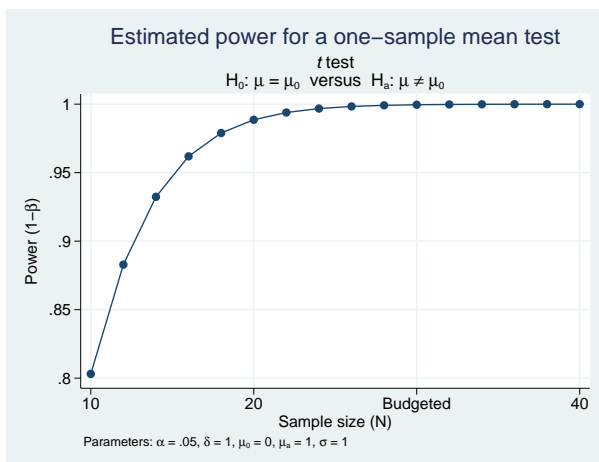

Figure 13.

We can use `ytitle()` and `xtitle()` to change the axis titles.

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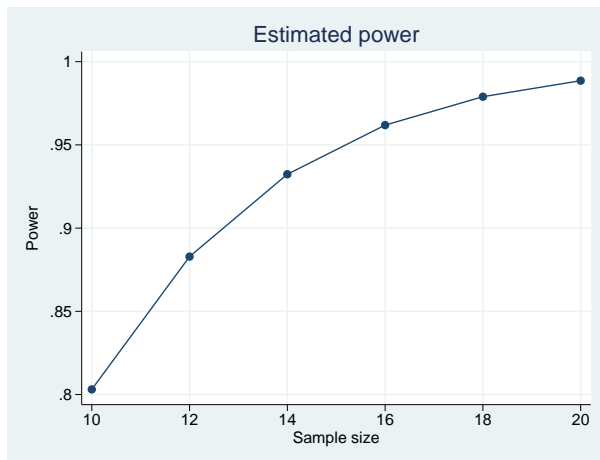

Figure 14.

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"))
```
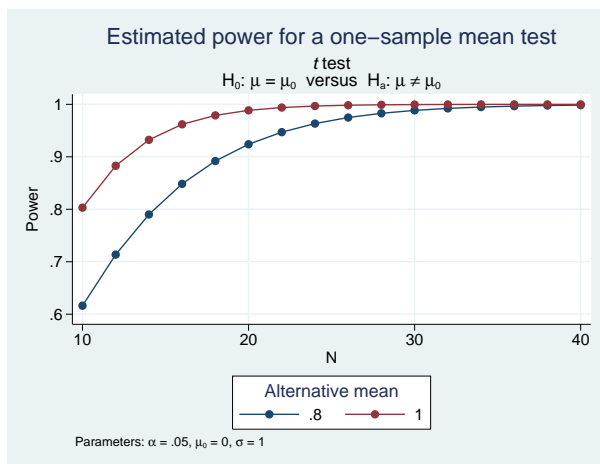

Figure 15.

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(" = "))
```
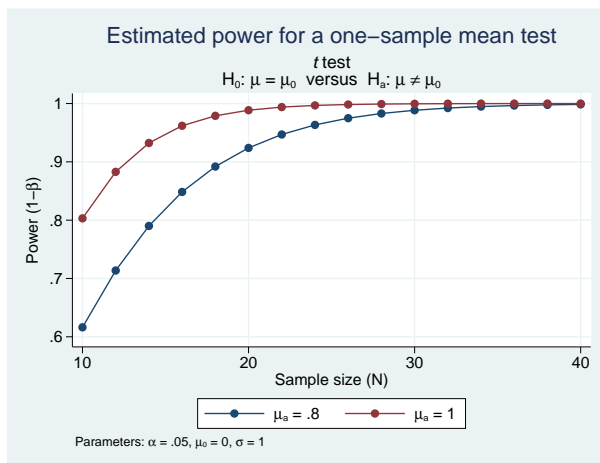

Figure 16.

◁

▷ 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)))
```
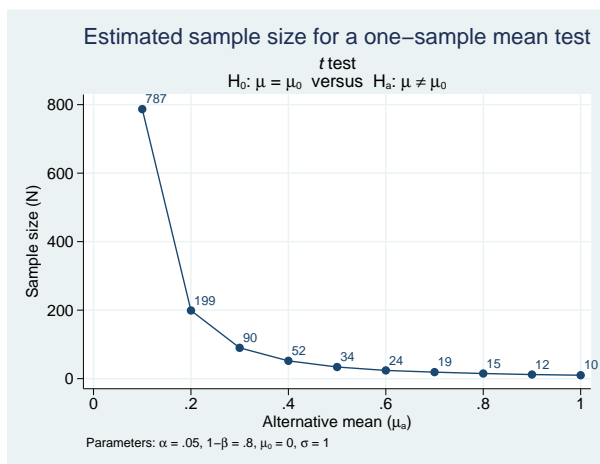

Figure 17.

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)))
```

Estimated power for a one–sample mean test
t test
$H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$
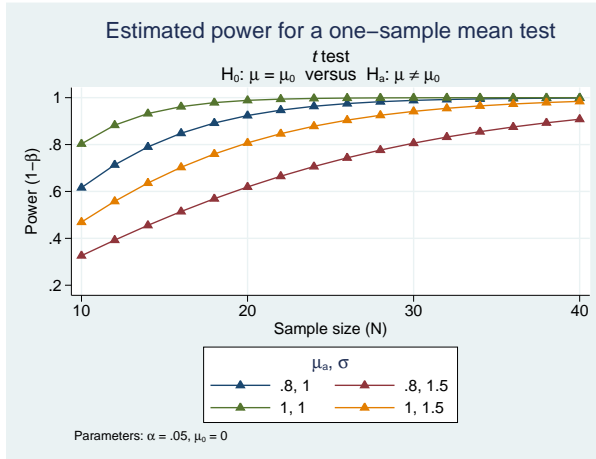
Figure 18.

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)))
```

Estimated power for a one–sample mean test
t test
$H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$
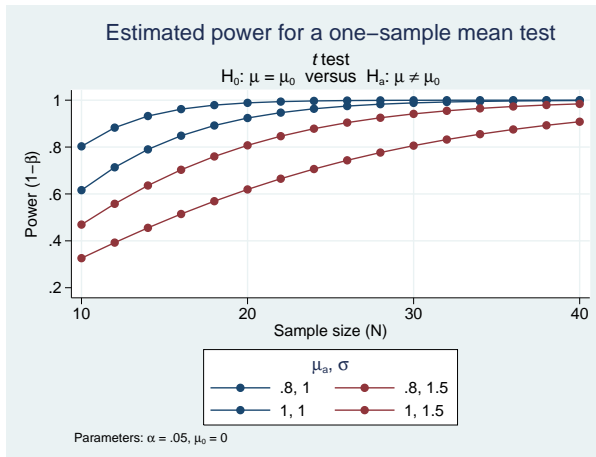
Figure 19.

▷ 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))
```
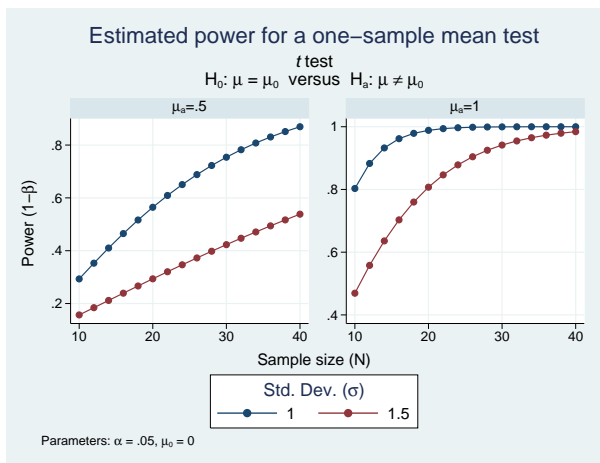

Figure 20.

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("")))
```
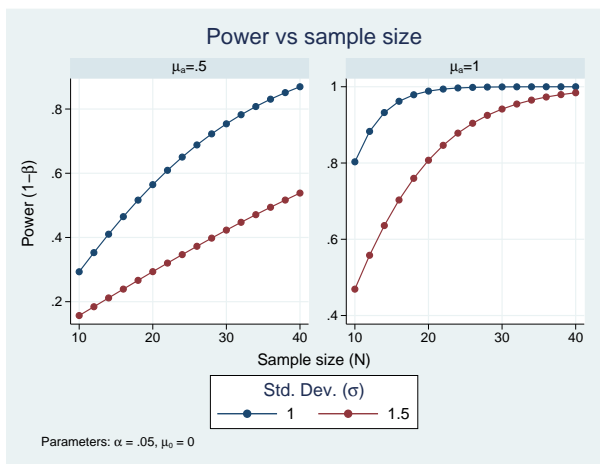

Figure 21.

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(""))
```
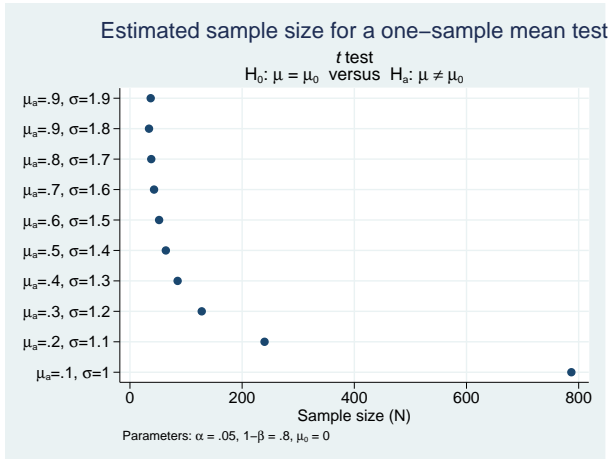

Figure 22.

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

Syntax   Menu   Description   Suboptions
Remarks and examples Stored results Also see

Syntax

Produce default table

 power ..., <u>table</u> ...

Suppress table

 power ..., <u>notable</u> ...

Produce custom table

 power ..., <u>table</u>([*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).

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

noheader and continue are not shown in the dialog box.

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

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.

Menu

Statistics > Power and sample size

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**.

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(*fmtspec*) specifies column formats. The default is %7.0g for integer-valued columns and %7.4g for real-valued columns. *fmtspec* 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 0.05-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

Syntax

Compute sample size

> power onemean m_0 m_a $\left[\,,\ \underline{\text{power}}(\textit{numlist})\ \textit{options}\,\right]$

Compute power

> power onemean m_0 m_a , n(*numlist*) $\left[\,\textit{options}\,\right]$

Compute effect size and target mean

> power onemean m_0 , n(*numlist*) $\underline{\text{power}}$(*numlist*) $\left[\,\textit{options}\,\right]$

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**).

options	Description
Main	
* <u>a</u>lpha(*numlist*)	significance level; default is `alpha(0.05)`
* <u>p</u>ower(*numlist*)	power; default is `power(0.8)`
* <u>b</u>eta(*numlist*)	probability of type II error; default is `beta(0.2)`
* n(*numlist*)	sample size; required to compute power or effect size
<u>nf</u>ractional	allow fractional sample size
* <u>d</u>iff(*numlist*)	difference between the alternative mean and the null mean, $m_a - m_0$; specify instead of the alternative mean m_a
* sd(*numlist*)	standard deviation; default is `sd(1)`
<u>k</u>nownsd	request computation assuming a known standard deviation; default is to assume an unknown standard deviation
* fpc(*numlist*)	finite population correction (FPC) as a sampling rate or as a population size
<u>dir</u>ection(<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>one</u>sided	one-sided test; default is two sided
<u>par</u>allel	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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>sav</u>ing(*filename* [, replace])	save the table data to *filename*; use `replace` to overwrite existing *filename*
Graph	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**
Iteration	
init(*#*)	initial value for sample size or mean; default is to use normal approximation
<u>iter</u>ate(*#*)	maximum number of iterations; default is `iterate(500)`
<u>tol</u>erance(*#*)	parameter tolerance; default is `tolerance(1e-12)`
<u>ftol</u>erance(*#*)	function tolerance; default is `ftolerance(1e-12)`
[no]log	suppress or display iteration log
[no]dots	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	number of subjects	N
delta	effect size	δ
m0	null mean	μ_0
ma	alternative mean	μ_a
diff	difference between the alternative and null means	$\mu_a - \mu_0$
sd	standard deviation	σ
fpc	FPC as population size	N_{pop}
	FPC as sampling rate	γ
target	target parameter; synonym for ma	
_all	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.

Menu

Statistics > Power and sample size

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.

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(*tablespec*), 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*

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 0.05-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 δ 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/r13/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: $\mu_{\text{SAT}} = 600$ versus a two-sided alternative H_a: $\mu_{\text{SAT}} \neq 600$ 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.

◁

Stored results

power onemean stores the following in r():

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

Macros
r(type)	test
r(method)	onemean
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 $x_1, \ldots, x_n$ be a sequence of n independent and identically distributed random variables drawn from a normal population with mean μ and variance σ^2. Let

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

denote the sample mean and the sample variance, respectively. Let μ_0 and μ_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).

Known standard deviation

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

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

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

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

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):

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

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

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

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 α.

Unknown standard deviation

In the case of an unknown standard deviation, an unbiased estimator s is used in place of σ in the definition of a z test statistic. The sampling distribution of the test statistic $t = \sqrt{n}(\overline{x} - \mu_0)/s$ 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 αth 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

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

where $T_{n-1,\lambda}\left(\cdot\right)$ 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 δ, 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 σ_{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: Marcel 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)

Title

<div style="border:1px solid">

power twomeans — Power analysis for a two-sample means test

</div>

Syntax

Compute sample size

> power twomeans m_1 m_2 $\left[\right.$, <u>power</u>(*numlist*) *options*$\left.\right]$

Compute power

> power twomeans m_1 m_2 , n(*numlist*) $\left[\right.$*options*$\left.\right]$

Compute effect size and experimental-group mean

> power twomeans m_1 , n(*numlist*) <u>power</u>(*numlist*) $\left[\right.$*options*$\left.\right]$

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**).

options	Description
[Main]	
*<u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
*<u>power</u>(*numlist*)	power; default is power(0.8)
*<u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	total sample size; required to compute power or effect size
*n1(*numlist*)	sample size of the control group
*n2(*numlist*)	sample size of the experimental group
*<u>nratio</u>(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
<u>compute</u>(n1|n2)	solve for N_1 given N_2 or for N_2 given N_1
<u>nfractional</u>	allow fractional sample sizes
*<u>diff</u>(*numlist*)	difference between the experimental-group mean and the control-group mean, $m_2 - m_1$; specify instead of the experimental-group mean m_2
*sd(*numlist*)	common standard deviation of the control and the experimental groups assuming equal standard deviations in both groups; default is sd(1)
*sd1(*numlist*)	standard deviation of the control group; requires sd2()
*sd2(*numlist*)	standard deviation of the experimental group; requires sd1()
<u>knownsds</u>	request computation assuming known standard deviations for both groups; default is to assume unknown standard deviations
<u>direction</u>(upper|lower)	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>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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph]	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration

<u>init</u>(#)	initial value for sample sizes or experimental-group mean
<u>iterate</u>(#)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(#)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(#)	function tolerance; default is ftolerance(1e-12)
[no] log	suppress or display iteration log
[no] dots	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
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	δ
m1	control-group mean	μ_1
m2	experimental-group mean	μ_2
diff	difference between the control-group mean and the experimental-group mean	$\mu_2 - \mu_1$
sd	common standard deviation	σ
sd1	control-group standard deviation	σ_1
sd2	experimental-group standard deviation	σ_2
target	target parameter; synonym for m2	
_all	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.

Menu

Statistics > Power and sample size

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.

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(*tablespec*), 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:

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 μ_1 and μ_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 δ 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

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N1)	sample size of the control group
r(N2)	sample size of the experimental group
r(nratio)	ratio of sample sizes, N_2/N_1
r(nratio_a)	actual ratio of sample sizes
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(m1)	control-group mean
r(m2)	experimental-group mean
r(diff)	difference between the experimental- and control-group means
r(sd)	common standard deviation of the control and experimental groups
r(sd1)	standard deviation of the control group
r(sd2)	standard deviation of the experimental group
r(knownsds)	1 if option knownsds is specified; 0 otherwise
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for sample sizes or experimental-group mean
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros
    r(type)   test
    r(method)   twomeans
    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 with n_1 subjects in the control group and n_2 subjects in the experimental group. Let $x_{11}, \ldots, x_{1n_1}$ be a random sample of size n_1 from a normal population with mean μ_1 and variance σ_1^2. Let $x_{21}, \ldots, x_{2n_2}$ be a random sample of size n_2 from a normal population with mean μ_2 and variance σ_2^2. Let effect size δ 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

$$
\overline{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \qquad \text{and} \qquad s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \overline{x}_1)^2
$$

$$
\overline{x}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i} \qquad \text{and} \qquad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \overline{x}_2)^2
$$

where $\overline{x}_j$ 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 σ_D denote the standard deviation of the difference between the two sample means. The test statistic of the form

$$
\text{TS} = \frac{(\overline{x}_2 - \overline{x}_1) - (\mu_2 - \mu_1)}{\sigma_D} \tag{1}
$$

is used in each of the four cases described above. Each case, however, determines the functional form of σ_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).

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

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

$$\pi = \begin{cases} \Phi\left(\frac{\delta}{\sigma_D} - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\ \Phi\left(-\frac{\delta}{\sigma_D} - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\ \Phi\left(\frac{\delta}{\sigma_D} - z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta}{\sigma_D} - z_{1-\alpha/2}\right) & \text{for a two-sided test} \end{cases} \tag{2}$$

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:

$$n_1 = \left(\frac{z_{1-\alpha} - z_\beta}{\mu_2 - \mu_1}\right)^2 \left(\sigma_1^2 + \frac{\sigma_2^2}{R}\right) \tag{3}$$

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:

$$n_1 = \frac{\sigma_1^2}{\left(\frac{\mu_2 - \mu_1}{z_{1-\alpha} - z_\beta}\right)^2 - \frac{\sigma_2^2}{n_2}} \tag{4}$$

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 α 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 μ_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

$$t = \frac{(\overline{x}_2 - \overline{x}_1) - (\mu_2 - \mu_1)}{s_D}$$

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 ν degrees of freedom. Parameters ν and s_D are defined below, separately for the case of equal and unequal standard deviations.

Let $t_{\nu,\alpha}$ denote the αth quantile of a Student's t distribution with ν degrees of freedom. Under the alternative hypothesis, the test statistic follows a noncentral Student's t distribution with ν degrees of freedom and noncentrality parameter λ.

The power is computed from the following equations:

$$\pi = \begin{cases} 1 - T_{\nu,\lambda}\left(t_{\nu,1-\alpha}\right) & \text{for an upper one-sided test} \\ T_{\nu,\lambda}\left(-t_{\nu,1-\alpha}\right) & \text{for a lower one-sided test} \\ 1 - T_{\nu,\lambda}\left(t_{\nu,1-\alpha/2}\right) + T_{\nu,\lambda}\left(-t_{\nu,1-\alpha/2}\right) & \text{for a two-sided test} \end{cases} \quad (5)$$

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 μ_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 ν 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 $s_p = \left\{ \sum_{i=1}^{n_1} (x_{1i} - \overline{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2i} - \overline{x}_2)^2 \right\} / (n_1 + n_2 - 2)$ is the pooled-sample standard deviation.

The degrees of freedom ν 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: Marcel 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

Syntax

Compute sample size

   power <u>pairedmeans</u> m_{a1} m_{a2} , *corrspec* [<u>power</u>(*numlist*) *options*]

Compute power

   power <u>pairedmeans</u> m_{a1} m_{a2} , *corrspec* n(*numlist*) [*options*]

Compute effect size and target mean difference

   power <u>pairedmeans</u> [m_{a1}] , *corrspec* n(*numlist*) <u>power</u>(*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**).

options	Description
[Main] **Main**	
* <u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
* <u>power</u>(*numlist*)	power; default is power(0.8)
* <u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample size
* <u>nulld</u>iff(*numlist*)	null difference, the difference between the posttreatment mean and the pretreatment mean under the null hypothesis; default is nulldiff(0)
* <u>altd</u>iff(*numlist*)	alternative difference $d_a = m_{a2} - m_{a1}$, the difference between the posttreatment mean and the pretreatment mean under the alternative hypothesis
* sddiff(*numlist*)	standard deviation σ_d of the differences; may not be combined with corr()
* corr(*numlist*)	correlation between paired observations; required unless sddiff() is specified
* sd(*numlist*)	common standard deviation; default is sd(1) and requires corr()
* sd1(*numlist*)	standard deviation of the pretreatment group; requires corr()
* sd2(*numlist*)	standard deviation of the posttreatment group; requires corr()
<u>knownsd</u>	request computation assuming a known standard deviation σ_d; default is to assume an unknown standard deviation
* fpc(*numlist*)	finite population correction (FPC) as a sampling rate or population size
<u>dir</u>ection(<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>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] **Table**	
[<u>no</u>]<u>table</u>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph] **Graph**	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration

init(#)	initial value for sample size or mean difference; default is to use normal approximation
iterate(#)	maximum number of iterations; default is iterate(500)
tolerance(#)	parameter tolerance; default is tolerance(1e-12)
ftolerance(#)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	suppress or display iterations as dots
notitle	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	number of subjects	N
delta	effect size	δ
d0	null mean difference	d_0
da	alternative mean difference	d_a
ma1	alternative pretreatment mean	μ_{a1}
ma2	alternative posttreatment mean	μ_{a2}
sd_d	standard deviation of the differences	σ_d
sd	common standard deviation	σ
sd1	standard deviation of the pretreatment group	σ_1
sd2	standard deviation of the posttreatment group	σ_2
corr	correlation between paired observations	ρ
fpc	FPC as a population size	N_{pop}
	FPC as a sampling rate	γ
target	target parameter; synonym for da	
_all	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.

Menu

Statistics > Power and sample size

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.

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 σ_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 σ_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(tablespec), 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*

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 μ_2 and pretreatment mean μ_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 σ_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 δ 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 `ttest` 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 `ttest` command to do this. We demonstrate the use of this command using the fictional `bpwide` dataset; see [R] **ttest** 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/r13/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 = d0  versus  Ha: d != d0
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.

◁

Stored results

power pairedmeans stores the following in r():

Scalars

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	sample size
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(d0)	difference between the posttreatment and pretreatment means under the null hypothesis
r(da)	difference between the posttreatment and pretreatment means under the alternative hypothesis
r(ma1)	pretreatment mean under the alternative hypothesis
r(ma2)	posttreatment mean under the alternative hypothesis
r(sd_d)	standard deviation of the differences
r(corr)	correlation between paired observations
r(sd1)	standard deviation of the pretreatment group
r(sd2)	standard deviation of the posttreatment group
r(sd)	common standard deviation
r(knownsd)	1 if option knownsd is specified; 0 otherwise
r(fpc)	finite population correction
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for the sample size or target mean difference
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros

r(type)	test
r(method)	pairedmeans
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 a sequence of n paired observations denoted by X_{ij} for $i = 1, \ldots, n$ 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 μ_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 $\overline{d} = \sum_{i=1}^{n} D_i / n$ 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 σ_1 and σ_2 are the pretreatment and posttreatment group standard deviations, respectively, and ρ 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: Marcel 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

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**).

options	Description
test(*test*)	specify the type of test; default is test(score)

Main

*alpha(*numlist*)	significance level; default is alpha(0.05)
*power(*numlist*)	power; default is power(0.8)
*beta(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	sample size; required to compute power or effect size
nfractional	allow fractional sample size
*diff(*numlist*)	difference between the alternative proportion and the null proportion, $p_a - p_0$; specify instead of the alternative proportion p_a
critvalues	show critical values for the binomial test
continuity	apply continuity correction to the normal approximation of the discrete distribution
direction(upper|lower)	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
onesided	one-sided test; default is two sided
parallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)

Table

[no]table[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
saving(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

Graph

graph[(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration

init(*#*)	initial value for the sample size or proportion
iterate(*#*)	maximum number of iterations; default is iterate(500)
tolerance(*#*)	parameter tolerance; default is tolerance(1e-12)
ftolerance(*#*)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	suppress or display iterations as dots
notitle	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.

test	Description
score	score test; the default
wald	Wald test
binomial	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).

column	Description	Symbol
alpha	significance level	α
alpha_a	actual (observed) significance level	α_a
power	power	$1 - \beta$
beta	type II error probability	β
N	number of subjects	N
delta	effect size	δ
p0	null proportion	p_0
pa	alternative proportion	p_a
diff	difference between the alternative and null proportions	$p_a - p_0$
C_l	lower critical value	C_l
C_u	upper critical value	C_u
target	target parameter; synonym for pa	
_all	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.

Menu

Statistics > Power and sample size

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.

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(*tablespec*), 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

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 bee 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(yline(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
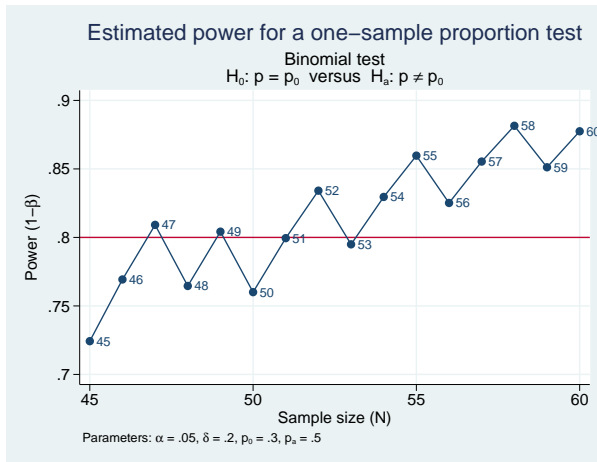
Figure 1. Saw-toothed power function

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 δ 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/r13/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 0.05-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.

◁

Stored results

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

Scalars

r(alpha)	significance level
r(alpha_a)	actual significance level of the binomial method
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	sample size
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(p0)	proportion under the null hypothesis
r(pa)	proportion under the alternative hypothesis
r(diff)	difference between the alternative and null proportions
r(C_l)	lower critical value of the binomial distribution
r(C_u)	upper critical value of the binomial distribution
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for the sample size or proportion
r(continuity)	1 if continuity correction is used; 0 otherwise
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros

r(type)	test
r(method)	oneproportion
r(test)	score, wald, or binomial
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 $x_1, \ldots, x_n$ 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

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \text{and} \qquad \text{se}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

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).

Large-sample normal approximation

For a large sample, the distribution of the sample proportion $\widehat{p}$ 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 $z = (\widehat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$ and the Wald test statistic $z = (\widehat{p} - p_0)/\sqrt{\widehat{p}(1-\widehat{p})/n}$. 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 α be the significance level, β be the probability of a type II error, and $z_{1-\alpha}$ and z_β be the $(1-\alpha)$th and the βth 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

$$
\pi = \begin{cases}
\Phi\left(\dfrac{\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha}\eta\right) & \text{upper one sided} \\[2ex]
\Phi\left(\dfrac{-\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha}\eta\right) & \text{lower one sided} \\[2ex]
\Phi\left(\dfrac{\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha/2}\eta\right) + \Phi\left(\dfrac{-\sqrt{n}(p_a-p_0)-c}{\sqrt{p_a(1-p_a)}} - z_{1-\alpha/2}\eta\right) & \text{two sided}
\end{cases} \quad (1)
$$

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 η 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

$$H_0: p = p_0 \qquad \text{versus} \qquad H_a: p > p_0$$

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 α.

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

$$2 \times \min \left\{ P(X \geq k; n, p_0), P(X \leq k; n, p_0) \right\}$$

For a one-sided test, the power of the test is computed from the following nonlinear equation:

$$\pi = 1 - \beta = \sum_{k=0}^{n} \binom{n}{k} p_a^k (1 - p_a)^{n-k} I \left\{ P(X \geq k; n, p_0) \leq \alpha \right\}$$

Power for a two-sided test can be obtained by replacing the indicator function above with $I\left[2 \times \min \left\{ P(X \geq k; n, p_0), P(X \leq k; n, p_0) \right\} \leq \alpha \right]$.

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: Marcel 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

Title

> **power twoproportions —** Power analysis for a two-sample proportions test

Syntax

Compute sample size

> power twoproportions p_1 p_2 $\left[\,,\ \text{power}(\textit{numlist})\ \textit{options}\,\right]$

Compute power

> power twoproportions p_1 p_2 , n(*numlist*) $\left[\textit{options}\right]$

Compute effect size and experimental-group proportion

> power twoproportions p_1 , n(*numlist*) power(*numlist*) $\left[\textit{options}\right]$

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**).

options	Description
test(*test*)	specify the type of test; default is test(chi2)

[Main]

* <u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
* <u>power</u>(*numlist*)	power; default is power(0.8)
* <u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	total sample size; required to compute power or effect size
* n1(*numlist*)	sample size of the control group
* n2(*numlist*)	sample size of the experimental group
* <u>nrat</u>io(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
<u>compute</u>(n1 | n2)	solve for N_1 given N_2 or for N_2 given N_1
<u>nfrac</u>tional	allow fractional sample sizes
* diff(*numlist*)	difference between the experimental-group and control-group proportions, $p_2 - p_1$; specify instead of the experimental-group proportion p_2
* <u>rat</u>io(*numlist*)	ratio of the experimental-group proportion to the control-group proportion, p_2/p_1; specify instead of the experimental-group proportion p_2
* <u>rd</u>iff(*numlist*)	risk difference, $p_2 - p_1$; synonym for diff()
* <u>rr</u>isk(*numlist*)	relative risk, p_2/p_1; synonym for ratio()
* <u>or</u>atio(*numlist*)	odds ratio, $\{p_2(1-p_1)\}/\{p_1(1-p_2)\}$
<u>eff</u>ect(*effect*)	specify the type of effect to display; default is effect(diff)
<u>contin</u>uity	apply continuity correction to the normal approximation of the discrete distribution
<u>dir</u>ection(<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>onesid</u>ed	one-sided test; default is two sided
<u>par</u>allel	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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename*[, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

[Graph]

<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration

<u>init</u>(#)	initial value for sample sizes or experimental-group proportion
<u>iterate</u>(#)	maximum number of iterations; default is `iterate(500)`
<u>tol</u>erance(#)	parameter tolerance; default is `tolerance(1e-12)`
<u>ftol</u>erance(#)	function tolerance; default is `ftolerance(1e-12)`
[no]log	suppress or display iteration log
[no]dots	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.

test	Description
chi2	Pearson's chi-squared test; the default
lrchi2	likelihood-ratio test
fisher	Fisher–Irwin's exact conditional test

`test()` does not appear in the dialog box. The dialog box selected is determined by the `test()` specification.

effect	Description
diff	difference between proportions, $p_2 - p_1$; the default
ratio	ratio of proportions, p_2/p_1
<u>rd</u>iff	risk difference, $p_2 - p_1$
<u>rr</u>isk	relative risk, p_2/p_1
<u>or</u>atio	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).

column	Description	Symbol
alpha	significance level	α
alpha_a	observed significance level	α_a
power	power	$1 - \beta$
beta	type II error probability	β
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	δ
p1	control-group proportion	p_1
p2	experimental-group proportion	p_2
diff	difference between the experimental-group proportion and the control-group proportion	$p_2 - p_1$
ratio	ratio of the experimental-group proportion to the control-group proportion	p_2/p_1
rdiff	risk difference	$p_2 - p_1$
rrisk	relative risk	p_2/p_1
oratio	odds ratio	θ
target	target parameter; synonym for p2	
_all	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.

Menu

Statistics > Power and sample size

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.

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 χ^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.

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(*tablespec*), notable; see [PSS] **power, table**.

saving(); see [PSS] **power**.

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

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*
> *Computing sample size*
> *Computing power*
> *Computing effect size and experimental-group proportion*
> *Testing a hypothesis about two independent proportions*

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 χ^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 χ^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` p_1 p_2 , ...

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` p_1 p_2 , `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` p_1 , `diff(`*numlist*`)` ...

Specify risk difference $p_2 - p_1$:

> `power twoproportions` p_1 , `rdiff(`*numlist*`)` ...

Specify ratio p_2/p_1 of the two proportions:

> `power twoproportions` p_1 , `ratio(`*numlist*`)` ...

Specify relative risk or risk ratio p_2/p_1:

> `power twoproportions` p_1 , `rrisk(`*numlist*`)` ...

Specify odds ratio $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$:

> `power twoproportions` p_1 , `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 respective 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 χ^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()` 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 $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$. 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 χ^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 χ^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
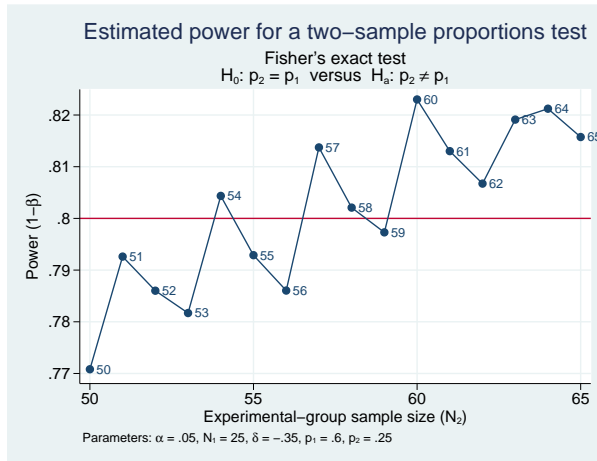
Figure 1. Saw-toothed power function

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 δ 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 2×3 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 χ^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.

◁

Stored results

power twoproportions stores the following in r():

Scalars

r(alpha)	significance level
r(alpha_a)	actual significance level of the Fisher's exact test
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N1)	sample size of the control group
r(N2)	sample size of the experimental group
r(nratio)	ratio of sample sizes, N_2/N_1
r(nratio_a)	actual ratio of sample sizes
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(p1)	control-group proportion
r(p2)	experimental-group proportion
r(diff)	difference between the experimental- and control-group proportions
r(ratio)	ratio of the experimental-group proportion to the control-group proportion
r(rdiff)	risk difference
r(rrisk)	relative risk
r(oratio)	odds ratio
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for sample sizes or experimental-group proportion
r(continuity)	1 if continuity correction is used; 0 otherwise
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros

r(type)	test
r(method)	twoproportions
r(test)	chi2, lrchi2, or fisher
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 of sizes n_1 and n_2. Let $x_{11}, \ldots, x_{1n_1}$ be a random sample from a binomial distribution with the success probability p_1. We refer to this sample as a control group. Let $x_{21}, \ldots, x_{2n_2}$ 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

$$\widehat{p}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \qquad \text{and} \qquad \text{se}(\widehat{p}_1) = \sqrt{n_1 \widehat{p}_1 (1 - \widehat{p}_1)}$$

$$\widehat{p}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i} \qquad \text{and} \qquad \text{se}(\widehat{p}_2) = \sqrt{n_2 \widehat{p}_2 (1 - \widehat{p}_2)}$$

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).

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(rdiff)` 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(oratio)` 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 `oratio()` is specified, the value of the experimental-group proportion p_2 is computed using the corresponding formula from above.

Pearson's chi-squared test

For a large sample size, a binomial process can be approximated by a normal distribution. The asymptotic sampling distribution of the test statistic

$$
z = \frac{(\widehat{p}_2 - \widehat{p}_1) - (p_2 - p_1)}{\sqrt{\widehat{\overline{p}}(1 - \widehat{\overline{p}})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}
$$

is standard normal, where $\overline{p} = (n_1 p_1 + n_2 p_2)/(n_1 + n_2)$ is the pooled proportion and $\widehat{\overline{p}}$ is its estimator. The square of this statistic, z^2, has an approximate χ^2 distribution with one degree of freedom, and the corresponding test is known as Pearson's χ^2 test.

Let α be the significance level, β be the probability of a type II error, and $z_{1-\alpha}$ and z_β be the $(1-\alpha)$th and the βth 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 $\sigma_p = \sqrt{\overline{p}(1-\overline{p})(1/n_1 + 1/n_2)}$ be the pooled standard deviation.

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

$$
\pi = \begin{cases} \Phi\left\{\frac{(p_2-p_1)-c-z_{1-\alpha}\sigma_p}{\sigma_D}\right\} & \text{for an upper one-sided test} \\[2mm] \Phi\left\{\frac{-(p_2-p_1)-c-z_{1-\alpha}\sigma_p}{\sigma_D}\right\} & \text{for a lower one-sided test} \\[2mm] \Phi\left\{\frac{(p_2-p_1)-c-z_{1-\alpha/2}\sigma_p}{\sigma_D}\right\} + \Phi\left\{\frac{-(p_2-p_1)-c-z_{1-\alpha/2}\sigma_p}{\sigma_D}\right\} & \text{for a two-sided test} \end{cases} \tag{1}
$$

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),

$$
n = \frac{\left\{z_{1-\alpha}\sqrt{\overline{p}(1-\overline{p})} - z_\beta\sqrt{w_2 p_1(1-p_1) + w_1 p_2(1-p_2)}\right\}^2}{w_1 w_2 (p_2 - p_1)^2} \tag{2}
$$

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 $c = 1/(2n w_1 w_2)$ (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 $\bar{q} = 1 - \bar{p} = 1 - (n_1 p_1 + n_2 p_2)/(n_1 + n_2)$. The likelihood-ratio test statistic is given by

$$G = \sqrt{2n \left\{ \frac{n_1 p_1}{n} \ln\left(\frac{p_1}{\bar{p}}\right) + \frac{n_1 q_1}{n} \ln\left(\frac{q_1}{\bar{q}}\right) + \frac{n_2 p_2}{n} \ln\left(\frac{p_2}{\bar{p}}\right) + \frac{n_2 q_2}{n} \ln\left(\frac{q_2}{\bar{q}}\right) \right\}}$$

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

$$\pi = \begin{cases} \Phi\left(G - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\ \Phi\left(-G - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\ \Phi\left(G - z_{1-\alpha/2}\right) + \Phi\left(-G - z_{1-\alpha/2}\right) & \text{for a two-sided test} \end{cases} \quad (3)$$

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),

$$n = \frac{\left(z_{1-\alpha} - z_\beta\right)^2}{2 \left\{ w_1 p_1 \ln\left(\frac{p_1}{\bar{p}}\right) + w_1 q_1 \ln\left(\frac{q_1}{\bar{q}}\right) + w_2 p_2 \ln\left(\frac{p_2}{\bar{p}}\right) + w_2 q_2 \ln\left(\frac{q_2}{\bar{q}}\right) \right\}} \quad (4)$$

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 $\widetilde{p}_1$ and $\widetilde{p}_2$ in place of p_1 and p_2 and η in place of θ. 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, $\widetilde{p}_1 = p_2$ is the proportion in the experimental group, $\widetilde{p}_2 = p_1$ is the proportion in the control group, and $\eta = 1/\theta$. Also we denote $\widetilde{n}_1 = n_2$ to be the sample size of the experimental group and $\widetilde{n}_2 = n_1$ 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

$$p(k|m, \eta) = \frac{\binom{\widetilde{n}_1}{k}\binom{\widetilde{n}_2}{m-k}\eta^k}{\sum_i \binom{\widetilde{n}_1}{i}\binom{\widetilde{n}_2}{m-i}\eta^i}$$

where $\eta = \{\widetilde{p}_1(1 - \widetilde{p}_2)\}/\{\widetilde{p}_2(1 - \widetilde{p}_1)\}$, and the range of i is given by $L = \max(0, m - \widetilde{n}_2)$ to $U = \min(\widetilde{n}_1, m)$.

Assume an upper one-sided test given by

$$H_0: \widetilde{p_1} = \widetilde{p_2} \qquad \text{versus} \qquad H_a: \widetilde{p_1} > \widetilde{p_2} \qquad\qquad (5)$$

The hypothesis (5) in terms of η can be expressed as follows:

$$H_0: \eta = 1 \qquad \text{versus} \qquad H_a: \eta > 1$$

Let k_u be the critical value of k such that the following inequalities are satisfied:

$$\sum_{i=k_u}^{U} p(i|m, \eta = 1) \leq \alpha \qquad \text{and} \qquad \sum_{i=k_u-1}^{U} p(i|m, \eta = 1) > \alpha \qquad\qquad (6)$$

The conditional power is

$$\beta(\eta|m) = \sum_{i=k_u}^{U} p(i|m, \eta)$$

For a lower one-sided hypothesis $H_a: \widetilde{p_1} < \widetilde{p_2}$, the corresponding hypothesis in terms of η is given by

$$H_0: \eta = 1 \qquad \text{versus} \qquad H_a: \eta < 1$$

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:

$$\sum_{i=L}^{k_l} p(i|m, \eta = 1) \leq \alpha \qquad \text{and} \qquad \sum_{i=L}^{k_l+1} p(i|m, \eta = 1) > \alpha \qquad\qquad (7)$$

.

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

Title

> **power pairedproportions —** Power analysis for a two-sample paired-proportions test

Syntax

Compute sample size

   Specify discordant proportions

   power pairedproportions p_{12} p_{21} $\big[$, power(*numlist*) *discordopts* $\big]$

   Specify marginal proportions

   power pairedproportions p_{1+} p_{+1} , corr(*numlist*) $\big[$ power(*numlist*) *margopts* $\big]$

Compute power

   Specify discordant proportions

   power pairedproportions p_{12} p_{21} , n(*numlist*) $\big[$ *discordopts* $\big]$

   Specify marginal proportions

   power pairedproportions p_{1+} p_{+1} , corr(*numlist*) n(*numlist*) $\big[$ *margopts* $\big]$

Compute effect size and target discordant proportions

   power pairedproportions, n(*numlist*) power(*numlist*) prdiscordant(*numlist*)

   $\big[$ *discordopts* $\big]$

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**.

discordopts	Description

[Main]

* <u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
* <u>power</u>(*numlist*)	power; default is power(0.8)
* <u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample size
* <u>prdiscordant</u>(*numlist*)	sum of the discordant proportions, $p_{12} + p_{21}$
* sum(*numlist*)	synonym for prdiscordant()
* diff(*numlist*)	difference between the discordant proportions, $p_{21} - p_{12}$
* ratio(*numlist*)	ratio of the discordant proportions, p_{21}/p_{12}
effect(*effect*)	specify the type of effect to display; default is effect(diff)
direction(<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>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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

[Graph]

<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**

[Iteration]

init(*#*)	initial value for the sample size or the difference between discordant proportions
<u>iterate</u>(*#*)	maximum number of iterations; default is iterate(500)
<u>tolerance</u>(*#*)	parameter tolerance; default is tolerance(1e-12)
<u>ftolerance</u>(*#*)	function tolerance; default is ftolerance(1e-12)
[no]<u>log</u>	suppress or display iteration log
[no]<u>dots</u>	suppress or display iterations as dots
notitle	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.

margopts	Description
[Main]	
*<u>alpha</u>(*numlist*)	significance level; default is `alpha(0.05)`
*<u>power</u>(*numlist*)	power; default is `power(0.8)`
*<u>beta</u>(*numlist*)	probability of type II error; default is `beta(0.2)`
*n(*numlist*)	sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample size
*<u>corr</u>(*numlist*)	correlation between the paired observations
*<u>diff</u>(*numlist*)	difference between the marginal proportions, $p_{+1} - p_{1+}$
*<u>ratio</u>(*numlist*)	ratio of the marginal proportions, p_{+1}/p_{1+}
*<u>rrisk</u>(*numlist*)	relative risk, p_{+1}/p_{1+}
*<u>or</u>atio(*numlist*)	odds ratio, $\{p_{+1}(1 - p_{+1})\}/\{p_{1+}(1 - p_{+1})\}$
<u>effect</u>(*effect*)	specify the type of effect to display; default is `effect(diff)`
<u>direction</u>(<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>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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph]	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**
<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.

effect	Description
diff	difference between the discordant proportions, $p_{21} - p_{12}$, or marginal proportions, $p_{+1} - p_{1+}$; the default
ratio	ratio of the discordant proportions, p_{21}/p_{12}, or of the marginal proportions, p_{+1}/p_{1+}
<u>rr</u>isk	relative risk, p_{+1}/p_{1+}; may only be specified with marginal proportions
<u>or</u>atio	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	number of subjects	N
delta	effect size	δ
p12	success–failure proportion	p_{12}
p21	failure–success proportion	p_{21}
pmarg1	success proportion in occasion 1	p_{1+}
pmarg2	success proportion in occasion 2	p_{+1}
corr	correlation between paired observations	ρ
prdiscordant	proportion of discordant pairs	$p_{12} + p_{21}$
sum	sum of discordant proportions	$p_{12} + p_{21}$
diff	difference between discordant proportions	$p_{21} - p_{12}$
	difference between marginal proportions	$p_{+1} - p_{1+}$
ratio	ratio of discordant proportions	p_{21}/p_{12}
	ratio of marginal proportions	p_{+1}/p_{1+}
rrisk	relative risk for marginal proportions	p_{+1}/p_{1+}
oratio	odds ratio for marginal proportions	θ
target	target parameter; synonym for p12	
_all	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.

Menu

Statistics > Power and sample size

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.

Options

Main

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

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.

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.

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(*tablespec*), 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:

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 2×2 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 2×2 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 χ_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 δ is the difference between the discordant proportions.

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

The above specification is equivalent to

    power pairedproportions p_{12} p_{21} , effect(diff) ...

Alternatively, you may request the effect size to be the ratio instead of the difference.

    power pairedproportions p_{12} p_{21} , 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 p_{1+} , 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 p_{1+} , corr(*numlist*) ratio(*numlist*) ...

    power pairedproportions p_{1+} , corr(*numlist*) rrisk(*numlist*) ...

or the odds ratio, $\{p_{+1}(1 - p_{1+})\}/\{p_{1+}(1 - p_{+1})\}$, in option oratio():

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

is equivalent to

> `power pairedproportions` p_{1+} p_{+1} `, corr(`*numlist*`) effect(diff)` ...

You may request other measures of effect size for marginal proportions such as the risk ratio,

> `power pairedproportions` p_{1+} p_{+1} `, corr(`*numlist*`) effect(rrisk)` ...

or odds ratio:

> `power pairedproportions` p_{1+} 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 2×2 table:

2012 Election	2016 Election Democratic	Republican	Total
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 δ 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, θ, 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 [ST] **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
                      Controls
Cases                 Exposed    Unexposed             Total

          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
r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	sample size
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(p12)	success–failure proportion (first discordant proportion)
r(p21)	failure–success proportion (second discordant proportion)
r(pmarg1)	success proportion for occasion 1 (first marginal proportion)
r(pmarg2)	success proportion for occasion 2 (second marginal proportion)
r(corr)	correlation between paired observations
r(diff)	difference between proportions
r(ratio)	ratio of proportions
r(prdiscordant)	proportion of discordant pairs
r(sum)	sum of discordant proportions
r(rrisk)	relative risk
r(oratio)	odds ratio
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for the sample size or the difference between discordant proportions
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros
r(effect)	diff, ratio, oratio, or rrisk
r(type)	test
r(method)	pairedproportions
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 a 2×2 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 ρ 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 χ^2 distribution with 1 degree of freedom under the null hypothesis. The square root of the χ^2 test statistic is approximately normal with zero mean and variance of one.

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

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

$$\pi = \begin{cases} \Phi\left\{\dfrac{p_{\text{diff}}\sqrt{n} - z_{1-\alpha}\sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}}\right\} & \text{for an upper one-sided test} \\[3ex] \Phi\left\{\dfrac{-p_{\text{diff}}\sqrt{n} - z_{1-\alpha}\sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}}\right\} & \text{for a lower one-sided test} \\[3ex] \Phi\left\{\dfrac{p_{\text{diff}}\sqrt{n} - z_{1-\alpha/2}\sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}}\right\} + \Phi\left\{\dfrac{-p_{\text{diff}}\sqrt{n} - z_{1-\alpha/2}\sqrt{p_{\text{disc}}}}{\sqrt{p_{\text{disc}} - p_{\text{diff}}^2}}\right\} & \text{for a two-sided test} \end{cases}$$

$$\tag{2}$$

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**

[ST] **epitab** — Tables for epidemiologists

Title

> **power onevariance** — Power analysis for a one-sample variance test

Syntax Menu Description Options
Remarks and examples Stored results Methods and formulas Reference
Also see

Syntax

Compute sample size

  Variance scale

    power <u>onevar</u>iance v_0 v_a $\left[\,,\ \underline{\text{power}}(\textit{numlist})\ \textit{options}\,\right]$

  Standard deviation scale

    power <u>onevar</u>iance s_0 s_a , sd $\left[\,\underline{\text{power}}(\textit{numlist})\ \textit{options}\,\right]$

Compute power

  Variance scale

    power <u>onevar</u>iance v_0 v_a , n(*numlist*) $\left[\,\textit{options}\,\right]$

  Standard deviation scale

    power <u>onevar</u>iance s_0 s_a , sd n(*numlist*) $\left[\,\textit{options}\,\right]$

Compute effect size and target parameter

  Target variance

    power <u>onevar</u>iance v_0 , n(*numlist*) <u>power</u>(*numlist*) $\left[\,\textit{options}\,\right]$

  Target standard deviation

    power <u>onevar</u>iance s_0 , sd n(*numlist*) <u>power</u>(*numlist*) $\left[\,\textit{options}\,\right]$

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**).

options	Description
sd	request computation using the standard deviation scale; default is the variance scale

Main

*<u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
*<u>power</u>(*numlist*)	power; default is power(0.8)
*<u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample size
*<u>ratio</u>(*numlist*)	ratio of variances, v_a/v_0 (or ratio of standard deviations, s_a/s_0, if option sd is specified); specify instead of the alternative variance v_a (or standard deviation s_a)
<u>dir</u>ection(<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>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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

Graph

<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration

init(*#*)	initial value for sample size or variance
<u>iter</u>ate(*#*)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(*#*)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(*#*)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	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**.

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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	number of subjects	N
delta	effect size	δ
v0	null variance	σ_0^2
va	alternative variance	σ_a^2
s0	null standard deviation	σ_0
sa	alternative standard deviation	σ_a
ratio	ratio of the alternative variance to the null variance	σ_a^2/σ_0^2
	or ratio of the alternative standard deviation to the null standard deviation (if sd is specified)	σ_a/σ_0
target	target parameter; synonym for va	
_all	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.

Menu

Statistics > Power and sample size

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.

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(*tablespec*), 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 χ^2 test statistic $(n-1)s^2/\sigma_0^2$, which has a chi-squared distribution χ_{n-1}^2 with $n-1$ degrees of freedom, is used to test hypotheses on variance, and the corresponding test is known as a χ^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 χ^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 χ^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 χ^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 δ 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/r13/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 $H_0: \sigma_{\mathrm{mpg}} = 3$ versus a two-sided alternative $H_a: \sigma_{\mathrm{mpg}} \neq 3$ 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

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	sample size
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(v0)	variance under the null hypothesis (for variance scale, default)
r(va)	variance under the alternative hypothesis (for variance scale, default)
r(s0)	standard deviation under the null hypothesis (if option sd is specified)
r(sa)	standard deviation under the alternative hypothesis (if option sd is specified)
r(ratio)	ratio of the alternative variance to the null variance (or the ratio of standard deviations if option sd is specified)
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for the sample size or the variance
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros

r(type)	test
r(method)	onevariance
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
r(scale)	variance or standard deviation

Matrix

r(pss_table)	table of results

Methods and formulas

Consider a random sample of size n from a normal population with mean μ and variance σ^2. Let σ_0^2 and σ_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 χ^2 distribution with $n-1$ degrees of freedom, where s^2 is the sample variance. The corresponding test is known as a χ^2 test.

The following formulas are based on Dixon and Massey (1983, 110–112).

Let α be the significance level, β 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 βth quantiles of the χ^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 χ^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 σ_a^2. The default initial value is obtained from (2) with α 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

Title

---

power twovariances — Power analysis for a two-sample variances test

---

Syntax

Compute sample size

  Variance scale

    power <u>twovar</u>iances v_1 v_2 [, <u>power</u>(*numlist*) *options*]

  Standard deviation scale

    power <u>twovar</u>iances s_1 s_2 , sd [<u>power</u>(*numlist*) *options*]

Compute power

  Variance scale

    power <u>twovar</u>iances v_1 v_2 , n(*numlist*) [*options*]

  Standard deviation scale

    power <u>twovar</u>iances s_1 s_2 , sd n(*numlist*) [*options*]

Compute effect size and target parameter

  Experimental-group variance

    power <u>twovar</u>iances v_1 , n(*numlist*) <u>power</u>(*numlist*) [*options*]

  Experimental-group standard deviation

    power <u>twovar</u>iances s_1 , sd n(*numlist*) <u>power</u>(*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**).

options	Description
sd	request computation using the standard deviation scale; default is the variance scale

Main

*alpha(*numlist*)	significance level; default is alpha(0.05)
*power(*numlist*)	power; default is power(0.8)
*beta(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	total sample size; required to compute power or effect size
*n1(*numlist*)	sample size of the control group
*n2(*numlist*)	sample size of the experimental group
*nratio(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
compute(n1 | n2)	solve for N_1 given N_2 or for N_2 given N_1
nfractional	allow fractional sample sizes
*ratio(*numlist*)	ratio of variances, v_2/v_1 (or ratio of standard deviations, s_2/s_1, if option sd is specified); specify instead of the experimental-group variance v_2 (or standard deviation s_2)
direction(upper|lower)	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
onesided	one-sided test; default is two sided
parallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)

Table

[no]table[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
saving(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

Graph

graph[(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration

init(*#*)	initial value for the sample sizes or experimental-group variance
iterate(*#*)	maximum number of iterations; default is iterate(500)
tolerance(*#*)	parameter tolerance; default is tolerance(1e-12)
ftolerance(*#*)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	suppress or display iterations as dots
notitle	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
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	δ
v1	control-group variance	σ_1^2
v2	experimental-group variance	σ_2^2
s1	control-group standard deviation	σ_1
s2	experimental-group standard deviation	σ_2
ratio	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 sd is specified)	σ_2^2/σ_1^2 σ_2/σ_1
target	target parameter; synonym for v2	
_all	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.

Menu

Statistics > Power and sample size

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.

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(*tablespec*), 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 χ^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:

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 δ 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 δ 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/r13/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

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N1)	sample size of the control group
r(N2)	sample size of the experimental group
r(nratio)	ratio of sample sizes, N_2/N_1
r(nratio_a)	actual ratio of sample sizes
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(v1)	control-group variance
r(v2)	experimental-group variance
r(ratio)	ratio of the experimental- to the control-group variances (or standard deviations if sd is specified)
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for sample sizes, experimental-group variance, or standard deviation
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	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 μ_1 and μ_2 and variances σ_1^2 and σ_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 σ_1^2 and σ_2^2 be the control-group and experimental-group variances, respectively.

A two-sample variances test involves testing the null hypothesis $H_0\colon \sigma_2^2 = \sigma_1^2$ versus the two-sided alternative hypothesis $H_a\colon \sigma_2^2 \neq \sigma_1^2$, the upper one-sided alternative $H_a\colon \sigma_2^2 > \sigma_1^2$, or the lower one-sided alternative $H_a\colon \sigma_2^2 < \sigma_1^2$.

Equivalently, the hypotheses can be expressed in terms of the ratio of the two variances: $H_0\colon \sigma_2^2/\sigma_1^2 = 1$ versus the two-sided alternative $H_a\colon \sigma_2^2/\sigma_1^2 \neq 1$, the upper one-sided alternative $H_a\colon \sigma_2^2/\sigma_1^2 > 1$, or the lower one-sided alternative $H_a\colon \sigma_2^2/\sigma_1^2 < 1$.

The following formulas are based on Dixon and Massey (1983, 116–119).

Let α be the significance level, β 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 αth and the βth 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

$$
\pi = \begin{cases}
1 - F_{n_1-1, n_2-1}\left(\frac{\sigma_1^2}{\sigma_2^2} F_{1-\alpha}\right) & \text{for an upper one-sided test} \\
F_{n_1-1, n_2-1}\left(\frac{\sigma_1^2}{\sigma_2^2} F_\alpha\right) & \text{for a lower one-sided test} \\
1 - F_{n_1-1, n_2-1}\left(\frac{\sigma_1^2}{\sigma_2^2} F_{1-\alpha/2}\right) + F_{n_1-1, n_2-1}\left(\frac{\sigma_1^2}{\sigma_2^2} F_{\alpha/2}\right) & \text{for a two-sided test}
\end{cases}
\tag{1}
$$

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:

$$
\sigma_2^2 = \begin{cases}
\sigma_1^2 \frac{F_{n_1-1, n_2-1, 1-\alpha}}{F_{n_1-1, n_2-1, \beta}} & \text{for an upper one-sided test} \\
\sigma_1^2 \frac{F_{n_1-1, n_2-1, \alpha}}{F_{n_1-1, n_2-1, 1-\beta}} & \text{for a lower one-sided test}
\end{cases}
\tag{2}
$$

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 α 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

Title

> **power onecorrelation** — Power analysis for a one-sample correlation test

Syntax

Compute sample size

> power <u>onecorr</u>elation r_0 r_a $\left[\,,\ \underline{\text{p}}\text{ower}(\textit{numlist})\ \textit{options}\,\right]$

Compute power

> power <u>onecorr</u>elation r_0 r_a , n(*numlist*) $\left[\,\textit{options}\,\right]$

Compute effect size and target correlation

> power <u>onecorr</u>elation r_0 , n(*numlist*) <u>p</u>ower(*numlist*) $\left[\,\textit{options}\,\right]$

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**).

options	Description
[Main]	
*<u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
*<u>power</u>(*numlist*)	power; default is power(0.8)
*<u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample size
*<u>diff</u>(*numlist*)	difference between the alternative correlation and the null correlation, $r_a - r_0$; specify instead of the alternative correlation r_a
<u>dir</u>ection(<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>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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph]	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**
[Iteration]	
<u>init</u>(*#*)	initial value for the sample size or correlation
<u>iterate</u>(*#*)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(*#*)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(*#*)	function tolerance; default is ftolerance(1e-12)
[<u>no</u>]log	suppress or display iteration log
[<u>no</u>]dots	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	number of subjects	N
delta	effect size	δ
r0	null correlation	ρ_0
ra	alternative correlation	ρ_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.

Menu

Statistics > Power and sample size

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.

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(*tablespec*), 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 ρ 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, ρ_0, is set to zero.

Statistical inference on the correlation coefficient requires a distributional assumption between two random variables—bivariate normality with correlation ρ. 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

$$\sqrt{n-3}\left\{\tanh^{-1}(\widehat{\rho}) - \tanh^{-1}(\rho)\right\} \sim N(0,1)$$

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 respective 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 ρ is zero and becomes negatively skewed as ρ approaches 1 and positively skewed as ρ 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, δ 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/r13/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
r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	sample size
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(r0)	correlation under the null hypothesis
r(ra)	correlation under the alternative hypothesis
r(diff)	difference between the alternative and null correlations
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for the sample size or the correlation
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros
r(type)	test
r(method)	onecorrelation
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 ρ denote Pearson's correlation and $\widehat{\rho}$ 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 $\widehat{\rho}$ 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 $\widehat{\rho}$ 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 $Z = \tanh^{-1}(\widehat{\rho}) = 0.5 \ln\{(1 + \widehat{\rho})/(1 - \widehat{\rho})\}$ denote Fisher's z transformation of $\widehat{\rho}$. 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 α be the significance level, β be the probability of a type II error, and $z_{1-\alpha/k}$ and z_β be the $(1 - \alpha/k)$th and the βth 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

$$
\pi = \begin{cases}
\Phi\left(\frac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\
\Phi\left(-\frac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\
\Phi\left(\frac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) + \Phi\left(-\frac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) & \text{for a two-sided test}
\end{cases}
\tag{1}
$$

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 μ_a and then applying the inverse Fisher's z transformation to the μ_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 μ_a are obtained by iteratively solving the two-sided power equation in (1) for n and μ_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

Syntax

Compute sample size

> power <u>twocorr</u>elations r_1 r_2 $\big[$, <u>p</u>ower(*numlist*) *options* $\big]$

Compute power

> power <u>twocorr</u>elations r_1 r_2 , n(*numlist*) $\big[$ *options* $\big]$

Compute effect size and experimental-group correlation

> power <u>twocorr</u>elations r_1 , n(*numlist*) <u>p</u>ower(*numlist*) $\big[$ *options* $\big]$

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**).

options	Description
[Main]	
* <u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
* <u>power</u>(*numlist*)	power; default is power(0.8)
* <u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	total sample size; required to compute power or effect size
* n1(*numlist*)	sample size of the control group
* n2(*numlist*)	sample size of the experimental group
* <u>nratio</u>(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
<u>compute</u>(n1 | n2)	solve for N_1 given N_2 or for N_2 given N_1
<u>nfrac</u>tional	allow fractional sample sizes
* <u>diff</u>(*numlist*)	difference between the experimental-group and control-group correlations, $r_2 - r_1$; specify instead of the experimental-group correlation r_2
<u>dir</u>ection(<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>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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph]	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**
[Iteration]	
init(#)	initial value for sample sizes or experimental-group correlation
<u>iterate</u>(#)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(#)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(#)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
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	δ
r1	control-group correlation	ρ_1
r2	experimental-group correlation	ρ_2
diff	difference between the experimental-group correlation and the control-group correlation	$\rho_2 - \rho_1$
target	target parameter; synonym for r2	
_all	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.

Menu

Statistics > Power and sample size

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.

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(`*tablespec*`)`, `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 ρ 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

$$\sqrt{n - 3} \left\{ \tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho) \right\} \sim N(0, 1)$$

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 ρ_F is the correlation between height and weight for females, and ρ_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 respective 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 δ 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/r13/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

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N1)	sample size of the control group
r(N2)	sample size of the experimental group
r(nratio)	ratio of sample sizes, N_2/N_1
r(nratio_a)	actual ratio of sample sizes
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(onesided)	1 for a one-sided test; 0 otherwise
r(r1)	control-group correlation
r(r2)	experimental-group correlation
r(diff)	difference between the experimental- and control-group correlations
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for sample sizes or experimental-group correlation
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	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 ρ_1 and ρ_2 denote Pearson's correlation for the control and the experimental groups, respectively. Let $\widehat{\rho}_1$ and $\widehat{\rho}_2$ 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 $\widehat{\rho}_1$ and $\widehat{\rho}_2$ 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 $Z_1 = \tanh^{-1}(\widehat{\rho}_1) = 0.5\ln\{(1 + \widehat{\rho}_1)/(1 - \widehat{\rho}_1)\}$ denote Fisher's z transformation of $\widehat{\rho}_1$, and let Z_2 denote the corresponding transformation of $\widehat{\rho}_2$. 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 α be the significance level, β be the probability of a type II error, and $z_{1-\alpha}$ and z_β be the $(1-\alpha)$th and the βth quantiles of the standard normal distribution.

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

$$
\pi = \begin{cases}
\Phi\left(\dfrac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for an upper one-sided test} \\[2mm]
\Phi\left(-\dfrac{\delta_z}{\sigma_z} - z_{1-\alpha}\right) & \text{for a lower one-sided test} \\[2mm]
\Phi\left(\dfrac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) + \Phi\left(-\dfrac{\delta_z}{\sigma_z} - z_{1-\alpha/2}\right) & \text{for a two-sided test}
\end{cases}
\tag{1}
$$

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 μ_2 and then applying the inverse Fisher's z transformation to that μ_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 μ_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

Title

> **power oneway** — Power analysis for one-way analysis of variance

Syntax

Compute sample size

> power oneway *meanspec* $\big[$, $\underline{\text{p}}$ower(*numlist*) *options* $\big]$

Compute power

> power oneway *meanspec*, n(*numlist*) $\big[$ *options* $\big]$

Compute effect size and target between-group variance

> power oneway, n(*numlist*) $\underline{\text{p}}$ower(*numlist*) $\underline{\text{ng}}$roups(#) $\big[$ $\underline{\text{varerr}}$or(*numlist*) *options* $\big]$

where *meanspec* is either a matrix *matname* containing group means or individual group means specified in a matrix form:

$m_1\ m_2\ [m_3\ \ldots\ m_J]$

m_j, where $j = 1, 2, \ldots, J$, is the alternative group mean or the group mean under the alternative hypothesis for the jth 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 jth group mean.

options	Description
[Main]	
*alpha(*numlist*)	significance level; default is alpha(0.05)
*power(*numlist*)	power; default is power(0.8)
*beta(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	total sample size; required to compute power or effect size
nfractional	allow fractional sample sizes
*npergroup(*numlist*)	number of subjects per group; implies balanced design
*n#(*numlist*)	number of subjects in a group #
grweights(*wgtspec*)	group weights; default is one for each group, meaning equal-group sizes
ngroups(*#*)	number of groups
*varmeans(*numlist*)	variance of the group means or between-group variance
*varerror(*numlist*)	error (within-group) variance; default is varerror(1)
contrast(*contrastspec*)	contrast specification for group means
parallel	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
[Table]	
[no]table[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
saving(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
[Graph]	
graph[(*graphopts*)]	graph results; see [PSS] **power, graph**
[Iteration]	
init(*#*)	initial value for the sample size or the effect size; default is to use a bisection algorithm to bound the solution
iterate(*#*)	maximum number of iterations; default is iterate(500)
tolerance(*#*)	parameter tolerance; default is tolerance(1e-12)
ftolerance(*#*)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	suppress or display iterations as dots
notitle	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.

wgtspec	Description
$\#_1 \; \#_2 \; \ldots \; \#_J$	J group weights. Weights must be positive and must be integers unless option nfractional is specified. Multiple values for each group weight $\#j$ can be specifed as a *numlist* enclosed in parentheses.
matname	matrix with J columns containing J group weights. Multiple rows are allowed, in which case each row corresponds to a different set of J weights or, equivalently, column j corresponds to a *numlist* for the jth 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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	total number of subjects	N
N_per_group	number of subjects per group	N/N_g
N_avg	average number of subjects per group	N_{avg}
N$\#_1$_$\#_2$	number of subjects in group ($\#_1$, $\#_2$)	$N_{\#_1, \#_2}$
delta	effect size	δ
N_g	number of groups	N_g
m#	group mean #	$\mu_\#$
Cm	mean contrast	$C \cdot \mu$
c0	null mean contrast	c_0
Var_m	group means (between-group) variance	σ_m^2
Var_Cm	contrast variance	$\sigma_{C\mu}^2$
Var_e	error (within-group) variance	σ_e^2
grwgt#	group weight #	$w_\#$
target	target parameter; synonym for target effect variance	
_all	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 grwgt# are not shown in the default table.

Menu

Statistics > Power and sample size

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.

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). If varerror() is not specified with effect-size determination, only effect size δ is reported; otherwise, the estimated between-group variance is also reported.

contrast(*contrastspec*) specifies a contrast for group means containing J contrast coefficients that must sum to zero. *contrastspec* is

$\#_1 \#_2 \left[\ \#_3 \ldots \#_J \right] \left[\ , \text{null}(\textit{numlist}) \ \underline{\text{onesided}} \right]$

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(*tablespec*), 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 δ 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*

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: $\mu_1 = \cdots = \mu_J$, 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, 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. If the varerror() option is also specified, the target between-group variance is reported in addition to the effect size.

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(#)` $\big[$...$\big]$

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 *numlist*s may contain different numbers of values, $K_1 \neq K_2 \neq ... \neq K_J$. `power oneway` will produce results for all possible combinations of values across *numlist*s. 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 jth group mean or a *numlist* for the jth 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* = $(m_{1,1}, \ m_{2,1} \ \backslash \ m_{1,2}, \ m_{2,2} \ \backslash \ m_{1,3}, \ m_{2,3})$

>   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* = $(m_{1,1}, \ m_{2,1} \ \backslash \ m_{1,2}, \ . \ \ \backslash \ m_{1,3}, \ .)$

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 $= \sqrt{233.556/4900} = 0.2183$, 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 δ 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/r13/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:  chi2(3) =   1.0063  Prob>chi2 = 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, varerror(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
```

| | Number of obs = | 58 | | R-squared | = 0.3355 |
| | Root MSE | = 10.7211 | | Adj R-squared = | 0.2985 |

Source	Partial SS	df	MS	F	Prob > F
Model	3133.23851	3	1044.41284	9.09	0.0001
drug	3133.23851	3	1044.41284	9.09	0.0001
Residual	6206.91667	54	114.942901		
Total	9340.15517	57	163.862371		

```
. 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.

◁

Stored results

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

Scalars
r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N_avg)	average sample size
r(N#)	number of subjects in group #
r(N_per_group)	number of subjects per group
r(N_g)	number of groups
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(balanced)	1 for a balanced design; 0 otherwise
r(grwgt#)	group weight #
r(onesided)	1 for a one-sided test of a mean contrast; 0 otherwise
r(m#)	group mean #
r(Cm)	mean contrast
r(c0)	null mean contrast
r(Var_m)	group-means (between-group) variance
r(Var_Cm)	contrast variance
r(Var_e)	error (within-group) variance
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for the sample size or effect size
r(maxiter)	maximum number of iterations

r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros

r(type)	test
r(method)	oneway
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

Methods and formulas

Consider a single factor A with J groups or levels, where each level comprises n_j observations for $j = 1, \ldots, J$. The total number of observations is $n = \sum_{j=1}^{J} n_j$. Let Y_{ij} denote the response for the jth level of the ith individual and μ_j denote the factor-level or group means. Individual responses from J populations are assumed to be normally distributed with mean μ_j and a constant variance σ_e^2.

The hypothesis of an overall F test of the equality of group means is

$$H_0: \mu_1 = \cdots = \mu_J$$

versus

$$H_a: \mu_j\text{'s are not all equal} \tag{1}$$

The hypothesis for a test of a mean contrast is

$$H_0: \sum_{j=1}^{J} c_j \mu_j = c_0 \tag{2}$$

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

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{\epsilon}$$

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 $\mathbf{b} = (b_1, \ldots, b_J)'$ are the μ_j, $j = 1, \ldots, J$.

A general linear hypothesis in this framework is given by

$$\mathbf{Cb} = \mathbf{c_0}$$

where $\mathbf{C}$ is a $\nu \times p$ matrix with rank$(\mathbf{C}) = \nu \le p$, 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 σ_e^2, respectively, are given by

$$\widehat{\mathbf{b}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

$$\widehat{\sigma}_e^2 = (\mathbf{y} - \mathbf{X}\widehat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\widehat{\mathbf{b}})/(n - p)$$

A general test statistic for testing hypotheses (1) and (2) is given by

$$F_C = \frac{SS_C}{(p-1)\widehat{\sigma}_e^2} \tag{3}$$

where $SS_C = (\mathbf{C}\widehat{\mathbf{b}} - \mathbf{c_0})'\{\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'}\}^{-1}(\mathbf{C}\widehat{\mathbf{b}} - \mathbf{c_0})$. Let α be the significance level, β 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 λ given by

$$\lambda = (\mathbf{Cb} - \mathbf{c_0})'\{\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'}\}^{-1}(\mathbf{Cb} - \mathbf{c_0})/\sigma_e^2$$
$$= n(\mathbf{Cb} - \mathbf{c_0})'\{\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C'}\}^{-1}(\mathbf{Cb} - \mathbf{c_0})/\sigma_e^2$$
$$= n\delta^2$$

where the matrix $\ddot{\mathbf{X}}$ contains the unique rows of $\mathbf{X}$ and $\mathbf{W} = \text{diag}(w_1, \ldots, w_p)$. We define δ 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 $\widetilde{w}_j$, then define a cell sample size multiplier as $n_c = n/\sum_j \widetilde{w}_j$ so that $n_j = n_c\widetilde{w}_j$. 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

$$1 - \beta = F_{\nu,n-p,\lambda}\left(F_{\nu,n-p,1-\alpha}\right) \tag{4}$$

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 $\widetilde{w}_j n_c$. 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

$$
1 - \beta = \begin{cases} 1 - T_{n-1,\widetilde{\lambda}}\left(t_{n-1,1-\alpha}\right) & \text{for an upper one-sided test} \\ T_{n-1,\widetilde{\lambda}}\left(-t_{n-1,1-\alpha}\right) & \text{for a lower one-sided test} \\ F_{1,n-p,\lambda}\left(F_{1,n-p,1-\alpha}\right) & \text{for a two-sided test} \end{cases} \tag{5}
$$

where $T_{.,\widetilde{\lambda}}(\cdot)$ is the cumulative of a noncentral Student's t distribution with the noncentrality parameter $\widetilde{\lambda}$ given by

$$
\begin{aligned}
\widetilde{\lambda} &= \sqrt{n}(\mathbf{Cb} - \mathbf{c_0})'\sqrt{\left\{\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}'\right\}^{-1}/\sigma_e^2} \\
&= \sqrt{n}\widetilde{\delta}
\end{aligned}
$$

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

Title

> **power twoway** — Power analysis for two-way analysis of variance

Syntax

Compute sample size

> power twoway *meanspec* $\big[$, $\underline{\text{p}}$ower(*numlist*) *options* $\big]$

Compute power

> power twoway *meanspec*, n(*numlist*) $\big[$ *options* $\big]$

Compute effect size and target effect variance

> power twoway, n(*numlist*) $\underline{\text{p}}$ower(*numlist*) $\underline{\text{nr}}$ows(#) $\underline{\text{nc}}$ols(#) $\big[$ *options* $\big]$

where *meanspec* is either a matrix *matname* containing cell means or individual cell means in a matrix form:

$$m_{1,1} \ m_{1,2} \ \big[\dots\big] \setminus m_{2,1} \ m_{2,2} \ \big[\dots\big] \ \big[\setminus \dots \setminus m_{J,1} \ \dots m_{J,K}\big]$$

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 jth row and kth 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.

options	Description

[Main]

* <u>a</u>lpha(*numlist*)	significance level; default is alpha(0.05)
* <u>p</u>ower(*numlist*)	power; default is power(0.8)
* <u>b</u>eta(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	total sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample sizes
* <u>nperc</u>ell(*numlist*)	number of subjects per cell; implies balanced design
<u>cellw</u>eights(*wgtspec*)	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 | <u>col</u>umn | rowcol)	tested effect
* <u>vareff</u>ect(*numlist*)	variance explained by the tested effect in factor()
* <u>varr</u>ow(*numlist*)	variance explained by the row effect; synonym for factor(row) and vareffect(*numlist*)
* <u>varc</u>olumn(*numlist*)	variance explained by the column effect; synonym for factor(column) and vareffect(*numlist*)
* <u>varr</u>owcolumn(*numlist*)	variance explained by the row–column interaction effect; synonym for factor(rowcol) and vareffect(*numlist*)
* <u>vare</u>rror(*numlist*)	error variance; default is varerror(1)
<u>showmat</u>rices	display cell means and sample sizes as matrices
<u>showme</u>ans	display cell means
<u>showc</u>ellsizes	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> [(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

[Graph]

<u>graph</u> [(*graphopts*)]	graph results; see [PSS] **power, graph**

[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>iter</u>ate(*#*)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(*#*)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(*#*)	function tolerance; default is ftolerance(1e-12)
[<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.

wgtspec	Description
$\#_{1,1} \ldots \#_{1,K} \setminus \ldots \setminus \#_{J,1} \ldots \#_{J,K}$	$J \times K$ cell weights; weights must be positive and must be integers unless option nfractional is specified
matname	$J \times K$ 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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	total number of subjects	N
N_per_cell	number of subjects per cell	N/N_{rc}
N_avg	average number of subjects per cell	N_{avg}
N$\#_1$_$\#_2$	number of subjects in cell ($\#_1$, $\#_2$)	$N_{\#_1,\#_2}$
delta	effect size	δ
N_rc	number of cells	N_{rc}
N_r	number of rows	N_r
N_c	number of columns	N_c
m$\#_1$_$\#_2$	cell mean ($\#_1$, $\#_2$)	$\mu_{\#_1,\#_2}$
Var_r	variance explained by the row effect	σ_r^2
Var_c	variance explained by the column effect	σ_c^2
Var_rc	variance explained by the row–column interaction	σ_{rc}^2
Var_e	error variance	σ_e^2
cwgt$\#_1$_$\#_2$	cell weight ($\#_1$, $\#_2$)	$w_{\#_1,\#_2}$
target	target parameter; synonym for target effect variance	
_all	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.

Menu

Statistics > Power and sample size

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.

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). If var-
error() is not specified with effect-size determination, only effect size δ is reported; otherwise,
the estimated variance of the tested effect is also reported.

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(*tablespec*), 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 δ 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 μ_{jk} be the cell mean of the jth row and the kth column in a two-way cell-means ANOVA model, $\mu_{j\cdot}$ be the marginal mean of the jth row, $\mu_{\cdot k}$ be the marginal mean of the kth column, and $\mu_{\cdot\cdot}$ be the grand mean. The jth-row-by-kth-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: $\mu_{1\cdot} = \ldots = \mu_{J\cdot}$, for testing the main row effect; 2) H_0: $\mu_{\cdot 1} = \ldots = \mu_{\cdot K}$, for testing the main column effect; and 3) H_0: 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, 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. If the `varerror()` option is also specified, the target variance explained by the tested effect is reported in addition to the effect size.

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` $m_{1,1}$ $m_{1,2}$ $m_{1,3}$ \ $m_{2,1}$ $m_{2,2}$ $m_{2,3}$ $\big[$ `, factor()` $...\big]$

or the variance of the tested effect and the number of rows J and columns K:

> `power twoway, factor() vareffect(#) nrows(2) ncols(3)` $\big[...\big]$

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* = $(m_{1,1},\ m_{1,2},\ m_{1,3}$ \ $m_{2,1},\ m_{2,2},\ m_{2,3})$

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 δ 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 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 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 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 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 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 is computed by default. If the error variance is specified in the `varerror()` option, the target variance of the effect is also reported.

▷ Example 8: Effect size for a two-way ANOVA—row effect

Continuing 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/r13/systolic
(Systolic Blood Pressure Data)

. anova systolic drug disease drug#disease
```

	Number of obs =	58	R-squared	=	0.4560
	Root MSE	= 10.5096	Adj R-squared =		0.3259

Source	Partial SS	df	MS	F	Prob > F
Model	4259.33851	11	387.212591	3.51	0.0013
drug	2997.47186	3	999.157287	9.05	0.0001
disease	415.873046	2	207.936523	1.88	0.1637
drug#disease	707.266259	6	117.87771	1.07	0.3958
Residual	5080.81667	46	110.452536		
Total	9340.15517	57	163.862371		

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)
```

	Patient's Disease		
Drug Used	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
r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N_avg)	average sample size
r(N#$_1$_#$_2$)	number of subjects in cell (#$_1$, #$_2$)
r(N_per_cell)	number of subjects per cell
r(N_rc)	number of cells
r(nfractional)	1 if nfractional is specified; 0 otherwise
r(balanced)	1 for a balanced design; 0 otherwise
r(cwgt#$_1$_#$_2$)	cell weight (#$_1$, #$_2$)
r(onesided)	1 for a one-sided test of a mean contrast; 0 otherwise
r(N_r)	number of rows
r(N_c)	number of columns
r(m#$_1$_#$_2$)	cell mean (#$_1$, #$_2$)
r(Var_r)	row variance
r(Var_c)	column variance
r(Var_rc)	row-by-column variance
r(Var_e)	error variance
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for sample size or effect size
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged; 0 otherwise

Macros
r(type)	test
r(method)	twoway
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
r(Nij)	cell-sizes matrix
r(means)	cell-means matrix
r(cwgt)	cell-weights matrix

Methods and formulas

Consider factor A with J levels and factor B with K levels. Let μ_{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 $k = 1$	$k = 2$	$k = 3$	Total
$j = 1$	μ_{11}	μ_{12}	μ_{13}	$\mu_{1\cdot}$
$j = 2$	μ_{21}	μ_{22}	μ_{23}	$\mu_{2\cdot}$
$j = 3$	μ_{31}	μ_{32}	μ_{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 jth level is $a_j = \mu_{j\cdot} - \mu_{\cdot\cdot}$. Similarly, the main effect of factor B at the kth 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 jth level and factor B at the kth 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 ith individual at the jth level of factor A and kth level of factor B for $i = 1, \ldots, n_{jk}$ 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 σ_e^2. Using the definitions of a_j, b_k, and $(ab)_{jk}$ from the previous sections, our linear model is expressed as

$$y_{ijk} = \mu_{..} + a_j + b_k + (ab)_{jk} + e_{ijk}$$
$$= \mu_{jk} + e_{ijk}$$

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:

$$H_0: \text{all } a_j = 0 \qquad \text{versus} \qquad H_a: \text{at least one } a_j \neq 0 \tag{1}$$
$$H_0: \text{all } b_k = 0 \qquad \text{versus} \qquad H_a: \text{at least one } b_k \neq 0 \tag{2}$$
$$H_0: \text{all } (ab)_{jk} = 0 \qquad \text{versus} \qquad H_a: \text{at least one } (ab)_{jk} \neq 0 \tag{3}$$

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

$$F_C = \frac{SS_C}{\nu \widehat{\sigma}_e^2} \tag{4}$$

where $SS_C = (\mathbf{C}\widehat{\mathbf{b}})'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}(\mathbf{C}\widehat{\mathbf{b}})$. 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, v is $J - 1$, $K - 1$, or $(J-1)(K-1)$ for hypotheses (1), (2), and (3), respectively.

Let α be the significance level and $F_{\nu, n-JK, 1-\alpha}$ be the $(1 - \alpha)$th quantile of an F distribution with ν 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 ν numerator and $n - JK$ denominator degrees of freedom and a noncentrality parameter λ given by

$$\lambda = n(\mathbf{C}\mathbf{b})'\{\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}'\}^{-1}(\mathbf{C}\mathbf{b})/\sigma_e^2$$
$$= n\delta^2$$

where the matrix $\ddot{\mathbf{X}}$ contains the unique rows of $\mathbf{X}$ such that $\boldsymbol{\mu} = \ddot{\mathbf{X}}\mathbf{b}$, $\mathbf{W} = \text{diag}(w_1, \ldots, w_{JK})$, and δ 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 $\ddot{\mathbf{X}}$ 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 3×3 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 $\widetilde{w}_i$, 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 $\widetilde{w}_j n_c$. 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

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

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:

$$m_{1,1} \ m_{1,2} \ \big[\ldots \ m_{1,K} \big] \ \big[\backslash \ldots \big[\backslash \ m_{J,1} \ m_{J,2} \ \big[\ldots \ m_{J,K} \big] \big] \big]$$

m_{jk}, where $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$, is the alternative cell mean or the cell mean of the jth row (group) and kth 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*)}.

options	Description
Main	
∗ <u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
∗ <u>power</u>(*numlist*)	power; default is power(0.8)
∗ <u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
∗ <u>n</u>(*numlist*)	total sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample sizes
∗ <u>nper</u>group(*numlist*)	number of subjects per group; implies balanced design
∗ <u>n</u>#(*numlist*)	number of subjects in group #
<u>grw</u>eights(*wgtspec*)	group weights; default is one for each group, meaning equal-group sizes
<u>ngr</u>oups(#)	number of groups
<u>nrep</u>eated(#)	number of repeated measures
∗ <u>corr</u>(*numlist*)	correlation between repeated measures; one of corr() or covmatrix() is required
<u>covm</u>atrix(*matname*)	covariance between repeated measures; one of corr() or covmatrix() is required
<u>factor</u>(<u>between</u>｜<u>within</u>｜<u>bwithin</u>)	tested effect: between, within, or between–within; default is factor(between)
∗ <u>vareff</u>ect(*numlist*)	variance explained by the tested effect specified in factor()
∗ <u>varb</u>etween(*numlist*)	variance explained by the between-subjects effect; synonym for factor(between) and vareffect(*numlist*)
∗ <u>varw</u>ithin(*numlist*)	variance explained by the within-subject effect; synonym for factor(within) and vareffect(*numlist*)
∗ <u>varbw</u>ithin(*numlist*)	variance explained by the between–within effect; synonym for factor(bwithin) and vareffect(*numlist*)
∗ <u>varer</u>ror(*numlist*)	error variance; default is varerror(1) when corr() is specified; not allowed with covmatrix()
<u>showmat</u>rices	display cell-means matrix and covariance matrix
<u>showm</u>eans	display cell means
<u>par</u>allel	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>[(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>sav</u>ing(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
Graph	
<u>graph</u>[(*graphopts*)]	graph results; see [PSS] **power, graph**

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 iterate(500)
<u>tol</u>erance(#)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(#)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	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.

wgtspec	Description
$\#_1 \; \#_2 \; \ldots \; \#_J$	J group weights. Weights must be positive and must be integers unless option nfractional is specified. Multiple values for each group weight $\#_j$ can be specifed as a *numlist* enclosed in parentheses.
matname	matrix with J columns containing J group weights. Multiple rows are allowed, in which case each row corresponds to a different set of J weights or, equivalently, column j corresponds to a *numlist* for the jth 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).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	total number of subjects	N
N_per_group	number of subjects per group	N/N_g
N_avg	average number of subjects per group	N_{avg}
N#	number of subjects in group #	$N_\#$
delta	effect size	δ
N_g	number of groups	N_g
N_rep	number of repeated measurements	N_{rep}
m#$_1$_#$_2$	cell mean (#$_1$, #$_2$): group #$_1$, occasion #$_2$	$\mu_{\#_1,\#_2}$
Var_b	between-subjects variance	σ_b^2
Var_w	within-subject variance	σ_w^2
Var_bw	between–within (group-by-occasion) variance	σ_{bw}^2
Var_be	between-subjects error variance	σ_{be}^2
Var_we	within-subject error variance	σ_{we}^2
Var_bwe	between–within (group-by-occasion) error variance	σ_{bwe}^2
Var_e	error variance	σ_e^2
corr	correlation between repeated measures	ρ
grwgt#	group weight #	$w_\#$
target	target parameter; synonym for target effect variance	
_all	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.

Menu

Statistics > Power and sample size

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.

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). If varerror() is not specified with effect-size determination, only effect size δ is reported, otherwise the estimated variance of the tested effect is also reported.

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(*tablespec*), 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 δ 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' 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, 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. If varerror() or covmatrix() is specified, the target variance explained by the tested effect is reported in addition to the effect size.

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 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 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 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 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 is computed by default. If the error variance is specified in `varerror()`, the target variance of the effect is also reported.

▷ Example 9: Effect size for a two-way repeated-measures ANOVA

Continuing 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/r13/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	232.733333	24.76	0.0000
Residual	112.8	12	9.4		
Total	1491.8	19	78.5157895		

```
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
```

| | | | | — Prob > F — | | |
Source	df	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

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N_avg)	average sample size
r(N#$_1$_#$_2$)	number of subjects in group (#$_1$, #$_2$)
r(N_per_group)	number of subjects per group

r(N_g)	number of groups	
r(nfractional)	1 if nfractional is specified; 0 otherwise	
r(balanced)	1 for a balanced design; 0 otherwise	
r(grwgt#$_1$_#$_2$)	group weight (#$_1$, #$_2$)	
r(N_rep)	number of rows	
r(m#$_1$_#$_2$)	cell mean (#$_1$, #$_2$)	
r(Var_b)	between-subjects variance	
r(Var_w)	within-subject variance	
r(Var_bw)	between–within subjects, interaction variance	
r(Var_be)	between-subjects error variance	
r(Var_we)	within-subject error variance	
r(Var_bwe)	between–within subjects, interaction error variance	
r(Var_e)	error variance	
r(spherical)	1 covariance is spherical; 0 otherwise	
r(epsilon)	nonsphericity correction	
r(epsilon_m)	mean nonsphericity correction	
r(separator)	number of lines between separator lines in the table	
r(divider)	1 if divider is requested in the table; 0 otherwise	
r(init)	initial value for sample size or effect size	
r(maxiter)	maximum number of iterations	
r(iter)	number of iterations performed	
r(tolerance)	requested parameter tolerance	
r(deltax)	final parameter tolerance achieved	
r(ftolerance)	requested distance of the objective function from zero	
r(function)	final distance of the objective function from zero	
r(converged)	1 if iteration algorithm converged; 0 otherwise	

Macros

r(type)	test
r(method)	repeated
r(factor)	between, within, or bwithin
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
r(means)	cell-means matrix
r(Cov)	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

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} \tag{1}$$

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 $\mathbf{\Sigma}$. 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

$$
\begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & y_{1,4} \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} \\ . & . & . & . \\ . & . & . & . \\ y_{n,1} & y_{n,2} & y_{n,3} & y_{n,4} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ . & . & . \\ . & . & . \\ 1 & x_{n,1} & x_{n,2} \end{bmatrix} \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} & \alpha_{1,4} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} & \alpha_{2,4} \end{bmatrix} + \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \\ . \\ . \\ \epsilon'_n \end{pmatrix}
$$

where $y_{i,k}$ is the response of the ith 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: μ_k is the mean-treatment response at time period k, $\alpha_{j,k}$ is the jth 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 $\boldsymbol{\epsilon}_i$ are independent normal vectors of length K with mean $\mathbf{0}$ and variance–covariance $\boldsymbol{\Sigma}$.

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

$$
H_0\colon \boldsymbol{\Theta} = \mathbf{0} \qquad H_a\colon \boldsymbol{\Theta} \neq \mathbf{0}
$$

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 $\mathrm{rank}(\mathbf{C}) = d_c \leq p$, and $\mathbf{U}$ is a $q \times d_u$ matrix of full rank, where $\mathrm{rank}(\mathbf{U}) = d_u \leq q$.

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

$$
\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}
$$

$$
\widehat{\boldsymbol{\Theta}} = \mathbf{C}\widehat{\mathbf{B}}\mathbf{U} \tag{2}
$$

$$
\widehat{\mathbf{H}} = \widehat{\boldsymbol{\Theta}}'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}\widehat{\boldsymbol{\Theta}} \tag{3}
$$

Define $\widehat{\mathbf{E}} = \mathbf{U}'\widehat{\boldsymbol{\Sigma}}\mathbf{U}(n-p)$. Then, under the assumption of sphericity, the test statistic is given by

$$
F_{C,U} = \frac{\mathrm{tr}(\widehat{\mathbf{H}})/d_c d_u}{\mathrm{tr}(\widehat{\mathbf{E}})/\{d_u(n-p)\}} \tag{4}
$$

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

$$\varepsilon = \frac{\text{tr}^2(\widehat{\boldsymbol{\Sigma}})}{d_u \, \text{tr}(\widehat{\boldsymbol{\Sigma}}^2)} = \frac{\left(\sum_{k=1}^{d_u} \lambda_k\right)^2}{d_u \sum_{k=1}^{d_u} \lambda_k^2}$$

Under the alternative hypothesis, the power is obtained using a noncentral F distribution with noncentrality parameter equal to

$$\lambda = d_c d_u \epsilon F_{C,U}$$

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

$$\boldsymbol{\Theta} = \mathbf{CBU}$$
$$\mathbf{H} = n\boldsymbol{\Theta}'\{\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}'\}^{-1}\boldsymbol{\Theta}$$
$$= n\mathbf{H}_*$$
$$\mathbf{E} = \mathbf{U}'\boldsymbol{\Sigma}\mathbf{U}(n-p)$$
$$= \boldsymbol{\Sigma}_*(n-p)$$

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 jth 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

$$F_{C,U} = \frac{\text{tr}(\mathbf{H})/d_c d_u}{\text{tr}(\mathbf{E})/\{d_u(n-p)\}}$$
$$= \frac{n}{d_c} \frac{\text{tr}(\mathbf{H}_*)}{\text{tr}(\boldsymbol{\Sigma}_*)}$$

from which we obtain the noncentrality parameter as

$$\lambda = d_u d_c \epsilon F_{C,U}$$
$$= n\epsilon\delta^2 \tag{5}$$

where the effect size δ is defined as $\delta = \sqrt{d_u \text{tr}(\mathbf{H}_*)/\text{tr}(\boldsymbol{\Sigma}_*)}$.

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 λ from (5).

The power of the overall F test is

$$1 - \beta = F_{d_c d_u \epsilon, d_u(n-p)\epsilon, \lambda}\left(F_{d_c d_u \epsilon_m, d_u(n-p)\epsilon_m, 1-\alpha}\right) \tag{6}$$

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 $\widetilde{w}_j n_c$, where $\widetilde{w}$ 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. Cushman, 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

Syntax

Two samples, compute sample size for unbalanced designs

  Compute total sample size

    power ..., <u>nrat</u>io(*numlist*) $\big[$<u>nfrac</u>tional$\big]$ ...

  Compute one group size given the other

    power ..., n#(*numlist*) compute(n1 | n2) $\big[$<u>nfrac</u>tional$\big]$ ...

Two samples, specify sample size for unbalanced designs

  Specify total sample size and allocation ratio

    power ..., n(*numlist*) <u>nrat</u>io(*numlist*) $\big[$<u>nfrac</u>tional$\big]$ ...

  Specify one of the group sizes and allocation ratio

    power ..., n#(*numlist*) <u>nrat</u>io(*numlist*) $\big[$<u>nfrac</u>tional$\big]$ ...

  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
*n(*numlist*)	total sample size; required to compute power or effect size
*n1(*numlist*)	sample size of the control group
*n2(*numlist*)	sample size of the experimental group
*<u>nrat</u>io(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
compute(n1 | n2)	solve for N_1 given N_2 or for N_2 given N_1
<u>nfrac</u>tional	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**.

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.

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, N_2/N_1 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*
> *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
```

◁

Also see

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

[PSS] **Glossary**

Glossary

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.

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.

actual alpha, **actual significance level**. This is an attained or observed significance level.

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, α, 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, β, 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*.

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, χ^2 **test**. This test for which either an asymptotic sampling distribution or a sampling distribution of a test statistic is χ^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.

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.

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 2×2 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. 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, δ, in the context of power and sample-size calculations, denotes the effect size.

directional test. See *one-sided test*.

discordant pairs. In a 2×2 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. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

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.

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 2×2 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 study. See *cohort study*.

Greenhouse–Geisser correction. See *nonsphericity correction*.

H_0. See *null hypothesis*.

H_a. See *alternative hypothesis*.

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**.

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.

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 an observation from another group with respect to one or more characteristics of interest. Also see *paired 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 2×2 contingency table as a hypothesis of marginal homogeneity. See [PSS] **power pairedproportions** and [ST] **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 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 ρ, 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 β 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 θ, defined as the probability that the observed sample belongs to the rejection region of a test for given θ. 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*.

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 α 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.

symmetry. In a 2×2 contingency table, symmetry refers to the equality of the off-diagonal elements. For a 2×2 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 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 I error probability. See *probability of a type I error*.

type II error probability. See *probability of a type II error*.

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**.

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 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

This is the subject and author index for the *Power and Sample-Size Reference Manual*. Readers interested in topics other than power and sample size should see the combined subject index (and the combined author index) in the *Glossary and Index*.

Symbols

α, [PSS] **intro**, [PSS] **GUI**, [PSS] **power**, [PSS] **power, graph**, [PSS] **power, table**, [PSS] **power onemean**, [PSS] **power twomeans**, [PSS] **power pairedmeans**, [PSS] **power oneproportion**, [PSS] **power twoproportions**, [PSS] **power pairedproportions**, [PSS] **power onevariance**, [PSS] **power twovariances**, [PSS] **power onecorrelation**, [PSS] **power twocorrelations**, [PSS] **power oneway**, [PSS] **power twoway**, [PSS] **power repeated**, [PSS] **unbalanced designs**, [PSS] **Glossary**

β, [PSS] **intro**, [PSS] **power**, [PSS] **power, graph**, [PSS] **power, table**, [PSS] **power onemean**, [PSS] **power twomeans**, [PSS] **power pairedmeans**, [PSS] **power oneproportion**, [PSS] **power twoproportions**, [PSS] **power pairedproportions**, [PSS] **power onevariance**, [PSS] **power twovariances**, [PSS] **power onecorrelation**, [PSS] **power twocorrelations**, [PSS] **power oneway**, [PSS] **power twoway**, [PSS] **power repeated**, [PSS] **Glossary**

δ, [PSS] **intro**, [PSS] **GUI**, [PSS] **power**, [PSS] **power, graph**, [PSS] **power, table**, [PSS] **power onemean**, [PSS] **power twomeans**, [PSS] **power pairedmeans**, [PSS] **power oneproportion**, [PSS] **power twoproportions**, [PSS] **power pairedproportions**, [PSS] **power onevariance**, [PSS] **power twovariances**, [PSS] **power onecorrelation**, [PSS] **power twocorrelations**, [PSS] **power oneway**, [PSS] **power twoway**, [PSS] **power repeated**, [PSS] **Glossary**

A

acceptance region, [PSS] **intro**, [PSS] **Glossary**
actual
    alpha, [PSS] **Glossary**, *also see* significance level, actual
    power, *see* power, actual
    sample size, *see* sample size, actual
    significance level, *see* significance level, actual
Agresti, A., [PSS] **intro**, [PSS] **power oneproportion**, [PSS] **power twoproportions**, [PSS] **power pairedproportions**
Alderman, M. H., [PSS] **intro**, [PSS] **power repeated**
ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group, [PSS] **intro**, [PSS] **power repeated**
allocation ratio, [PSS] **intro**, [PSS] **power twomeans**, [PSS] **power twoproportions**, [PSS] **power twovariances**, [PSS] **power twocorrelations**, [PSS] **unbalanced designs**, [PSS] **Glossary**

alpha, *see* α
alternative
    correlation, [PSS] **intro**, [PSS] **power**, [PSS] **power onecorrelation**
    hypothesis, *see* null hypothesis and alternative hypothesis
    mean, [PSS] **intro**, [PSS] **GUI**, [PSS] **power**, [PSS] **power, graph**, [PSS] **power, table**, [PSS] **power onemean**, [PSS] **unbalanced designs**
    mean difference, [PSS] **intro**, [PSS] **power**, [PSS] **power pairedmeans**
    parameter, *see* alternative value
    proportion, [PSS] **intro**, [PSS] **power**, [PSS] **power oneproportion**
    space, [PSS] **intro**
    standard deviation, [PSS] **intro**, [PSS] **power**, [PSS] **power onevariance**
    value, [PSS] **Glossary**, *also see* postulated value
    variance, [PSS] **intro**, [PSS] **power**, [PSS] **power onevariance**
analysis of variance, [PSS] **power**, [PSS] **power oneway**, [PSS] **power twoway**, [PSS] **power repeated**, [PSS] **Glossary**
    repeated measures, [PSS] **power repeated**
Anderson, T. W., [PSS] **intro**, [PSS] **power onecorrelation**, [PSS] **power twocorrelations**
ANOVA, *see* analysis of variance
Armitage, P., [PSS] **intro**, [PSS] **power twomeans**, [PSS] **power pairedmeans**

B

balanced design, [PSS] **intro**, [PSS] **power twomeans**, [PSS] **power twoproportions**, [PSS] **power twovariances**, [PSS] **power twocorrelations**, [PSS] **power oneway**, [PSS] **power twoway**, [PSS] **power repeated**, [PSS] **unbalanced designs**, [PSS] **Glossary**
Barrett, J. H., [PSS] **intro**
Barthel, F. M.-S., [PSS] **intro**
Barton, C. N., [PSS] **power repeated**
Batistatou, E., [PSS] **power**
Berger, R. L., [PSS] **intro**
Berry, G., [PSS] **intro**, [PSS] **power twomeans**, [PSS] **power pairedmeans**
beta, *see* β
between-group variance, [PSS] **power oneway**
between-subjects
    design, [PSS] **power oneway**, [PSS] **power twoway**, [PSS] **power repeated**, [PSS] **Glossary**
    factor, [PSS] **power repeated**, [PSS] **Glossary**
    variance, [PSS] **power repeated**
binary outcome, [PSS] **intro**, [PSS] **GUI**, [PSS] **power**, [PSS] **power oneproportion**, [PSS] **power twoproportions**, [PSS] **power pairedproportions**
binomial test, [PSS] **power oneproportion**, [PSS] **Glossary**
bisection method, *see* iteration, bisection method

Z