

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

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

Description

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

Quick start

Sample size for a test of $H_0: \pi_1 = \pi_2$ vs $H_a: \pi_1 \neq \pi_2$ given alternative control-group proportion $p_1 = 0.8$, alternative experimental-group proportion $p_2 = 0.65$ with default power of 0.8 and significance level $\alpha = 0.05$

```
power twoproportions .8 .65
```

Same as above, but specified as p_1 and difference between proportions $p_2 - p_1 = -0.15$

```
power twoproportions .8, diff(-.15)
```

Same as above, but specified as $p_1 = 0.8$ and ratio $p_2/p_1 = 0.8125$

```
power twoproportions .8, ratio(.8125)
```

Same as above, but specified as $p_1 = 0.8$ and odds ratio $\{p_2/(1-p_2)\}/\{p_1/(1-p_1)\} = 0.464$

```
power twoproportions .8, oratio(.464)
```

Power for sample sizes of 50 and 80 in groups 1 and 2, respectively

```
power twoproportions 0.8 0.65, n1(50) n2(80)
```

Power for total sample sizes of 150, 170, 190, 210, and 230

```
power twoproportions .8 .65, n(150(20)230)
```

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

```
power twoproportions .8 .65, n(150(20)230) graph
```

As above, but with difference equal to 0.1, 0.15, and 0.2

```
power twoproportions .65, n(150(20)200) graph diff(.1(.05).2)
```

Display results in a table showing total sample size, difference, and power

```
power twoproportions .65, n(150(20)200) table(N diff power)
```

Effect size and target p_2 for $p_1 = 0.6$ with sample size of 200, power of 0.8, and $\alpha = 0.01$

```
power twoproportions .6, n(200) power(.8) alpha(.01)
```

Menu

Statistics > Power and sample size

Syntax

Compute sample size

```
power twoproportions  $p_1$   $p_2$  [ , power(numlist) options ]
```

Compute power

```
power twoproportions  $p_1$   $p_2$  , n(numlist) [options ]
```

Compute effect size and experimental-group proportion

```
power twoproportions  $p_1$  , n(numlist) power(numlist) [options ]
```

where p_1 is the proportion in the control (reference) group, and p_2 is the proportion in the experimental (comparison) group. p_1 and p_2 may each be specified either as one number or as a list of values in parentheses (see [U] [11.1.8 numlist](#)).

<i>options</i>	Description
<code>test(<i>test</i>)</code>	specify the type of test; default is <code>test(chi2)</code>
Main	
* <code>alpha(<i>numlist</i>)</code>	significance level; default is <code>alpha(0.05)</code>
* <code>power(<i>numlist</i>)</code>	power; default is <code>power(0.8)</code>
* <code>beta(<i>numlist</i>)</code>	probability of type II error; default is <code>beta(0.2)</code>
* <code>n(<i>numlist</i>)</code>	total sample size; required to compute power or effect size
* <code>n1(<i>numlist</i>)</code>	sample size of the control group
* <code>n2(<i>numlist</i>)</code>	sample size of the experimental group
* <code>nratio(<i>numlist</i>)</code>	ratio of sample sizes, $N2/N1$; default is <code>nratio(1)</code> , meaning equal group sizes
<code>compute(n1 n2)</code>	solve for $N1$ given $N2$ or for $N2$ given $N1$
<code>nfractional</code>	allow fractional sample sizes
* <code>diff(<i>numlist</i>)</code>	difference between the experimental-group and control-group proportions, $p_2 - p_1$; specify instead of the experimental-group proportion p_2
* <code>ratio(<i>numlist</i>)</code>	ratio of the experimental-group proportion to the control-group proportion, p_2/p_1 ; specify instead of the experimental-group proportion p_2
* <code>rdiff(<i>numlist</i>)</code>	risk difference, $p_2 - p_1$; synonym for <code>diff()</code>
* <code>rrisk(<i>numlist</i>)</code>	relative risk, p_2/p_1 ; synonym for <code>ratio()</code>
* <code>oratio(<i>numlist</i>)</code>	odds ratio, $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$
<code>effect(<i>effect</i>)</code>	specify the type of effect to display; default is <code>effect(diff)</code>
<code>continuity</code>	apply continuity correction to the normal approximation of the discrete distribution
<code>direction(<u>upper</u> <u>lower</u>)</code>	direction of the effect for effect-size determination; default is <code>direction(upper)</code> , which means that the postulated value of the parameter is larger than the hypothesized value
<code>onesided</code>	one-sided test; default is two sided
<code>parallel</code>	treat number lists in starred options or in command arguments as parallel when multiple values per option or argument are specified (do not enumerate all possible combinations of values)
Table	
<code>[no]table [(<i>tablespec</i>)]</code>	suppress table or display results as a table; see [PSS] power, table
<code>saving(<i>filename</i> [, replace])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph [(<i>graphopts</i>)]</code>	graph results; see [PSS] power, graph

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

*Specifying a list of values in at least two starred options, or at least two command arguments, or at least one starred option and one argument results in computations for all possible combinations of the values; see [U] 11.1.8 `numlist`. Also see the `parallel` option.

`notitle` does not appear in the dialog box.

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

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

<i>effect</i>	Description
<code>diff</code>	difference between proportions, $p_2 - p_1$; the default
<code>ratio</code>	ratio of proportions, p_2/p_1
<code>rdiff</code>	risk difference, $p_2 - p_1$
<code>rrisk</code>	relative risk, p_2/p_1
<code>oratio</code>	odds ratio, $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$

where *tablespec* is

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

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

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	α
<code>alpha_a</code>	observed significance level	α_a
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	β
<code>N</code>	total number of subjects	N
<code>N1</code>	number of subjects in the control group	N_1
<code>N2</code>	number of subjects in the experimental group	N_2
<code>nratio</code>	ratio of sample sizes, experimental to control	N_2/N_1
<code>delta</code>	effect size	δ
<code>p1</code>	control-group proportion	p_1
<code>p2</code>	experimental-group proportion	p_2
<code>diff</code>	difference between the experimental-group proportion and the control-group proportion	$p_2 - p_1$
<code>ratio</code>	ratio of the experimental-group proportion to the control-group proportion	p_2/p_1
<code>rdiff</code>	risk difference	$p_2 - p_1$
<code>rrisk</code>	relative risk	p_2/p_1
<code>oratio</code>	odds ratio	θ
<code>target</code>	target parameter; synonym for <code>p2</code>	
<code>_all</code>	display all supported columns	

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

Column `alpha_a` is available when the `test(fisher)` option is specified.

Columns `nratio`, `diff`, `ratio`, `rdiff`, `rrisk`, and `oratio` are shown in the default table if specified.

Options

`test(test)` specifies the type of the test for power and sample-size computations. *test* is one of `chi2`, `lrchi2`, or `fisher`.

`chi2` requests computations for the Pearson's 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.

Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `compute()`, `nfractional`; see [PSS] [power](#).

`diff(numlist)` specifies the difference between the experimental-group proportion and the control-group proportion, $p_2 - p_1$. You can specify either the experimental-group proportion p_2 as a command argument or the difference between the two proportions in `diff()`. If you specify `diff(#)`, the experimental-group proportion is computed as $p_2 = p_1 + \#$. This option is not allowed with the effect-size determination and may not be combined with `ratio()`, `rdiff()`, `rrisk()`, or `oratio()`.

`ratio(numlist)` specifies the ratio of the experimental-group proportion to the control-group proportion, p_2/p_1 . You can specify either the experimental-group proportion p_2 as a command argument or

the ratio of the two proportions in `ratio()`. If you specify `ratio(#)`, the experimental-group proportion is computed as $p_2 = p_1 \times \#$. This option is not allowed with the effect-size determination and may not be combined with `diff()`, `rdiff()`, `rrisk()`, or `oratio()`.

`rdiff(numlist)` specifies the risk difference $p_2 - p_1$. This is a synonym for the `diff()` option, except the results are labeled as risk differences. This option is not allowed with the effect-size determination and may not be combined with `diff()`, `ratio()`, `rrisk()`, or `oratio()`.

`rrisk(numlist)` specifies the relative risk or risk ratio $p_2 - p_1$. This is a synonym for the `ratio()` option, except the results are labeled as relative risks. This option is not allowed with the effect-size determination and may not be combined with `diff()`, `ratio()`, `rdiff()`, or `oratio()`.

`oratio(numlist)` specifies the odds ratio $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$. You can specify either the experimental-group proportion p_2 as a command argument or the odds ratio in `oratio()`. If you specify `oratio(#)`, the experimental-group proportion is computed as $p_2 = 1/\{1 + (1 - p_1)/(p_1 \times \#)\}$. This option is not allowed with the effect-size determination and may not be combined with `diff()`, `ratio()`, `rdiff()`, or `rrisk()`.

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. *effect* is one of `diff`, `ratio`, `rdiff`, `rrisk`, or `oratio`. By default, the effect size `delta` is the difference between proportions. If `diff()`, `ratio()`, `rdiff()`, `rrisk()`, or `oratio()` is specified, the effect size `delta` will contain the effect corresponding to the specified option. For example, if `oratio()` is specified, `delta` will contain the odds ratio.

`continuity` requests that continuity correction be applied to the normal approximation of the discrete distribution. `continuity` cannot be specified with `test(fisher)` or `test(lrchi2)`.

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

Table

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

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

Graph

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

Iteration

`init(#)` specifies the initial value for the estimated parameter. For sample-size determination, the estimated parameter is either the control-group size n_1 or, if `compute(n2)` is specified, the experimental-group size n_2 . For the effect-size determination, the estimated parameter is the experimental-group proportion p_2 . The default initial values for sample sizes for a two-sided test are based on the corresponding one-sided large-sample z test with the significance level $\alpha/2$. The default initial value for the experimental-group proportion is computed using the bisection method.

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

The following option is available with `power twoproportions` but is not shown in the dialog box: `notitle`; see [PSS] [power](#).

Remarks and examples

Remarks are presented under the following headings:

Introduction

Using power twoproportions

Alternative ways of specifying effect

Computing sample size

Computing power

Computing effect size and experimental-group proportion

Testing a hypothesis about two independent proportions

Video examples

This entry describes the `power twoproportions` command and the methodology for power and sample-size analysis for a two-sample proportions test. See [\[PSS\] intro](#) for a general introduction to power and sample-size analysis and [\[PSS\] power](#) for a general introduction to the `power` command using hypothesis tests.

Introduction

The comparison of two independent proportions arises in studies involving two independent binomial populations. There are many examples of studies where a researcher would like to compare two independent proportions. A pediatrician might be interested in the relationship between low birthweight and the mothers' use of a particular drug during pregnancy. He or she would like to test the null hypothesis that there is no difference in the proportion of low-birthweight babies for mothers who took the drug and mothers who did not. A drug manufacturer may want to test the developed new topical treatment for a foot fungus by testing the null hypothesis that the proportion of successfully treated patients is the same in the treatment and placebo groups.

Hypothesis testing of binomial outcomes relies on a set of assumptions: 1) a Bernoulli outcome is observed a fixed number of times; 2) the probability p of observing an event of interest in one trial is fixed across all trials; and 3) individual trials are independent. Each of the two populations must conform to the assumptions of a binomial distribution.

This entry describes power and sample-size analysis for the inference about two population proportions performed using hypothesis testing. Specifically, we consider the null hypothesis $H_0: p_2 = p_1$ versus the two-sided alternative hypothesis $H_a: p_2 \neq p_1$, the upper one-sided alternative $H_a: p_2 > p_1$, or the lower one-sided alternative $H_a: p_2 < p_1$.

The large-sample Pearson's 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 p1 p2, ...
```

For this specification, the default effect size `delta` displayed by the `power` command is the difference $p_2 - p_1$ between the proportions. You can use the `effect()` option to request another type of effect. For example, if you specify `effect(oratio)`,

```
power twoproportions p1 p2, effect(oratio) ...
```

the effect size `delta` will correspond to the odds ratio.

Alternatively, you can specify the control-group proportion p_1 and one of the options `diff()`, `ratio()`, `rdiff()`, `rrisk()`, or `oratio()`. For these specifications, the effect size `delta` will contain the effect corresponding to the option. If desired, you can change this by specifying the `effect()` option.

Specify difference $p_2 - p_1$ between the two proportions:

```
power twoproportions p1, diff(numlist) ...
```

Specify risk difference $p_2 - p_1$:

```
power twoproportions p1, rdiff(numlist) ...
```

Specify ratio p_2/p_1 of the two proportions:

```
power twoproportions p1, ratio(numlist) ...
```

Specify relative risk or risk ratio p_2/p_1 :

```
power twoproportions p1, rrisk(numlist) ...
```

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

```
power twoproportions p1, oratio(numlist) ...
```

In the following sections, we describe the use of `power twoproportions` accompanied by examples for computing sample size, power, and experimental-group proportions.

Computing sample size

To compute sample size, you must specify the control-group proportion p_1 , the experimental-group proportion p_2 , and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

► Example 1: Sample size for a two-sample proportions test

Consider a study investigating the effectiveness of aspirin in reducing the mortality rate due to myocardial infarction (heart attacks). Let p_A denote the proportion of deaths for aspirin users in the population and p_N denote the corresponding proportion for nonusers. We are interested in testing the null hypothesis $H_0: p_A - p_N = 0$ against the two-sided alternative hypothesis $H_a: p_A - p_N \neq 0$.

Previous studies indicate that the proportion of deaths due to heart attacks is 0.015 for nonusers and 0.001 for users. Investigators wish to determine the minimum sample size required to detect an absolute difference of $|0.001 - 0.015| = 0.014$ with 80% power using a two-sided 5%-level test.

To compute the required sample size, we specify the values 0.015 and 0.001 as the control- and experimental-group proportions after the command name. We omit options `alpha(0.05)` and `power(0.8)` because the specified values are their defaults.

```
. power twoproportions 0.015 0.001
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.0140 (difference)
      p1 =     0.0150
      p2 =     0.0010
Estimated sample sizes:
      N =      1270
      N per group =    635
```

A total sample of 1,270 individuals, 635 individuals per group, must be obtained to detect an absolute difference of 0.014 between proportions of aspirin users and nonusers with 80% power using a two-sided 5%-level Pearson's 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(oratio)` option.

```
. power twoproportions 0.015 0.001, effect(oratio)
Performing iteration ...
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.0657 (odds ratio)
      p1 =     0.0150
      p2 =     0.0010
Estimated sample sizes:
      N =      1270
      N per group =    635
```

The effect size `delta` now contains the estimated odds ratio and is labeled correspondingly.

Instead of the estimate of the proportion in the experimental group, we may have an estimate of the odds ratio $\{p_2(1 - p_1)\}/\{p_1(1 - p_2)\}$. 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

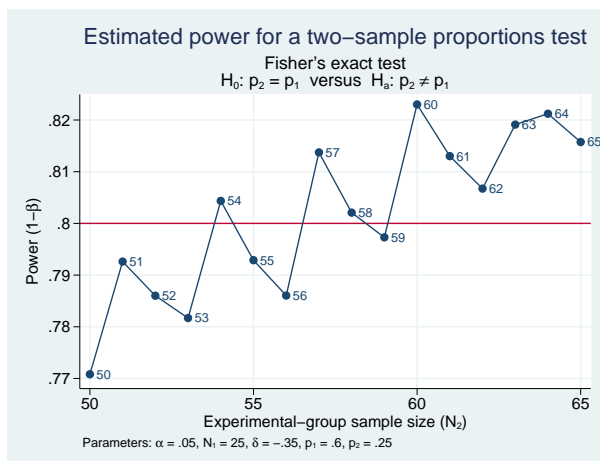


Figure 1. Saw-toothed power function

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.

◀

Video examples

[How to calculate sample size for two independent proportions](#)

[How to calculate power for two independent proportions](#)

[How to calculate minimum detectable effect size for two independent proportions](#)

Stored results

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

Scalars

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

Macros

<code>r(type)</code>	<code>test</code>
<code>r(method)</code>	<code>twoproportions</code>
<code>r(test)</code>	<code>chi2</code> , <code>lrchi2</code> , or <code>fisher</code>
<code>r(effect)</code>	specified effect: <code>diff</code> , <code>ratio</code> , etc.
<code>r(direction)</code>	<code>upper</code> or <code>lower</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

Matrix

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

Methods and formulas

Consider two independent samples of sizes n_1 and n_2 . Let x_{11}, \dots, x_{1n_1} be a random sample from a binomial distribution with the success probability p_1 . We refer to this sample as a control group. Let x_{21}, \dots, x_{2n_2} be a random sample from a binomial distribution with the success probability p_2 , the experimental group. Let $x_{ij} = 1$ denote a success and $x_{ij} = 0$ denote a failure. The two samples are assumed to be independent.

The sample proportions and their respective standard errors are

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

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

A two-sample proportions test involves testing the null hypothesis $H_0: p_2 = p_1$ versus the two-sided alternative hypothesis $H_a: p_2 \neq p_1$, the upper one-sided alternative $H_a: p_2 > p_1$, or the lower one-sided alternative $H_a: p_2 < p_1$.

Let $R = n_2/n_1$ denote the allocation ratio. Then $n_2 = R \times n_1$ and power can be viewed as a function of n_1 . Therefore, for sample-size determination, the control-group sample size n_1 is computed first. The experimental-group size n_2 is then computed as $R \times n_1$, and the total sample size is computed as $n = n_1 + n_2$. By default, sample sizes are rounded to integer values; see *Fractional sample sizes* in [PSS] **unbalanced designs** for details.

The formulas below are based on [Fleiss, Levin, and Paik \(2003\)](#) and [Agresti \(2013\)](#).

Methods and formulas are presented under the following headings:

Effect size
Pearson's χ^2 test
Likelihood-ratio test
Fisher's exact conditional test

Effect size

The measures of risk or effect size can be defined in a number of ways for the two-sample proportions test. By default, the effect size is the difference between the experimental-group proportion and the control-group proportion. Other available risk measures include risk difference, risk ratio or relative risk, and odds ratio.

By default or when `effect(diff)` or `effect(rdif)` is specified, the effect size is computed as

$$\delta = p_2 - p_1$$

When `effect(ratio)` or `effect(rrisk)` is specified, the effect size is computed as

$$\delta = p_2/p_1$$

When `effect(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 χ^2 test

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

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

is standard normal, where $\bar{p} = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$ is the pooled proportion and \widehat{p} is its estimator. The square of this statistic, z^2 , has an approximate 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 β th quantiles of the standard normal distribution.

Let $\sigma_D = \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$ be the standard deviation of the difference between proportions and $\sigma_p = \sqrt{\bar{p}(1 - \bar{p})(1/n_1 + 1/n_2)}$ be the pooled standard deviation.

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

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

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, and c is the normal-approximation continuity correction. For equal sample sizes, $n_1 = n_2 = n/2$, the continuity correction is expressed as $c = 2/n$ (Levin and Chen 1999).

For a one-sided test, given the allocation ratio $R = n_2/n_1$, the total sample size n is computed by inverting the corresponding power equation in (1),

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

where $w_1 = 1/(1 + R)$ and $w_2 = R/(1 + R)$. Then n_1 and n_2 are computed as $n_1 = n/(1 + R)$ and $n_2 = R \times n_1$, respectively. If the `continuity` option is specified, the sample size n_c for a one-sided test is computed as

$$n_c = \frac{n}{4} \left(1 + \sqrt{1 + \frac{2}{n w_1 w_2 |p_2 - p_1|}} \right)^2$$

where n is the sample size computed without the correction. For unequal sample sizes, the continuity correction generalizes to $c = 1/(2nw_1w_2)$ (Fleiss, Levin, and Paik 2003).

For a two-sided test, the sample size is computed by iteratively solving the two-sided power equation in (1). The initial values for the two-sided computations are obtained from (2) with the significance level $\alpha/2$.

If one of the group sizes is known, the other one is computed by iteratively solving the corresponding power equation in (1). The initial values are obtained from (2) by assuming that $R = 1$.

The experimental-group proportion p_2 is computed by iteratively solving the corresponding power equations in (1). The default initial values are obtained by using a bisection method.

Likelihood-ratio test

Let $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, and $\bar{q} = 1 - \bar{p} = 1 - (n_1p_1 + n_2p_2)/(n_1 + n_2)$. The likelihood-ratio test statistic is given by

$$G = \sqrt{2n \left\{ \frac{n_1p_1}{n} \ln \left(\frac{p_1}{\bar{p}} \right) + \frac{n_1q_1}{n} \ln \left(\frac{q_1}{\bar{q}} \right) + \frac{n_2p_2}{n} \ln \left(\frac{p_2}{\bar{p}} \right) + \frac{n_2q_2}{n} \ln \left(\frac{q_2}{\bar{q}} \right) \right\}}$$

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

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

For a one-sided test, given the allocation ratio R , the total sample size n is computed by inverting the corresponding power equation in (3),

$$n = \frac{(z_{1-\alpha} - z_\beta)^2}{2 \left\{ w_1p_1 \ln \left(\frac{p_1}{\bar{p}} \right) + w_1q_1 \ln \left(\frac{q_1}{\bar{q}} \right) + w_2p_2 \ln \left(\frac{p_2}{\bar{p}} \right) + w_2q_2 \ln \left(\frac{q_2}{\bar{q}} \right) \right\}} \quad (4)$$

where $w_1 = 1/(1 + R)$ and $w_2 = R/(1 + R)$. Then n_1 and n_2 are computed as $n_1 = n/(1 + R)$ and $n_2 = R \times n_1$, respectively.

For a two-sided test, the sample size is computed by iteratively solving the two-sided power equation in (3). The initial values for the two-sided computations are obtained from equation (4) with the significance level $\alpha/2$.

If one of the group sizes is known, the other one is computed by iteratively solving the corresponding power equation in (3). The initial values are obtained from (4) by assuming that $R = 1$.

The experimental-group proportion p_2 is computed by iteratively solving the corresponding power equations in (3). The default initial values are obtained by using a bisection method.

Fisher's exact conditional test

Power computation for Fisher's exact test is based on [Casagrande, Pike, and Smith \(1978\)](#). We present formulas from the original paper with a slight change in notation: we use \tilde{p}_1 and \tilde{p}_2 in place of p_1 and p_2 and 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, $\tilde{p}_1 = p_2$ is the proportion in the experimental group, $\tilde{p}_2 = p_1$ is the proportion in the control group, and $\eta = 1/\theta$. Also we denote $\tilde{n}_1 = n_2$ to be the sample size of the experimental group and $\tilde{n}_2 = n_1$ to be the sample size of the control group.

Let k be the number of successes in the experimental group, and let m be the total number of successes in both groups. The conditional distribution of k is given by

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

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

Assume an upper one-sided test given by

$$H_0: \tilde{p}_1 = \tilde{p}_2 \quad \text{versus} \quad H_a: \tilde{p}_1 > \tilde{p}_2 \quad (5)$$

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

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

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

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

The conditional power is

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

For a lower one-sided hypothesis $H_a: \tilde{p}_1 < \tilde{p}_2$, the corresponding hypothesis in terms of η is given by

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

The conditional power in this case is

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

where k_l is the critical value of k such that the following inequalities are satisfied:

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

For a two-sided test, the critical values k_l and k_u are calculated using the inequalities (6) and (7) with $\alpha/2$, respectively.

Finally, the unconditional power is calculated as

$$\beta(\eta) = \sum_j \beta(\eta|j)P(j)$$

where j takes the value from 0 to n , and

$$P(j) = \sum_{i=L}^U \binom{\widetilde{n}_1}{i} \widetilde{p}_1^i (1 - \widetilde{p}_1)^{\widetilde{n}_1 - i} \binom{\widetilde{n}_2}{j - i} \widetilde{p}_2^{j - i} (1 - \widetilde{p}_2)^{\widetilde{n}_2 - j + i}$$

where $L = \max(0, j - \widetilde{n}_2)$ and $U = \min(\widetilde{n}_1, j)$.

References

- Agresti, A. 2013. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- Casagrande, J. T., M. C. Pike, and P. G. Smith. 1978. The power function of the “exact” test for comparing two binomial distributions. *Journal of the Royal Statistical Society, Series C* 27: 176–180.
- Chernick, M. R., and C. Y. Liu. 2002. The saw-toothed behavior of power versus sample size and software solutions: Single binomial proportion using exact methods. *American Statistician* 56: 149–155.
- Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh: Oliver & Boyd.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley.
- Irwin, J. O. 1935. Tests of significance for differences between percentages based on small numbers. *Metron* 12: 83–94.
- Levin, B., and X. Chen. 1999. Is the one-half continuity correction used once or twice to derive a well-known approximate sample size formula to compare two independent binomial distributions? *American Statistician* 53: 62–66.

Also see

- [PSS] **power** — Power and sample-size analysis for hypothesis tests
- [PSS] **power, graph** — Graph results from the power command
- [PSS] **power, table** — Produce table of results from the power command
- [PSS] **Glossary**
- [R] **bitest** — Binomial probability test
- [R] **prtest** — Tests of proportions