Title

> **power oneproportion** — Power analysis for a one-sample proportion test

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. For power and sample-size analysis in a cluster randomized design, see [PSS] **power oneproportion, cluster**. Also see [PSS] **power** for a general introduction to the power command using hypothesis tests.

Quick start

Sample size for a test of H_0: $\pi = 0.2$ versus H_a: $\pi \neq 0.2$ with null proportion $p_0 = 0.2$, alternative proportion $p_a = 0.1$, default power of 0.8, and significance level $\alpha = 0.05$

    power oneproportion .2 .1

As above, but for p_a equal to 0.05, 0.075, 0.1, 0.125, and 0.15

    power oneproportion .2 (.05(.025).15)

As above, but display results as a graph of sample size versus alternative proportion

    power oneproportion .2 (.05(.025).15), graph

Sample size for one-sided test with power of 0.9

    power oneproportion .2 .1, power(.9) onesided

Sample size for a Wald test instead of the default score test

    power oneproportion .2 .1, test(wald)

Power for a sample size of 120

    power oneproportion .2 .1, n(120)

As above, but for sample sizes of 110, 120, 130, and 140

    power oneproportion .2 .1, n(110(10)140)

Effect size and target proportion for $p_0 = 0.3$, sample size of 75, and power of 0.8

    power oneproportion .3, n(75) power(.8)

Menu

Statistics > Power and sample size

Syntax

Compute sample size

> power oneproportion p_0 p_a $\left[\right.$, <u>power</u>(*numlist*) *options* $\left.\right]$

Compute power

> power oneproportion p_0 p_a , n(*numlist*) $\left[\right.$*options*$\left.\right]$

Compute effect size and target proportion

> power oneproportion p_0 , n(*numlist*) <u>power</u>(*numlist*) $\left[\right.$*options*$\left.\right]$

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 or in command arguments as parallel when multiple values per option or argument 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 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
cluster	perform computations for a CRD; see [PSS] **power oneproportion, cluster**
notitle	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.

cluster and notitle do 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	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.

Options

test(*test*) specifies the type of the test for power and sample-size computations. *test* is one of
   score, wald, or binomial.

   score requests computations for the score test. This is the default test.

   wald requests computations for the Wald test. This corresponds to computations using the value
   of the alternative proportion instead of the default null proportion in the formula for the standard
   error of the estimator of the proportion.

   binomial requests computations for the binomial test. The computation using the binomial
   distribution is not available for sample-size and effect-size determinations; see example 7 for
   details. Iteration options are not allowed with this test.

Main

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

diff(*numlist*) specifies the difference between the alternative proportion and the null proportion, $p_a - p_0$. You can specify either the alternative proportion p_a as a command argument or the difference between the two proportions in diff(). If you specify diff(#), the alternative proportion is computed as $p_a = p_0 + \#$. This option is not allowed with the effect-size determination.

critvalues requests that the critical values be reported when the computation is based on the binomial distribution.

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

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

Table

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

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

Graph

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

Iteration

init(#) specifies the initial value of the sample size for the sample-size determination or the initial value of the proportion for the effect-size determination.

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

The following options are available with power oneproportion but are not shown in the dialog box:

cluster; see [PSS] **power oneproportion, cluster**.

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

Remarks and examples   stata.com

Remarks are presented under the following headings:

> *Introduction*
> *Using power oneproportion*
> *Computing sample size*
> *Computing power*
> *Computing effect size and target proportion*
> *Performing hypothesis tests on proportion*
> *Video examples*

This entry describes the power oneproportion command and the methodology for power and sample-size analysis for a one-sample proportion test. See [PSS] **intro** for a general introduction to power and sample-size analysis and [PSS] **power** for a general introduction to the power command using hypothesis tests. Also see [PSS] **power oneproportion, cluster** for power and sample-size analysis in a cluster randomized design.

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
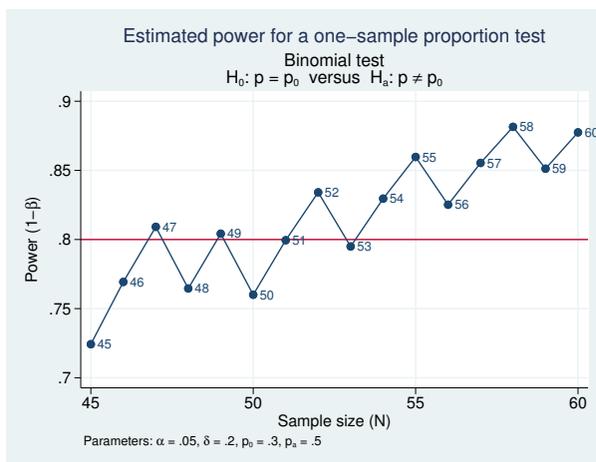
Figure 1. Saw-toothed power function

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,

$$\Pr(X \le C_l | p = p_0) \le \alpha/2 \text{ and } \Pr(X \ge C_u | p = p_0) \le \alpha/2$$

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/r15/lbw
(Hosmer & Lemeshow data)
. prtest ui==0.2
```

One-sample test of proportion   Number of obs   =   189

Variable	Mean	Std. Err.		[95% Conf. Interval]
ui	.1481481	.0258404	.0975019	.1987944

```
    p = proportion(ui)                                          z =  -1.7821
Ho: p = 0.2
```

Ha: p < 0.2	Ha: p != 0.2	Ha: p > 0.2
Pr(Z < z) = 0.0374	Pr(|Z| > |z|) = 0.0747	Pr(Z > z) = 0.9626

We do not have statistical evidence to reject the null hypothesis of H_0: $p = 0.2$ versus a two-sided alternative H_a: $p \neq 0.2$ at least at the 5% significance level; the p-value $= 0.0747 > 0.05$.

If our true objective were to study uterine irritability in the population of females, we would have performed the corresponding power and sample-size analysis before collecting the data. For example, using the estimates of lbw.dta, we can use power oneproportion to compute the required sample size for a 5%-level two-sided large-sample z test to detect the change in proportion from the reference value of 0.2 to approximately 0.148 with a power of, say, 80%:

```
. power oneproportion 0.2 0.148
Performing iteration ...
Estimated sample size for a one-sample proportion test
Score z test
Ho: p = p0  versus  Ha: p != p0
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =   -0.0520
           p0 =    0.2000
           pa =    0.1480
Estimated sample size:
            N =       434
```

We find that we need 434 subjects, many more than the current sample of 189, to detect the specified change in proportions.

◁

Video examples

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

Power calculation for comparing a sample proportion to a reference value

Minimum detectable effect size for comparing a sample proportion to a reference value using Stata

Stored results

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

Scalars

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(continuity)	1 if continuity correction is used; 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 size or proportion
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

Matrices

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).

Methods and formulas are presented under the following headings:

> *Large-sample normal approximation*
> *Binomial test*

Large-sample normal approximation

For a large sample, the distribution of the sample proportion $\widehat{p}$ 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

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

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: Dekker.

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

Krishnamoorthy, K., and J. Peng. 2007. Some properties of the exact and score methods for binomial proportion and sample size calculation. *Communications in Statistics—Simulation and Computation* 36: 1171–1186.

Levin, B., and X. Chen. 1999. Is the one-half continuity correction used once or twice to derive a well-known approximate sample size formula to compare two independent binomial distributions? *American Statistician* 53: 62–66.

Also see

[PSS] **power oneproportion, cluster** — Power analysis for a one-sample proportion test, CRD

[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