Title

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

Description

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

Quick start

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

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

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

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

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

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

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

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

Sample size for one-sided test with power of 0.9

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

Same as above, specified as μ_1 and difference between means $m_2 - m_1 = 4$

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

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

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

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

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

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

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

Menu

Statistics > Power and sample size

Syntax

Compute sample size

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

Compute power

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

Compute effect size and experimental-group mean

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

where m_1 is the mean of the control (reference) group, and m_2 is the mean of the experimental (comparison) group. m_1 and m_2 may each be specified either as one number or as a list of values in parentheses (see **[U] 11.1.8 numlist**).

options	Description
Main	
* alpha(*numlist*)	significance level; default is alpha(0.05)
* power(*numlist*)	power; default is power(0.8)
* beta(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	total sample size; required to compute power or effect size
* n1(*numlist*)	sample size of the control group
* n2(*numlist*)	sample size of the experimental group
* nratio(*numlist*)	ratio of sample sizes, N2/N1; default is nratio(1), meaning equal group sizes
compute(n1 | n2)	solve for N1 given N2 or for N2 given N1
nfractional	allow fractional sample sizes
* diff(*numlist*)	difference between the experimental-group mean and the control-group mean, $m_2 - m_1$; specify instead of the experimental-group mean m_2
* sd(*numlist*)	common standard deviation of the control and the experimental groups assuming equal standard deviations in both groups; default is sd(1)
* sd1(*numlist*)	standard deviation of the control group; requires sd2()
* sd2(*numlist*)	standard deviation of the experimental group; requires sd1()
knownsds	request computation assuming known standard deviations for both groups; default is to assume unknown standard deviations
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 sizes or experimental-group mean

`iterate(#)`	maximum number of iterations; default is `iterate(500)`
`tolerance(#)`	parameter tolerance; default is `tolerance(1e-12)`
`ftolerance(#)`	function tolerance; default is `ftolerance(1e-12)`
`[no]log`	suppress or display iteration log
`[no]dots`	suppress or display iterations as dots
`notitle`	suppress the title

*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.

where *tablespec* is

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

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

column	Description	Symbol
`alpha`	significance level	α
`power`	power	$1-\beta$
`beta`	type II error probability	β
`N`	total number of subjects	N
`N1`	number of subjects in the control group	N_1
`N2`	number of subjects in the experimental group	N_2
`nratio`	ratio of sample sizes, experimental to control	N_2/N_1
`delta`	effect size	δ
`m1`	control-group mean	μ_1
`m2`	experimental-group mean	μ_2
`diff`	difference between the control-group mean and the experimental-group mean	$\mu_2-\mu_1$
`sd`	common standard deviation	σ
`sd1`	control-group standard deviation	σ_1
`sd2`	experimental-group standard deviation	σ_2
`target`	target parameter; synonym for `m2`	
`_all`	display all supported columns	

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

Columns `nratio`, `diff`, `sd`, `sd1`, and `sd2` are shown in the default table if specified.

Options

Main

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

diff(*numlist*) specifies the difference between the experimental-group mean and the control-group mean, $m_2 - m_1$. You can specify either the experimental-group mean m_2 as a command argument or the difference between the two means in diff(). If you specify diff(#), the experimental-group mean is computed as $m_2 = m_1 + \#$. This option is not allowed with the effect-size determination.

sd(*numlist*) specifies the common standard deviation of the control and the experimental groups assuming equal standard deviations in both groups. The default is sd(1).

sd1(*numlist*) specifies the standard deviation of the control group. If you specify sd1(), you must also specify sd2().

sd2(*numlist*) specifies the standard deviation of the experimental group. If you specify sd2(), you must also specify sd1().

knownsds requests that standard deviations of each group be treated as known in the computations. By default, standard deviations are treated as unknown, and the computations are based on a two-sample t test, which uses a Student's t distribution as a sampling distribution of the test statistic. If knownsds is specified, the computation is based on a two-sample z test, which uses a normal distribution as the sampling distribution of the test statistic.

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

Table

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

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

Graph

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

Iteration

init(#) specifies the initial value for the estimated parameter. For sample-size determination, the estimated parameter is either the control-group size n_1 or, if compute(n2) is specified, the experimental-group size n_2. For the effect-size determination, the estimated parameter is the experimental-group mean m_2. The default initial values for a two-sided test are obtained as a closed-form solution for the corresponding one-sided test with the significance level $\alpha/2$. The default initial values for the t test computations are based on the corresponding large-sample normal approximation.

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

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

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

Remarks and examples

Remarks are presented under the following headings:

This entry describes the `power twomeans` command and the methodology for power and sample-size analysis for a two-sample means test. See [PSS] **intro** for a general introduction to power and sample-size analysis and [PSS] **power** for a general introduction to the `power` command using hypothesis tests.

Introduction

The analysis of means is one of the most commonly used approaches in a wide variety of statistical studies. Many applications lead to the study of two independent means, such as studies comparing the average mileage of foreign and domestic cars, the average SAT scores obtained from two different coaching classes, the average yields of a crop due to a certain fertilizer, and so on. The two populations of interest are assumed to be independent.

This entry describes power and sample-size analysis for the inference about two population means performed using hypothesis testing. Specifically, we consider the null hypothesis H_0: $\mu_2 = \mu_1$ versus the two-sided alternative hypothesis H_a: $\mu_2 \neq \mu_1$, the upper one-sided alternative H_a: $\mu_2 > \mu_1$, or the lower one-sided alternative H_a: $\mu_2 < \mu_1$.

The considered two-sample tests rely on the assumption that the two random samples are normally distributed or that the sample size is large. Suppose that the two samples are normally distributed. If variances of the considered populations are known a priori, the test statistic has a standard normal distribution under the null hypothesis, and the corresponding test is referred to as a two-sample z test. If variances of the two populations are not known, then the null sampling distribution of the test statistic depends on whether the two variances are assumed to be equal. If the two variances are assumed to be equal, the test statistic has an exact Student's t distribution under the null hypothesis. The corresponding test is referred to as a two-sample t test. If the two variances are not equal, then the distribution can only be approximated by a Student's t distribution; the degrees of freedom is approximated using Satterthwaite's method. We refer to this test as Satterthwaite's t test. For a large sample, the distribution of the test statistic is approximately normal, and the corresponding test is a large-sample z test.

The `power twomeans` command provides power and sample-size analysis for the above tests.

Using power twomeans

`power twomeans` computes sample size, power, or experimental-group mean for a two-sample means test. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced- or equal-allocation design; see [PSS] **unbalanced designs** for a description of how to specify an unbalanced design.

By default, all computations are for a two-sample t test, which assumes equal and unknown standard deviations. By default, the common standard deviation is set to one but may be changed by specifying the `sd()` option. To specify different standard deviations, use the respective `sd1()` and `sd2()` options. These options must be specified together and may not be used in combination with `sd()`. When `sd1()` and `sd2()` are specified, the computations are based on Satterthwaite's t test, which assumes unequal and unknown standard deviations. If standard deviations are known, use the `knownsds` option to request that computations be based on a two-sample z test.

To compute the total sample size, you must specify the control-group mean m_1, the experimental-group mean m_2, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

Instead of the total sample size, you can compute one of the group sizes given the other one. To compute the control-group sample size, you must specify the `compute(n1)` option and the sample size of the experimental group in the `n2()` option. Likewise, to compute the experimental-group sample size, you must specify the `compute(n2)` option and the sample size of the control group in the `n1()` option.

To compute power, you must specify the total sample size in the `n()` option, the control-group mean m_1, and the experimental-group mean m_2.

Instead of the experimental-group mean m_2, you may specify the difference $m_2 - m_1$ between the experimental-group mean and the control-group mean in the `diff()` option when computing sample size or power.

To compute effect size, the difference between the experimental-group mean and the null mean, and the experimental-group mean, you must specify the total sample size in the `n()` option, the power in the `power()` option, the control-group mean m_1, and, optionally, the direction of the effect. The direction is upper by default, `direction(upper)`, which means that the experimental-group mean is assumed to be larger than the specified control-group value. You can change the direction to be lower, which means that the experimental-group mean is assumed to be smaller than the specified control-group value, by specifying the `direction(lower)` option.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see *Two samples* in [PSS] **unbalanced designs** for more details.

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

Computing sample size

To compute sample size, you must specify the control-group mean m_1, the experimental-group mean m_2, and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified.

▷ Example 1: Sample size for a two-sample means test

Consider a study investigating the effects of smoking on lung function of males. The response variable is forced expiratory volume (FEV), measured in liters (L), where better lung function implies higher values of FEV. We wish to test the null hypothesis H_0: $\mu_1 = \mu_2$ versus a two-sided alternative hypothesis H_a: $\mu_1 \neq \mu_2$, where μ_1 and μ_2 are the mean FEV for nonsmokers and smokers, respectively.

Suppose that the mean FEV from previous studies was reported to be 3 L for nonsmokers and 2.7 L for smokers. We are designing a new study and wish to find out how many subjects we need to enroll so that the power of a 5%-level two-sided test to detect the specified difference between means is at least 80%. We assume equal numbers of subjects in each group and a common standard deviation of 1.

```
. power twomeans 3 2.7
Performing iteration ...
Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
           sd =    1.0000
Estimated sample sizes:
            N =       352
  N per group =       176
```

We need a total sample of 352 subjects, 176 per group, to detect the specified mean difference between the smoking and nonsmoking groups with 80% power using a two-sided 5%-level test.

The default computation is for the case of equal and unknown standard deviations, as indicated by the output. You can specify the knownsds option to request the computation assuming known standard deviations.

◁

▷ Example 2: Sample size assuming unequal standard deviations

Instead of assuming equal standard deviations as in example 1, we use the estimates of the standard deviations from previous studies as our hypothetical values. The standard deviation of FEV for the nonsmoking group was reported to be 0.8 L and that for the smoking group was reported to be 0.7 L. We specify standard deviations in the sd1() and sd2() options.

```
. power twomeans 3 2.7, sd1(0.8) sd2(0.7)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
          sd1 =    0.8000
          sd2 =    0.7000
Estimated sample sizes:
            N =       200
  N per group =       100
```

The specified standard deviations are smaller than one, so we obtain a smaller required total sample size of 200 compared with example 1.

◁

▷ Example 3: Specifying difference between means

Instead of the mean FEV of 2.7 for the smoking group as in example 2, we can specify the difference between the two means of $2.7 - 3 = -0.3$ in the `diff()` option.

```
. power twomeans 3, sd1(0.8) sd2(0.7) diff(-0.3)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
         diff =   -0.3000
          sd1 =    0.8000
          sd2 =    0.7000
Estimated sample sizes:
            N =       200
  N per group =       100
```

We obtain the same results as in example 2. The difference between means is now also reported in the output following the individual means.

◁

▷ Example 4: Computing one of the group sizes

Suppose we anticipate a sample of 120 nonsmoking subjects. We wish to compute the required number of subjects in the smoking group, keeping all other study parameters as in example 2. We specify the number of subjects in the nonsmoking group in the `n1()` option and specify the `compute(n2)` option.

```
. power twomeans 3 2.7, sd1(0.8) sd2(0.7) n1(120) compute(n2)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
          sd1 =    0.8000
          sd2 =    0.7000
           N1 =       120
Estimated sample sizes:
            N =       202
           N2 =        82
```

We need a sample of 82 smoking subjects given a sample of 120 nonsmoking subjects.

◁

▷ Example 5: Unbalanced design

By default, `power twomeans` computes sample size for a balanced- or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size for an unbalanced design by specifying the `nratio()` option.

Continuing with example 2, we will suppose that we anticipate to recruit twice as many smokers than nonsmokers; that is, $n_2/n_1 = 2$. We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```
. power twomeans 3 2.7, sd1(0.8) sd2(0.7) nratio(2)
Performing iteration ...
Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
          sd1 =    0.8000
          sd2 =    0.7000
        N2/N1 =    2.0000
Estimated sample sizes:
            N =       237
           N1 =        79
           N2 =       158
```

We need a total sample size of 237 subjects, which is larger than the required total sample size for the corresponding balanced design from example 2.

Also see *Two samples* in [PSS] **unbalanced designs** for more examples of unbalanced designs for two-sample tests.

◁

Computing power

To compute power, you must specify the total sample size in the `n()` option, the control-group mean m_1, and the experimental-group mean m_2.

▷ Example 6: Power of a two-sample means test

Continuing with example 1, we will suppose that we have resources to enroll a total of only 250 subjects, assuming equal-sized groups. To compute the power corresponding to this sample size given the study parameters from example 1, we specify the total sample size in `n()`:

```
. power twomeans 3 2.7, n(250)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1
Study parameters:
        alpha =    0.0500
            N =       250
  N per group =       125
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
           sd =    1.0000
Estimated power:
        power =    0.6564
```

With a total sample of 250 subjects, we obtain a power of only 65.64%.

◁

▷ Example 7: Multiple values of study parameters

In this example, we assess the effect of varying the common standard deviation (assuming equal standard deviations in both groups) of FEV on the power of our study.

Continuing with example 6, we compute powers for a range of common standard deviations between 0.5 and 1.5 with the step size of 0.1. We specify the corresponding *numlist* in the `sd()` option.

```
. power twomeans 3 2.7, sd(0.5(0.1)1.5) n(250)
Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1
```

alpha	power	N	N1	N2	delta	m1	m2	sd
.05	.9972	250	125	125	-.3	3	2.7	.5
.05	.976	250	125	125	-.3	3	2.7	.6
.05	.9215	250	125	125	-.3	3	2.7	.7
.05	.8397	250	125	125	-.3	3	2.7	.8
.05	.747	250	125	125	-.3	3	2.7	.9
.05	.6564	250	125	125	-.3	3	2.7	1
.05	.5745	250	125	125	-.3	3	2.7	1.1
.05	.5036	250	125	125	-.3	3	2.7	1.2
.05	.4434	250	125	125	-.3	3	2.7	1.3
.05	.3928	250	125	125	-.3	3	2.7	1.4
.05	.3503	250	125	125	-.3	3	2.7	1.5

The power decreases from 99.7% to 35.0% as the common standard deviation increases from 0.5 to 1.5 L.

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

◁

Computing effect size and experimental-group mean

Effect size δ for a two-sample means test is defined as the difference between the experimental-group mean and the control-group mean $\delta = \mu_2 - \mu_1$.

Sometimes, we may be interested in determining the smallest effect and the corresponding experimental-group mean that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and control-group mean must be specified. In addition, you must also decide on the direction of the effect: upper, meaning $m_2 > m_1$, or lower, meaning $m_2 < m_1$. The direction may be specified in the `direction()` option; `direction(upper)` is the default.

▷ Example 8: Minimum detectable change in the experimental-group mean

Continuing with example 6, we compute the smallest change in the mean of the smoking group that can be detected given a total sample of 250 subjects and 80% power, assuming equal-group allocation. To solve for the mean FEV of the smoking group, after the command name, we specify the nonsmoking-group mean of 3, total sample size `n(250)`, and power `power(0.8)`.

Because our initial study was based on the hypothesis that FEV for the smoking group is lower than that of the nonsmoking group, we specify the `direction(lower)` option to compute the smoking-group mean that is lower than the specified nonsmoking-group mean.

```
. power twomeans 3, n(250) power(0.8) direction(lower)
Performing iteration ...
Estimated experimental-group mean for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1; m2 < m1
Study parameters:
        alpha =    0.0500
        power =    0.8000
            N =       250
  N per group =       125
           m1 =    3.0000
           sd =    1.0000
Estimated effect size and experimental-group mean:
        delta =   -0.3558
           m2 =    2.6442
```

We find that the minimum detectable value of the effect size is −0.36, which corresponds to the mean FEV of 2.64 for the smoking group.

◁

Testing a hypothesis about two independent means

After data are collected, we can use the `ttest` command to test the equality of two independent means using a t test; see [R] **ttest** for details. In this section, we demonstrate the use of `ttesti`, the immediate form of the `test` command, which can be used to test a hypothesis using summary statistics instead of the actual data values.

▷ Example 9: Two-sample t test

Consider an example from van Belle et al. (2004, 129), where newborn infants were divided into two groups: a treatment group, where infants received daily "walking stimulus" for eight weeks, and a control group, where no stimulus was provided. The goal of this study was to test whether receiving the walking stimulus during stages of infancy induces the walking ability to develop sooner.

The average number of months before the infants started walking was recorded for both groups. The authors provide estimates of the average of 10.125 months for the treatment group with estimated standard deviation of 1.447 months and 12.35 months for the control group with estimated standard deviation of 0.9618 months. The sample sizes for treatment and control groups were 6 and 5, respectively. We supply these estimates to the `ttesti` command and use the `unequal` option to perform a t test assuming unequal variances.

```
. ttesti 6 10.125 1.447 5 12.35 0.9618, unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       6      10.125    .5907353       1.447    8.606467    11.64353
       y |       5       12.35      .43013       .9618    11.15577    13.54423
---------+--------------------------------------------------------------------
combined |      11    11.13636     .501552     1.66346    10.01884    12.25389
---------+--------------------------------------------------------------------
    diff |              -2.225    .7307394               -3.887894    -.562106
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =  -3.0449
Ho: diff = 0                     Satterthwaite's degrees of freedom =  8.66326

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0073         Pr(|T| > |t|) = 0.0145          Pr(T > t) = 0.9927
```

We reject the null hypothesis of H_0: $\mu_C = \mu_T$ against the two-sided alternative H_a: $\mu_C \neq \mu_T$ at the 5% significance level; the p-value $= 0.0145$.

We use the estimates of this study to perform a sample-size analysis we would have conducted before a new study. In our analysis, we assume equal-group allocation.

```
. power twomeans 10.125 12.35, power(0.8) sd1(1.447) sd2(0.9618)
Performing iteration ...

Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =    2.2250
           m1 =   10.1250
           m2 =   12.3500
          sd1 =    1.4470
          sd2 =    0.9618

Estimated sample sizes:

            N =        14
  N per group =         7
```

We find that the sample size required to detect a difference of 2.225 ($12.35 - 10.125 = 2.225$) given the control-group standard deviation of 1.447 and the experimental-group standard deviation of 0.9618 using a 5%-level two-sided test is 7 in each group.

◁

Stored results

power twomeans stores the following in r():

Scalars

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

Macros

r(type)	test
r(method)	twomeans
r(direction)	upper or lower
r(columns)	displayed table columns
r(labels)	table column labels
r(widths)	table column widths
r(formats)	table column formats

Matrix

r(pss_table)	table of results

Methods and formulas

Consider two independent samples with n_1 subjects in the control group and n_2 subjects in the experimental group. Let $x_{11}, \ldots, x_{1n_1}$ be a random sample of size n_1 from a normal population with mean μ_1 and variance σ_1^2. Let $x_{21}, \ldots, x_{2n_2}$ be a random sample of size n_2 from a normal population with mean μ_2 and variance σ_2^2. Let effect size δ be the difference between the experimental-group mean and the control-group mean, $\delta = \mu_2 - \mu_1$. The sample means and variances for the two independent samples are

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

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

where $\overline{x}_j$ and s_j^2 are the respective sample means and sample variances of the two samples.

A two-sample means test involves testing the null hypothesis H_0: $\mu_2 = \mu_1$ versus the two-sided alternative hypothesis H_a: $\mu_2 \neq \mu_1$, the upper one-sided alternative H_a: $\mu_2 > \mu_1$, or the lower one-sided alternative H_a: $\mu_2 < \mu_1$.

The two-sample means test can be performed under four different assumptions: 1) population variances are known and not equal; 2) population variances are known and equal; 3) population variances are unknown and not equal; and 4) population variances are unknown and equal.

Let σ_D denote the standard deviation of the difference between the two sample means. The test statistic of the form

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

is used in each of the four cases described above. Each case, however, determines the functional form of σ_D and the sampling distribution of the test statistic (1) under the null hypothesis.

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

The following formulas are based on Armitage, Berry, and Matthews (2002); Chow, Shao, and Wang (2008); and Dixon and Massey (1983).

Methods and formulas are presented under the following headings:

Known standard deviations

Below we present formulas for the computations that assume unequal standard deviations. When standard deviations are equal, the corresponding formulas are special cases of the formulas below with $\sigma_1 = \sigma_2 = \sigma$.

When the standard deviations of the control and the experimental groups are known, the test statistic in (1) is a z test statistic

$$z = \frac{(\overline{x}_2 - \overline{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

with $\sigma_D = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$. The sampling distribution of this test statistic under the null hypothesis is standard normal. The corresponding test is referred to as a z test.

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

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

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

where $\Phi(\cdot)$ is the cdf of a standard normal distribution.

For a one-sided test, the control-group sample size n_1 is computed as follows:

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

For a one-sided test, if one of the group sizes is known, the other one is computed using the following formula. For example, to compute n_1 given n_2, we use the following formula:

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

For a two-sided test, sample sizes are computed by iteratively solving the two-sided power equation in (2). The default initial values for the iterative procedure are calculated from the respective equations (3) and (4), with α replaced with $\alpha/2$.

The absolute value of the effect size for a one-sided test is obtained by inverting the corresponding one-sided power equation in (2):

$$|\delta| = \sigma_D(z_{1-\alpha} - z_\beta)$$

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

The experimental-group mean for a one-sided test is then computed as

$$\mu_2 = \begin{cases} \mu_1 + (z_{1-\alpha} - z_\beta)\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} & \text{when } \mu_2 > \mu_1 \\ \mu_1 - (z_{1-\alpha} - z_\beta)\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} & \text{when } \mu_2 < \mu_1 \end{cases}$$

For a two-sided test, the experimental-group mean is computed by iteratively solving the two-sided power equation in (2) for μ_2. The default initial value is obtained from the corresponding one-sided computation with $\alpha/2$.

Unknown standard deviations

When the standard deviations of the control group and the experimental group are unknown, the test statistic in (1) is a t test statistic

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

where s_D is the estimated standard deviation of the sample mean difference. The sampling distribution of this test statistic under the null hypothesis is (approximately) a Student's t distribution with ν degrees of freedom. Parameters ν and s_D are defined below, separately for the case of equal and unequal standard deviations.

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

The power is computed from the following equations:

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

In the equations above, $\lambda = |\mu_2 - \mu_1|/s_D$.

Sample sizes and the experimental-group mean are obtained by iteratively solving the nonlinear equation (5) for n_1, n_2, and μ_2, respectively. For sample-size and effect-size computations, the default initial values for the iterative procedure are calculated using the corresponding formulas assuming known standard deviations from the previous subsection.

Unequal standard deviations

In the case of unequal standard deviations,

$$s_D = \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

and the degrees of freedom ν of the test statistic is obtained by Satterthwaite's formula:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

The sampling distribution of the test statistic under the null hypothesis is an approximate Student's t distribution. We refer to the corresponding test as Satterthwaite's t test.

Equal standard deviations

In the case of equal standard deviations,

$$s_D = s_p\sqrt{1/n_1 + 1/n_2}$$

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

The degrees of freedom ν is

$$\nu = n_1 + n_2 - 2$$

The sampling distribution of the test statistic under the null hypothesis is exactly a Student's t distribution. We refer to the corresponding test as a two-sample t test.

References

Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.

Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research*. 2nd ed. New York: Dekker.

Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw–Hill.

van Belle, G., L. D. Fisher, P. J. Heagerty, and T. S. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences*. 2nd ed. New York: Wiley.

Also see

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

[PSS] **power oneway** — Power analysis for one-way analysis of variance

[PSS] **power twoway** — Power analysis for two-way analysis of variance

[PSS] **power, graph** — Graph results from the power command

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

[PSS] **Glossary**

[R] **ttest** — t tests (mean-comparison tests)