power pairedmeans — Power analysis for a two-sample paired-means test

Description

power pairedmeans computes sample size, power, or target mean difference for a two-sample paired-means test. By default, it computes sample size for given power and the values of the null and alternative mean differences. Alternatively, it can compute power for given sample size and the values of the null and alternative mean differences or the target mean difference for given sample size, power, and the null mean difference. Also see [PSS-2] power for a general introduction to the power command using hypothesis tests.

For precision and sample-size analysis for a CI for the difference between two means from paired samples, see [PSS-3] ciwidth pairedmeans.

Quick start

Sample size for a test of \( H_0: \mu_2 - \mu_1 = d = 0 \) versus \( H_a: d \neq 0 \) given alternative pretreatment mean \( m_{a1} = 73 \) and alternative posttreatment mean \( m_{a2} = 57 \) with standard deviation of the differences \( \sigma_d = 36 \) using default power of 0.8 and significance level \( \alpha = 0.05 \)

\[
\text{power pairedmeans 73 57, sddiff(36)}
\]

Same as above, specified using the difference between means of \(-16\)

\[
\text{power pairedmeans, altdiff(-16) sddiff(36)}
\]

As above, but instead of standard deviation of the differences, specify correlation between paired observations of 0.5 with pretreatment standard deviation of 29 and posttreatment standard deviation of 40

\[
\text{power pairedmeans 73 57, corr(.5) sd1(29) sd2(40)}
\]

For differences in means of \(-20, -18, -16, -14, -12, \) and \(-10\)

\[
\text{power pairedmeans, altdiff(-20(2)-10) sddiff(36)}
\]

Power for a sample size of 23

\[
\text{power pairedmeans 73 57, sddiff(36) n(23)}
\]

Effect size and target mean difference for sample sizes 20, 30, and 40 with power of 0.85

\[
\text{power pairedmeans 73, sdiff(36) power(.85) n(20(10))40}
\]

As above, but display results as a graph of target mean difference versus sample size

\[
\text{power pairedmeans 73, sdiff(36) power(.85) n(20(10))40 graph}
\]
Menu
Statistics > Power, precision, and sample size

Syntax

Compute sample size

```
power pairedmeans m_{a1} m_{a2}, corrspec [power(numlist) options]
```

Compute power

```
power pairedmeans m_{a1} m_{a2}, corrspec n(numlist) [options]
```

Compute effect size and target mean difference

```
power pairedmeans [m_{a1}], corrspec n(numlist) power(numlist) [options]
```

where corrspec is one of

- `sddiff()`
- `corr() [sd()]`
- `corr() [sd1() sd2()]`

$m_{a1}$ is the alternative pretreatment mean or the pretreatment mean under the alternative hypothesis, and $m_{a2}$ is the alternative posttreatment mean or the value of the posttreatment mean under the alternative hypothesis. $m_{a1}$ and $m_{a2}$ may each be specified either as one number or as a list of values in parentheses (see `[u] 11.1.8 numlist`).
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) sample size; required to compute power or effect size
nfractional allow fractional sample size
*nulldiff(numlist) null difference, the difference between the posttreatment mean and the pretreatment mean under the null hypothesis; default is nulldiff(0)
*altdiff(numlist) alternative difference $d_a = m_{a2} - m_{a1}$, the difference between the posttreatment mean and the pretreatment mean under the alternative hypothesis
*sddiff(numlist) standard deviation $\sigma_d$ of the differences; may not be combined with corr()
*corr(numlist) correlation between paired observations; required unless sddiff() is specified
*sd(numlist) common standard deviation; default is sd(1) and requires corr()
*sd1(numlist) standard deviation of the pretreatment group; requires corr()
*sd2(numlist) standard deviation of the posttreatment group; requires corr()
knownsd request computation assuming a known standard deviation $\sigma_d$; default is to assume an unknown standard deviation
*fpc(numlist) finite population correction (FPC) as a sampling rate or population size

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-2] power, table
saving(filename [, replace]) save the table data to filename; use replace to overwrite existing filename

Graph
graph (graphopts) graph results; see [PSS-2] power, graph
Iteration

init(#)  initial value for sample size or mean difference; default is to use normal approximation

iterate(#)  maximum number of iterations; default is iterate(500)

tolerance(#)  parameter tolerance; default is tolerance(1e-12)

ftolerance(#)  function tolerance; default is ftolerance(1e-12)

[no]log  suppress or display iteration log

[no]dots  suppress or display iterations as dots

notitle  suppress the title

*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).

<table>
<thead>
<tr>
<th>column</th>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>significance level</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>power</td>
<td>power</td>
<td>( 1 - \beta )</td>
</tr>
<tr>
<td>beta</td>
<td>type II error probability</td>
<td>( \beta )</td>
</tr>
<tr>
<td>N</td>
<td>number of subjects</td>
<td>( N )</td>
</tr>
<tr>
<td>delta</td>
<td>effect size</td>
<td>( \delta )</td>
</tr>
<tr>
<td>d0</td>
<td>null mean difference</td>
<td>( d_0 )</td>
</tr>
<tr>
<td>da</td>
<td>alternative mean difference</td>
<td>( d_a )</td>
</tr>
<tr>
<td>ma1</td>
<td>alternative pretreatment mean</td>
<td>( \mu_{a1} )</td>
</tr>
<tr>
<td>ma2</td>
<td>alternative posttreatment mean</td>
<td>( \mu_{a2} )</td>
</tr>
<tr>
<td>sd_d</td>
<td>standard deviation of the differences</td>
<td>( \sigma_d )</td>
</tr>
<tr>
<td>sd</td>
<td>common standard deviation</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>sd1</td>
<td>standard deviation of the pretreatment group</td>
<td>( \sigma_1 )</td>
</tr>
<tr>
<td>sd2</td>
<td>standard deviation of the posttreatment group</td>
<td>( \sigma_2 )</td>
</tr>
<tr>
<td>corr</td>
<td>correlation between paired observations</td>
<td>( \rho )</td>
</tr>
<tr>
<td>fpc</td>
<td>FPC as a population size</td>
<td>( N_{\text{pop}} )</td>
</tr>
<tr>
<td>target</td>
<td>target parameter; synonym for da</td>
<td></td>
</tr>
<tr>
<td>_all</td>
<td>display all supported columns</td>
<td></td>
</tr>
</tbody>
</table>

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

Columns ma1, ma2, sd, sd1, sd2, corr, and fpc are shown in the default table if specified.
Options

**Main**

- `alpha()`, `power()`, `beta()`, `n()`, `nfractional`; see [PSS-2] power. The `nfractional` option is allowed only for sample-size determination.

- `nulldiff(numlist)` specifies the difference between the posttreatment mean and the pretreatment mean under the null hypothesis. The default is `nulldiff(0)`, which means that the pretreatment mean equals the posttreatment mean under the null hypothesis.

- `altdiff(numlist)` specifies the alternative difference $d_a = m_{a2} - m_{a1}$, the difference between the posttreatment mean and the pretreatment mean under the alternative hypothesis. This option is the alternative to specifying the alternative means $m_{a1}$ and $m_{a2}$. If $m_{a1}$ is specified in combination with `altdiff(#)`, then $m_{a2} = # + m_{a1}$.

- `sddiff(numlist)` specifies the standard deviation $\sigma_d$ of the differences. Either `sddiff()` or `corr()` must be specified.

- `corr(numlist)` specifies the correlation between paired, pretreatment and posttreatment, observations. This option along with `sd1()` and `sd2()` or `sd()` is used to compute the standard deviation of the differences unless that standard deviation is supplied directly in the `sddiff()` option. Either `corr()` or `sddiff()` must be specified.

- `sd(numlist)` specifies the common standard deviation of the pretreatment and posttreatment groups. Specifying `sd(#)` implies that both `sd1()` and `sd2()` are equal to #. Options `corr()` and `sd()` are used to compute the standard deviation of the differences unless that standard deviation is supplied directly with the `sddiff()` option. The default is `sd(1)`.

- `sd1(numlist)` specifies the standard deviation of the pretreatment group. Options `corr()`, `sd1()`, and `sd2()` are used to compute the standard deviation of the differences unless that standard deviation is supplied directly with the `sddiff()` option.

- `sd2(numlist)` specifies the standard deviation of the posttreatment group. Options `corr()`, `sd1()`, and `sd2()` are used to compute the standard deviation of the differences unless that standard deviation is supplied directly with the `sddiff()` option.

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

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

**Table**

- `table()`, `notable`; see [PSS-2] power, table.

- `saving()`; see [PSS-2] power.
Introduction

The analysis of paired means is commonly used in settings such as repeated-measures designs with before and after measurements on the same individual or cross-sectional studies of paired measurements from twins. For example, a company might initiate a voluntary exercise program and would like to test that the average weight loss of participants from beginning to six months is greater than zero. Or a school district might design an intensive remedial program for students with low math scores and would like to know if the students’ math scores improve from the pretest to the posttest. For paired data, the inference is made on the mean difference accounting for the dependence between the two groups.

This entry describes power and sample-size analysis for the inference about the population mean difference performed using hypothesis testing. Specifically, we consider the null hypothesis \( H_0: d = d_0 \) versus the two-sided alternative hypothesis \( H_a: d \neq d_0 \), the upper one-sided alternative \( H_a: d > d_0 \), or the lower one-sided alternative \( H_a: d < d_0 \). The parameter \( d \) is the mean difference between the posttreatment mean \( \mu_2 \) and pretreatment mean \( \mu_1 \).

A two-sample paired-means test assumes that the two correlated samples are drawn from two normal populations or that the sample size is large. When the population variances are known, the sampling distribution of the test statistic under the null hypothesis is standard normal, and the corresponding test is known as a paired \( z \) test. If the population variances are unknown, the sampling distribution of the test statistic under the null hypothesis is Student’s \( t \), and the corresponding test is known as a paired \( t \) test.
The random sample is typically drawn from an infinite population. When the sample is drawn from a population of a fixed size, sampling variability must be adjusted for a finite population size.

The **power pairedmeans** command provides power and sample-size analysis for the comparison of two correlated means using a paired t test or a paired z test.

### Using power pairedmeans

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

By default, all computations are based on a paired t test, which assumes an unknown standard deviation of the differences. For a known standard deviation, you can specify the `knownsd` option to request a paired z test.

For all computations, you must specify either the standard deviation of the differences in the `sddiff()` option or the correlation between the paired observations in the `corr()` option. If you specify the `corr()` option, then individual standard deviations of the pretreatment and posttreatment groups may also be specified in the respective `sd1()` and `sd2()` options. By default, their values are set to 1. When the two standard deviations are equal, you may specify the common standard deviation in the `sd()` option instead of specifying them individually.

To compute sample size, you must specify the pretreatment and posttreatment means under the alternative hypothesis, $m_{a1}$ and $m_{a2}$, respectively, and, optionally, the power of the test in the `power()` option. The default power is set to 0.8.

To compute power, you must specify the sample size in the `n()` option and the pretreatment and posttreatment means under the alternative hypothesis, $m_{a1}$ and $m_{a2}$, respectively.

Instead of the alternative means $m_{a1}$ and $m_{a2}$, you can specify the difference $m_{a2} - m_{a1}$ between the alternative posttreatment mean and the alternative pretreatment mean in the `altdiff()` option when computing sample size or power.

By default, the difference between the posttreatment mean and the pretreatment mean under the null hypothesis is set to zero. You may specify other values in the `nulldiff()` option.

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

By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS-4] *Unbalanced designs* for an example. The `nfractional` option is allowed only for sample-size determination.

Some of *power pairedmeans*’s computations require iteration. For example, when the standard deviation of the differences is unknown, computations use a noncentral Student’s t distribution. Its degrees of freedom depends on the sample size, and the noncentrality parameter depends on the sample size and effect size. Therefore, the sample-size and effect-size determinations require iteration. The default initial values of the estimated parameters are obtained by using a closed-form normal
approximation. They may be changed by specifying the init() option. See [PSS-2] power for the descriptions of other options that control the iteration procedure.

All computations assume an infinite population. For a finite population, use the fpc() option to specify a sampling rate or a population size. When this option is specified, a finite population correction is applied to the standard deviation of the differences. The correction factor depends on the sample size; therefore, computing sample size in this case requires iteration. The initial value for sample-size determination in this case is based on the corresponding normal approximation with a finite population size.

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

Computing sample size

To compute sample size, you must specify the pretreatment and posttreatment means under the alternative hypothesis, $m_{a1}$ and $m_{a2}$, respectively, or the difference between them in altdiff() and, optionally, the power of the test in the power() option. A default power of 0.8 is assumed if power() is not specified.

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

Consider a study of low birthweight (LBW) infants as in Howell (2002, 186). The variable of interest is the Bayley mental development index (MDI) of infants when they are 6-, 12-, and 24-months old. Previous research suggested that the MDI scores for LBW children might decline significantly between 6 and 24 months of age. Suppose we would like to conduct a similar study where the null hypothesis of interest is no difference between 6-month and 24-month MDI scores, $H_0: d = 0$, and the two-sided alternative is $H_a: d \neq 0$, implying the existence of a difference.

In this example, we use the estimates from Howell (2002, 193) as our study parameters. The mean MDI score of a 6-month group was estimated to be 111. We want to obtain the minimum sample size that is required to detect the mean MDI score of 106.71 in a 24-month group with a power of 80% using a 5%-level two-sided test. The standard deviation of the differences was previously estimated to be 16.04. To compute the sample size, we specify the alternative means after the command name and standard deviation of the differences in sdiff().
. power pairedmeans 111 106.71, sddiff(16.04)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
\[
\begin{align*}
\alpha &= 0.0500 & m_{a1} &= 111.0000 \\
power &= 0.8000 & m_{a2} &= 106.7100 \\
delta &= -0.2675 \\
d0 &= 0.0000 \\
da &= -4.2900 \\
sd_d &= 16.0400
\end{align*}
\]
Estimated sample size:
\[N = 112\]

As we mentioned in the previous section, sample-size determination requires iteration in the case of an unknown standard deviation. By default, `power pairedmeans` suppresses the iteration log, which may be displayed by specifying the `log` option.

A sample of 112 subjects is required for the test to detect the resulting difference of \(-4.29\) with a power of 80%.

Study parameters are divided into two columns. The parameters that are always displayed are listed in the first column, and the parameters that are displayed only if they are specified are listed in the second column.

In this example, we specified optional command arguments containing the alternative pretreatment mean `m_{a1}` and the alternative posttreatment mean `m_{a2}`. Because these arguments are optional, they are listed in the second column.

Example 2: Specifying mean differences

Instead of the individual alternative means, we can specify their difference, \(106.71 - 111 = -4.29\), in the `altdiff()` option.

. power pairedmeans, altdiff(-4.29) sddiff(16.04)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
\[
\begin{align*}
\alpha &= 0.0500 \\
power &= 0.8000 \\
delta &= -0.2675 \\
d0 &= 0.0000 \\
da &= -4.2900 \\
sd_d &= 16.0400
\end{align*}
\]
Estimated sample size:
\[N = 112\]

We obtain the same results as in example 1.
Example 3: Specifying individual standard deviations

Howell (2002) also reported the group-specific standard deviations: 13.85 in the 6-month group and 12.95 in the 24-month group. Using the values of individual standard deviations and the standard deviation of the differences from the previous example, we obtain the correlation between the 6-month group and the 24-month group to be \((\frac{13.85^2 + 12.95^2 - 16.04^2}{2 \times 13.85 \times 12.95}) = 0.285\). To compute the sample size, we specify the group-specific standard deviations in `sd1()` and `sd2()` and the correlation in `corr()`.

```
. power pairedmeans 111 106.71, corr(0.285) sd1(13.85) sd2(12.95)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0
Study parameters:
  alpha = 0.0500        ma1 = 111.0000
  power = 0.8000        ma2 = 106.7100
  delta = -0.2675       sd1 = 13.8500
  d0 = 0.0000           sd2 = 12.9500
  da = -4.2900          corr = 0.2850
  sd_d = 16.0403
Estimated sample size:
  N = 112
```

We obtain the same sample size as in example 1.

The correlation and standard deviations are reported in the second column.

Example 4: Specifying common standard deviation

If standard deviations in both groups are equal, we may specify the common standard deviation in option `sd()`. As a demonstration, we use the average of the individual standard deviations \((\frac{13.85 + 12.95}{2} = 13.4)\) as our common standard deviation.

```
. power pairedmeans 111 106.71, corr(0.285) sd(13.4)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test assuming sd1 = sd2 = sd
Ho: d = d0 versus Ha: d != d0
Study parameters:
  alpha = 0.0500        ma1 = 111.0000
  power = 0.8000        ma2 = 106.7100
  delta = -0.2677       sd = 13.4000
  d0 = 0.0000           corr = 0.2850
  da = -4.2900          sd_d = 16.0241
Estimated sample size:
  N = 112
```

The resulting standard deviation of the differences of 16.0241 is close to our earlier estimate of 16.04, so the computed sample size is the same as the sample size in example 1.
Example 5: Nonzero null

In all the previous examples, we assumed that the difference between the 6-month and 24-month means is zero under the null hypothesis. For a nonzero null hypothesis, you can specify the corresponding null value in the `nulldiff()` option.

Continuing with example 2, we will suppose that we are testing the nonzero null hypothesis of $H_0: d = d_0 = -1$. We compute the sample size as follows:

```
. power pairedmeans, nulldiff(-1) altdiff(-4.29) sddiff(16.04)
```

Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0

Study parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>0.0500</td>
</tr>
<tr>
<td>power</td>
<td>0.8000</td>
</tr>
<tr>
<td>delta</td>
<td>-0.2051</td>
</tr>
<tr>
<td>d0</td>
<td>-1.0000</td>
</tr>
<tr>
<td>da</td>
<td>-4.2900</td>
</tr>
<tr>
<td>sd_d</td>
<td>16.0400</td>
</tr>
</tbody>
</table>

Estimated sample size:

$N = 189$

Compared with example 2, the absolute value of the effect size $\delta$ decreases to 0.2051, and thus a larger sample of 189 subjects is required to detect this smaller effect.

Computing power

To compute power, you must specify the sample size in the `n()` option and the pretreatment and posttreatment means under the alternative hypothesis, $m_{a1}$ and $m_{a2}$, respectively, or the difference between them in the `altdiff()` option.

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

Continuing with example 1, we will suppose that because of limited resources, we anticipate to obtain a sample of only 100 subjects. To compute power, we specify the sample size in the `n()` option:

```
. power pairedmeans 111 106.71, n(100) sddiff(16.04)
```

Estimated power for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0

Study parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>0.0500</td>
</tr>
<tr>
<td>ma1</td>
<td>111.0000</td>
</tr>
<tr>
<td>N</td>
<td>100</td>
</tr>
<tr>
<td>ma2</td>
<td>106.7100</td>
</tr>
<tr>
<td>delta</td>
<td>-0.2675</td>
</tr>
<tr>
<td>d0</td>
<td>0.0000</td>
</tr>
<tr>
<td>da</td>
<td>-4.2900</td>
</tr>
<tr>
<td>sd_d</td>
<td>16.0400</td>
</tr>
</tbody>
</table>

Estimated power:

power = 0.7545

Compared with example 1, the power decreases to 75.45%.
Example 7: Known standard deviation

In the case of a known standard deviation $\sigma_d$, you can specify the knownsd option to request a paired $z$ test. Using the same study parameters as in example 6, we can compute the power as follows:

```
. power pairedmeans 111 106.71, n(100) sddiff(16.04) knownsd
```

Estimated power for a two-sample paired-means test

Paired $z$ test

Ho: $d = d_0$ versus Ha: $d \neq d_0$

Study parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.0500</td>
</tr>
<tr>
<td>$N$</td>
<td>100</td>
</tr>
<tr>
<td>$d_0$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\delta$</td>
<td>-0.2675</td>
</tr>
<tr>
<td>$\sigma_d$</td>
<td>16.0400</td>
</tr>
</tbody>
</table>

Estimated power:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>0.7626</td>
</tr>
</tbody>
</table>

The power of 76.26% of a paired $z$ test is close to the power of 75.45% of a paired $t$ test obtained in example 6.

Example 8: Multiple values of study parameters

Continuing with example 3, we will suppose that we would like to assess the effect of varying correlation on the power of our study. The standard deviation of the MDI scores for infants aged 6 months is 13.85 and that for infants aged 24 months is 12.95, which are obtained from Howell (2002, 193). We believe the data on pairs to be positively correlated because we expect a 6-month-old infant with a high score to have a high score at 24 months of age as well. We specify a range of correlations between 0.1 and 0.9 with the step size of 0.1 in the corr() option:

```
. power pairedmeans 111 106.71, n(100) sd1(13.85) sd2(12.95) corr(0.1(0.1)0.9)
```

Estimated power for a two-sample paired-means test

Paired $t$ test

Ho: $d = d_0$ versus Ha: $d \neq d_0$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
</tr>
<tr>
<td>$N$</td>
<td>100</td>
</tr>
<tr>
<td>$d_0$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\delta$</td>
<td>-4.2900</td>
</tr>
<tr>
<td>$\sigma_d$</td>
<td>16.0400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>0.656</td>
</tr>
<tr>
<td>Corr</td>
<td>0.1</td>
</tr>
<tr>
<td>$\sigma_d$</td>
<td>17.99</td>
</tr>
<tr>
<td>$\delta$</td>
<td>-0.2385</td>
</tr>
</tbody>
</table>

As the correlation increases, the power also increases. This is because the standard deviation of the differences is negatively related to correlation when the correlation is positive. As the correlation increases, the standard deviation of the differences decreases, thus resulting in higher power. Likewise, the opposite is true when the correlation is negative.
Computing effect size and target mean difference

Effect size $\delta$ for a two-sample paired-means test is defined as a standardized difference between the alternative mean difference $d_a$ and the null mean difference $d_0$, $\delta = (d_a - d_0)/\sigma_d$.

Sometimes, we may be interested in determining the smallest effect and the corresponding mean difference that yield a statistically significant result for prespecified sample size and power. In this case, power, sample size, and the alternative pretreatment mean must be specified. By default, the null mean difference is set to 0. In addition, you must also decide on the direction of the effect: upper, meaning $d_a > d_0$, or lower, meaning $d_a < d_0$. The direction may be specified in the direction() option; direction(upper) is the default.

Example 9: Minimum detectable value of the effect size

Continuing with example 6, we may be interested to find the minimum effect size with a power of 80% given a sample of 100 subjects. To compute the smallest effect size and the corresponding target mean difference, we specify the sample size n(100), power power(0.8), and the standard deviation of the differences sddiff(16.04):

```
. power pairedmeans 111, n(100) power(0.8) sddiff(16.04)
Performing iteration ...
Estimated target parameters for a two-sample paired-means test
Paired t test
Ho: d = d0 versus Ha: d != d0; da > d0
Study parameters:
  alpha = 0.0500     ma1 = 111.0000
  power = 0.8000
  N = 100
  d0 = 0.0000
  sd_d = 16.0400
Estimated effect size and target parameters:
  delta = 0.2829
  da = 4.5379
  ma2 = 115.5379
```

The smallest detectable value of the effect size is 0.28, which corresponds to the alternative mean difference of 4.54. Compared with example 1, for the same power of 80%, the target mean difference increased to 4.54 when the sample size was reduced to 100 subjects.

Testing a hypothesis about two correlated means

In this section, we demonstrate the use of the ttest command for testing hypotheses about paired means. Suppose we wish to test the hypothesis that the means of the paired samples are the same. We can use the ttest command to do this. We demonstrate the use of this command using the fictional bpwide dataset; see [R] ttest for details.
Example 10: Testing means from paired data

Suppose that we have a sample of 120 patients. We are interested in investigating whether a certain drug induces a change in the systolic blood pressure. We record blood pressures for each patient before and after the drug is administered. In this case, each patient serves as his or her own control. We wish to test whether the mean difference between the posttreatment and pretreatment systolic blood pressures are significantly different from zero.

```
. use https://www.stata-press.com/data/r16/bpwide
(fictional blood-pressure data)
. ttest bp_before == bp_after
```

```
Paired t test
Mean | Obs  | Mean | Std. Err. | Std. Dev. | [95% Conf. Interval]
-----|------|------|-----------|-----------|------------------
bp_bef| 120  | 156.45 | 1.039746   | 11.38985   | 154.3912  158.5088
bp_after | 120  | 151.3583 | 1.294234  | 14.17762   | 148.7956  153.921
```

```
diff | 120  | 5.091667  | 1.525736 | 16.7136   | 2.070557  8.112776
```

```
mean(diff) = mean(bp_before - bp_after)  t = 3.3372
Ho: mean(diff) = 0  degrees of freedom = 119
Ha: mean(diff) < 0  Ha: mean(diff) != 0  Ha: mean(diff) > 0
Pr(T < t) = 0.9994  Pr(|T| > |t|) = 0.0011  Pr(T > t) = 0.0006
```

We find statistical evidence to reject the null hypothesis of \( H_0: d = 0 \) versus the two-sided alternative \( H_a: d \neq 0 \) at the 5% significance level; the \( p \)-value = 0.0011.

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

```
. power pairedmeans, altdiff(5.09) sddiff(16.71)
Performing iteration ...
Estimated sample size for a two-sample paired-means test
Paired t test
Ho: d = d0  versus  Ha: d != d0
```

```
Study parameters:
alpha = 0.0500
power = 0.8000
delta = 0.3046
d0 = 0.0000
da = 5.0900
sd_d = 16.7100
```

```
Estimated sample size:
N = 87
```

We find that the sample size required to detect a mean difference of 5.09 for given standard deviation of the differences of 16.71 with 80% power using a 5%-level two-sided test is 87.
Video examples

Sample-size calculation for comparing sample means from two paired samples
Power calculation for comparing sample means from two paired samples
Minimum detectable effect size for comparing sample means from two paired samples

Stored results

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

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

Macros
- `r(type)`    : test
- `r(method)`  : `pairedmeans`
- `r(direction)` : upper or lower
- `r(columns)` : displayed table columns
- `r(labels)`  : table column labels
- `r(widths)` : table column widths
- `r(formats)` : table column formats

Matrices
- `r(pss_table)` : table of results

Methods and formulas

Consider a sequence of `n` paired observations denoted by `X_{ij}` for `i = 1, \ldots, n` and groups `j = 1, 2`. Individual observation corresponds to the pair `(X_{i1}, X_{i2})`, and inference is made on the differences within the pairs. Let `d = \mu_2 - \mu_1` denote the mean difference, where `\mu_j` is the population mean of group `j`, and `D_i = X_{i2} - X_{i1}` denote the difference between individual observations. Let `d_0` and `d_a` denote the null and alternative values of the mean difference `d`. Let \( \bar{d} = \sum_{i=1}^{n} D_i / n \) denote the sample mean difference.
Unlike a two-sample means test where we consider two independent samples, a paired-means test allows the two groups to be dependent. As a result, the standard deviation of the differences is given by \( \sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \), where \( \sigma_1 \) and \( \sigma_2 \) are the pretreatment and posttreatment group standard deviations, respectively, and \( \rho \) is the correlation between the paired measurements.

Power, sample-size, and effect-size determination for a paired-means test is analogous to a one-sample mean test where the sample of differences \( D_i \)’s is treated as a single sample. See Methods and formulas in [PSS-2] power onemean.

Also see Armitage, Berry, and Matthews (2002); Dixon and Massey (1983); and Chow et al. (2018) for more details.

References


Also see

[PSS-2] power — Power and sample-size analysis for hypothesis tests

[PSS-2] power repeated — Power analysis for repeated-measures analysis of variance

[PSS-2] power, graph — Graph results from the power command

[PSS-2] power, table — Produce table of results from the power command

[PSS-3] ciwidth pairedmeans — Precision analysis for a paired-means-difference CI

[PSS-5] Glossary

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