

power pcorr — Power analysis for a partial-correlation test in a multiple linear regression

Description
Options
Reference

Quick start
Remarks and examples
Also see

Menu
Stored results

Syntax
Methods and formulas

Description

`power pcorr` computes sample size, power, or target squared partial correlation for a partial-correlation test in a multiple linear regression. A partial-correlation test is an F test of the squared partial multiple correlation that is used to test the significance of a subset of coefficients in a regression model. By default, `power pcorr` computes sample size given power and the squared partial correlation. Alternatively, it computes power given sample size and the squared partial correlation, or it computes the squared partial correlation given sample size and power. See [PSS-2] [power](#) for a general introduction to the `power` command using hypothesis tests.

Quick start

Sample size for a test of $H_0: \rho_p^2 = 0$ versus $H_a: \rho_p^2 \neq 0$ given squared partial correlation of 0.1, 3 tested covariates, and 5 control covariates using default power of 0.8 and significance level $\alpha = 0.05$

```
power pcorr 0.1, ntested(3) ncontrol(5)
```

Power for sample size of 100

```
power pcorr 0.1, ntested(3) ncontrol(5) n(100)
```

Same as above, but for sample sizes of 45, 60, 75, and 90

```
power pcorr 0.1, ntested(3) ncontrol(5) n(45(15)90)
```

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

```
power pcorr 0.1, ntested(3) ncontrol(5) n(45(15)90) graph
```

Effect size and target squared partial correlation for sample size of 100 with power of 0.8

```
power pcorr, ntested(3) ncontrol(5) n(100) power(0.8)
```

Menu

Statistics > Power, precision, and sample size

Syntax

Compute sample size

```
power pcorr rho2_p [ , power(numlist) options ]
```

Compute power

```
power pcorr rho2_p, n(numlist) [options]
```

Compute effect size and target squared partial correlation

```
power pcorr, n(numlist) power(numlist) [options]
```

where *rho2_p* is the hypothesized squared partial correlation in a multiple linear regression. *rho2_p* may be specified either as one number or as a list of values in parentheses (see [U] **11.1.8 numlist**).

<i>options</i>	Description
Main	
* <u>alpha</u> (<i>numlist</i>)	significance level; default is <code>alpha(0.05)</code>
* <u>power</u> (<i>numlist</i>)	power; default is <code>power(0.8)</code>
* <u>beta</u> (<i>numlist</i>)	probability of type II error; default is <code>beta(0.2)</code>
* <u>n</u> (<i>numlist</i>)	sample size; required to compute power or effect size
<u>nfractional</u>	allow fractional sample size
* <u>ntested</u> (<i>numlist</i>)	number of tested covariates; default is <code>ntested(1)</code>
* <u>ncontrol</u> (<i>numlist</i>)	number of control covariates; default is <code>ncontrol(1)</code>
<u>parallel</u>	treat number lists in starred options as parallel when multiple values per option are specified (do not enumerate all possible combinations of values)
Table	
<u>[no]</u> <u>table</u> [(<i>tablespec</i>)]	suppress table or display results as a table; see [PSS-2] power, table
<u>saving</u> (<i>filename</i> [, <i>replace</i>])	save the table data to <i>filename</i> ; use <i>replace</i> to overwrite existing <i>filename</i>
Graph	
<u>graph</u> [(<i>graphopts</i>)]	graph results; see [PSS-2] power, graph
Iteration	
<u>init</u> (#)	initial value for sample size or squared partial correlation
<u>iterate</u> (#)	maximum number of iterations; default is <code>iterate(500)</code>
<u>tolerance</u> (#)	parameter tolerance; default is <code>tolerance(1e-12)</code>
<u>ftolerance</u> (#)	function tolerance; default is <code>ftolerance(1e-12)</code>
<u>[no]</u> <u>log</u>	suppress or display iteration log
<u>[no]</u> <u>dots</u>	suppress or display iterations as dots
<u>notitle</u>	suppress the title

* 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.

`collect` is allowed; see [U] 11.1.10 **Prefix commands**.

`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).

<i>column</i>	Description	Symbol
<code>alpha</code>	significance level	α
<code>power</code>	power	$1 - \beta$
<code>beta</code>	type II error probability	β
<code>N</code>	number of subjects	N
<code>delta</code>	effect size	δ
<code>rho2_p</code>	squared partial multiple correlation	ρ_p^2
<code>ntested</code>	number of tested covariates	N_T
<code>ncontrol</code>	number of control covariates	N_C
<code>target</code>	target parameter; synonym for <code>rho2_p</code>	
<code>_all</code>	display all supported columns	

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

Options

Main

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

`ntested(numlist)` specifies the number of tested covariates. The default is `ntested(1)`.

`ncontrol(numlist)` specifies the number of control covariates or the number of covariates in the reduced model. The default is `ncontrol(1)`.

`parallel`; see [PSS-2] **power**.

Table

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

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

Graph

`graph`, `graph()`; see [PSS-2] **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 squared partial correlation for the effect-size determination. The default is to use a bisection search method to compute an initial value.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS-2] **power**.

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

`notitle`; see [PSS-2] **power**.

Remarks and examples

Remarks are presented under the following headings:

[Introduction](#)
[Using power pcorr](#)
[Computing sample size](#)
[Computing power](#)
[Computing effect size and target squared partial correlation](#)
[Performing hypothesis tests on the partial correlation](#)

`power pcorr` computes sample size, power, and the target squared partial correlation for a multiple regression partial-correlation test. See [PSS-2] [Intro \(power\)](#) for a general introduction to power and sample-size analysis, and see [PSS-2] [power](#) for a general introduction to the `power` command using hypothesis tests.

Introduction

In contrast to a simple linear regression, a multiple regression framework allows researchers to control for additional predictors that may add information to better predict or explain the variation in the dependent variable of interest. There are several scenarios in which we may be interested in knowing whether additional predictors improve the model. Consider an example where a researcher is interested in whether variables such as experience, x_{exp} , and location, x_{loc} , add additional information in explaining variation in wage, y_{wage} , after controlling for education, x_{edu} , and gender, x_{gender} ,

$$y_{\text{wage}} = \beta_0 + \beta_{\text{exp}}x_{\text{exp}} + \beta_{\text{loc}}x_{\text{loc}} + \beta_{\text{edu}}x_{\text{edu}} + \beta_{\text{gender}}x_{\text{gender}} + \varepsilon$$

where ε is an independently and normally distributed error term with mean zero and constant standard deviation σ . One way to test whether experience and location add additional information is to test the joint significance of the coefficients on experience, β_{exp} , and location, β_{loc} . In this case, we perform an F test using the null $H_0: \beta_{\text{exp}} = \beta_{\text{loc}} = 0$. x_{exp} and x_{loc} are the tested covariates; x_{edu} and x_{gender} are the control covariates.

An equivalent test, what we refer to as a partial-correlation test, can be constructed based on the squared partial correlation, ρ_p^2 . The null hypothesis is $H_0: \rho_p^2 = 0$ and, as in the joint test of coefficients, can be tested with an F test. ρ_p^2 is a function of the coefficient of determination, R^2 , which is a measure of the variation of the dependent variable explained by the model that is used to assess overall model fit for a linear regression. Specifically, $\rho_p^2 = (R_F^2 - R_R^2)/(1 - R_R^2)$. R_F^2 is the R^2 of the full model that includes the tested and control covariates, and R_R^2 is the R^2 of the reduced model that includes only the control covariates.

The `power pcorr` command provides power and sample-size analysis for a partial-correlation test. For power analysis for an R^2 test in a multiple linear regression, see [PSS-2] [power rsquared](#). For power analysis for a slope test in a simple linear regression, see [PSS-2] [power oneslope](#).

Using power pcorr

`power pcorr` computes sample size, power, or the target squared partial correlation `rho2_p` for a partial-correlation test in a multiple linear regression. By default, all computations are performed at the significance level of 0.05. You may change the significance level by specifying the `alpha()` option.

By default, the numbers of tested covariates and of control covariates are set to 1. You may change the respective values with the `ntested()` and `ncontrol()` options.

To compute sample size, you must specify the squared partial correlation `rho2_p` 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 squared partial correlation `rho2_p`.

To compute the target partial correlation and effect size, which is defined in terms of the partial correlation as $\delta = \rho_p^2 / (1 - \rho_p^2)$, you must specify the sample size in the `n()` option and the power in the `power()` 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-4] **Unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

`power pcorr`'s computations of sample size and effect size require iteration because the denominator degrees of freedom of the noncentral F distribution depends on the sample size, and the noncentrality parameter depends on the sample size and effect size. The default initial values are obtained using a bisection search method. You may use the `init()` option to specify your own value. The initial value of the sample size must be greater than the number of parameters in the multiple regression model. See [PSS-2] **power** for the descriptions of other options that control the iteration procedure.

Computing sample size

To compute sample size, you must specify the squared partial correlation `rho2_p` 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 partial-correlation test

Consider an example from [Cohen \(1988, 436\)](#) where a psychologist investigates a selection procedure based on job candidates' demographic characteristics used to predict success in a sales position. Suppose we want to conduct a similar study. Data for age, education, and prior experience, our control covariates, are readily available. However, data on verbal aptitude and extraversion, the tested covariates, are costly to obtain, and we decide that these variables are worth including only if the squared partial correlation with the dependent variable is at least 0.0426.

We will conduct a test-case study to determine the minimum sample size required to detect a squared partial correlation of 0.0426 between the dependent variable and the two extra variables, verbal aptitude and extraversion. We compute the minimum sample size required with 80% power at a 5% significance level:

```
. power pcorr 0.0426, ntested(2) ncontrol(3)
Performing iteration ...
Estimated sample size for multiple linear regression
F test for partial correlation
H0: rho2_p = 0 versus Ha: rho2_p != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.0445
      rho2_p =    0.0426
      ncontrol =    3
      ntested =    2
Estimated sample size:
      N =      220
```

We find that a sample of 220 subjects is required to detect a squared partial correlation of 0.0426 associated with the two extra variables, verbal aptitude and extraversion, with 80% power using a 5% level test. The effect size `delta` is calculated using the given information about the hypothesized squared partial correlation; see [Methods and formulas](#) in [\[PSS-2\] power rsquared](#) for details.

As we mentioned in [Using power pcorr](#), sample-size computation requires iteration. The iteration log is suppressed by default, but you can display it by specifying the `log` option.

◀

Computing power

To compute power, you must specify the sample size in the `n()` option and the squared partial correlation `rho2_p`.

► Example 2: Power for a partial-correlation test

Continuing with [example 1](#), suppose that we are designing a new study and anticipate a sample of 200 subjects. Given the study parameters from example 1, we compute the power by specifying the sample size of 200 in the `n()` option:

```
. power pcorr 0.0426, ntested(2) ncontrol(3) n(200)
Estimated power for multiple linear regression
F test for partial correlation
H0: rho2_p = 0 versus Ha: rho2_p != 0
Study parameters:
      alpha =    0.0500
      N =      200
      delta =    0.0445
      rho2_p =    0.0426
      ncontrol =    3
      ntested =    2
Estimated power:
      power =    0.7588
```

With this smaller sample size, the power of the test decreases to about 76%.

◀

➤ Example 3: Multiple values of study parameters

Continuing with [example 2](#), we want to see the effect of sample size on power. We specify a list of sample sizes in the `n()` option:

```
. power pcorr 0.0426, n(50 100 200 400 800) ntested(2) ncontrol(3)
Estimated power for multiple linear regression
F test for partial correlation
H0: rho2_p = 0 versus Ha: rho2_p != 0
```

alpha	power	N	delta	rho2_p	ntested	ncontrol
.05	.2331	50	.0445	.0426	2	3
.05	.4435	100	.0445	.0426	2	3
.05	.7588	200	.0445	.0426	2	3
.05	.972	400	.0445	.0426	2	3
.05	.9999	800	.0445	.0426	2	3

As expected, when the sample size increases, the power tends to get closer to 1.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [\[PSS-2\] power, table](#). If you wish to produce a power plot, see [\[PSS-2\] power, graph](#).



Computing effect size and target squared partial correlation

Effect size δ for a multiple regression, defined in terms of the partial correlation, is $\delta = \rho_p^2 / (1 - \rho_p^2)$. To compute the effect size and target squared partial correlation, you must specify the sample size in the `n()` option and the power in the `power()` option.

➤ Example 4: Minimum detectable squared partial correlation

Continuing with [example 2](#), we may also be interested in finding the minimum value of the squared partial correlation that can be detected with a sample of 200 subjects and 80% power. To compute this, we specify `n(200)` and `power(0.8)`. As before, we use 2 tested covariates and 3 control covariates.

```
. power pcorr, n(200) power(0.8) ntested(2) ncontrol(3)
Performing iteration ...
Estimated squared partial correlation for multiple linear regression
F test for partial correlation
H0: rho2_p = 0 versus Ha: rho2_p != 0
Study parameters:
    alpha =    0.0500
    power =    0.8000
      N =      200
ncontrol =      3
ntested  =      2
Estimated effect size and squared partial correlation:
    delta =    0.0489
    rho2_p =    0.0466
```

The minimum detectable squared partial correlation is 0.0466, which corresponds to an effect size of 0.0489. These values are slightly larger than the values of 0.0426 and 0.0445 that can be detected with the larger sample of 220 subjects from [example 1](#) for the same 80% power.



Performing hypothesis tests on the partial correlation

In this section, we briefly demonstrate the use of the `pcorr` command for estimating partial correlations.

► Example 5: Partial-correlation test for one variable

Suppose that our study goal is to investigate whether the mileage of a car (`mpg`) has an effect on its price (`price`) after controlling for headroom (`headroom`) and trunk space (`trunk`). We compute the partial correlation by using `pcorr` on data about cars from 1978 in `auto.dta`. This preliminary investigation will tell us how many modern cars we should select for our study.

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)

. pcorr price mpg headroom trunk
(obs=74)
```

Partial and semipartial correlations of price with

Variable	Partial corr.	Semipartial corr.	Partial corr. ²	Semipartial corr. ²	Significance value
mpg	-0.3800	-0.3576	0.1444	0.1279	0.0010
headroom	-0.1606	-0.1416	0.0258	0.0201	0.1779
trunk	0.1397	0.1228	0.0195	0.0151	0.2418

We obtain a squared partial correlation of around 0.14 for `mpg`. The significance value is less than 0.05.

Suppose we wish to design a new similar study. We use the estimated squared partial correlation from this study to perform a sample-size analysis. `pcorr` computes the partial correlation for only a single variable, not a group of variables, so our power calculation here uses only one tested covariate; see [R] `pcorr` for details. We use the default value of 1 for `ntested()` and specify 2 control covariates.

```
. power pcorr 0.14, ntested(1) ncontrol(2)

Performing iteration ...

Estimated sample size for multiple linear regression
F test for partial correlation
H0: rho2_p = 0 versus Ha: rho2_p != 0

Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.1628
      rho2_p =    0.1400
      ncontrol =      2
      ntested =      1

Estimated sample size:
      N =      51
```

We find that a sample size of 51 is required to detect a squared partial correlation of 0.14 with 80% power using a 5%-level test.

Stored results

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

Scalars

<code>r(alpha)</code>	significance level
<code>r(power)</code>	power
<code>r(beta)</code>	probability of a type II error
<code>r(delta)</code>	effect size
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified, 0 otherwise
<code>r(rho2_p)</code>	squared partial correlation
<code>r(ntested)</code>	number of tested covariates
<code>r(ncontrol)</code>	number of control covariates
<code>r(separator)</code>	number of lines between separator lines in the table
<code>r(divider)</code>	1 if <code>divider</code> is requested in the table, 0 otherwise
<code>r(init)</code>	initial value for sample size or for squared partial correlation
<code>r(maxiter)</code>	maximum number of iterations
<code>r(iter)</code>	number of iterations performed
<code>r(tolerance)</code>	requested parameter tolerance
<code>r(deltax)</code>	final parameter tolerance achieved
<code>r(ftolerance)</code>	requested distance of the objective function from zero
<code>r(function)</code>	final distance of the objective function from zero
<code>r(converged)</code>	1 if iteration algorithm converged, 0 otherwise

Macros

<code>r(type)</code>	test
<code>r(method)</code>	<code>pcorr</code>
<code>r(columns)</code>	displayed table columns
<code>r(labels)</code>	table column labels
<code>r(widths)</code>	table column widths
<code>r(formats)</code>	table column formats

Matrices

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

Methods and formulas

See *Testing a subset of coefficients: Partial multiple correlation* under *Methods and formulas* in [PSS-2] **power rsquared**.

Reference

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.

Also see

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

[PSS-2] **power oneslope** — Power analysis for a slope test in a simple linear regression

[PSS-2] **power rsquared** — Power analysis for an R^2 test in a multiple linear regression

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

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

[PSS-5] **Glossary**

[R] **pcorr** — Partial and semipartial correlation coefficients

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

