Title

ciwidth onevariance — Precision analysis for a one-variance CI

Description

`ciwidth onevariance` computes sample size, CI width, and probability of CI width for a CI for a population variance. It can compute sample size for a given CI width and probability of CI width. Alternatively, it can compute CI width for a given sample size and probability of CI width. It can also compute probability of CI width for a given sample size and CI width. The computation is available for the variance or the standard deviation. Also see [PSS-3] **ciwidth** for PrSS analysis for other CI methods.

For power and sample-size analysis for a one-sample variance test, see [PSS-2] **power onevariance**.

Quick start

Sample size required for a two-sided 95% CI for a population variance to have a width no larger than 2 with a probability of 90%, using population-variance estimate $v = 4$,

```
ciwidth onevariance 4, width(2) probwidth(0.9)
```

As above, but specify multiple widths and graph the result

```
ciwidth onevariance 4, width(2 3 4) probwidth(0.9) graph
```

CI width for a sample size of 30, with a 90% probability that the CI width will be no larger than the estimated value

```
ciwidth onevariance 4, n(30) probwidth(0.9)
```

As above, but specify standard deviations rather than variances

```
ciwidth onevariance 4, sd n(30) probwidth(0.9)
```

As above, but specify an upper one-sided CI

```
ciwidth onevariance 4, sd n(30) probwidth(0.9) upper
```

Menu

Statistics > Power, precision, and sample size

Syntax

Compute sample size

Variance scale

`ciwidth onevariance` v, `width(`*numlist*`) probwidth(`*numlist*`)` [*options*]

Standard deviation scale

`ciwidth onevariance` s, `sd width(`*numlist*`) probwidth(`*numlist*`)` [*options*]

Compute CI width

Variance scale

`ciwidth onevariance` v, `probwidth(`*numlist*`) n(`*numlist*`)` [*options*]

Standard deviation scale

`ciwidth onevariance` s, `sd probwidth(`*numlist*`) n(`*numlist*`)` [*options*]

Compute probability of CI width

Variance scale

`ciwidth onevariance` v, `width(`*numlist*`) n(`*numlist*`)` [*options*]

Standard deviation scale

`ciwidth onevariance` s, `sd width(`*numlist*`) n(`*numlist*`)` [*options*]

where v and s are variance and standard deviation, respectively. Each argument may be specified either as one number or as a list of values in parentheses (see **[U] 11.1.8 numlist**).

options	Description
`sd`	request computation using the standard-deviation scale; default is the variance scale
Main	
*`level(`*numlist*`)`	confidence level; default is `level(95)`
*`alpha(`*numlist*`)`	significance level; default is `alpha(0.05)`
*`probwidth(`*numlist*`)`	probability of CI width; required to compute sample size and CI width
*`width(`*numlist*`)`	CI width; required to compute sample size and probability of CI width
*`n(`*numlist*`)`	sample size; required to compute CI width and probability of CI width
`nfractional`	allow fractional sample sizes
`lower`	lower one-sided CI; default is two-sided CI
`upper`	upper one-sided CI; default is two-sided CI
`onesided`	synonym for option `upper`
`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-3] **ciwidth, table**
`saving(`*filename* `[, replace])`	save the table data to *filename*; use `replace` to overwrite existing *filename*
Graph	
`graph[(`*graphopts*`)]`	graph results; see [PSS-3] **ciwidth, graph**
Iteration	
`init(`*#*`)`	initial value for sample size; default is to use a closed-form 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.

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

`sd` does not appear in the dialog box; specification of `sd` is done automatically by the dialog box selected.

`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
`level`	confidence level	$100(1-\alpha)$
`alpha`	significance level	α
`N`	number of subjects	N
`Pr_width`	probability of CI width	p_{width}
`width`	CI width	w
`v`	variance	σ^2
`s`	standard deviation	σ
`_all`	display all supported columns	

Column `alpha` is shown in the default table in place of column `level` if `alpha()` is specified.

Column `s` is shown in the default table in place of column `v` if option `sd` is specified.

Options

`sd` specifies that the computation be performed using the standard-deviation scale. The default is to use the variance scale. Specification of the `sd` option is done automatically when the dialog box for standard deviation is selected.

Main

`level()`, `alpha()`, `probwidth()`, `width()`, `n()`, `nfractional`; see [PSS-3] **ciwidth**. The `nfractional` option is allowed only for sample-size determination.

`lower`, `upper`, `onesided`, `parallel`; see [PSS-3] **ciwidth**.

Table

`table`, `table()`, `notable`; see [PSS-3] **ciwidth, table**.

`saving()`; see [PSS-3] **ciwidth**.

Graph

`graph`, `graph()`; see [PSS-3] **ciwidth, graph**. Also see the *column* table for a list of symbols used by the graphs.

Iteration

`init(#)` specifies an initial value for the sample size when iteration is used to compute the sample size. The default is to use a closed-form normal approximation to compute an initial sample size.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS-3] **ciwidth**.

The following option is available with `ciwidth onevariance` but is not shown in the dialog box:

`notitle`; see [PSS-3] **ciwidth**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

This entry describes the `ciwidth onevariance` command and the methodology for PrSS analysis for a CI for a population variance. See [PSS-3] **Intro (ciwidth)** for a general introduction to PrSS analysis, and see [PSS-3] **ciwidth** for a general introduction to the `ciwidth` command. For PSS analysis for hypothesis tests, see [PSS-2] **power**.

Introduction

The study of variance arises in cases where investigators are interested in measuring the variability of a process. For example, the accuracy of a thermometer in taking measurements, the variation in the weights of potato chips from one bag to another, and the variation in mileage across automobiles of the same model. Before undertaking the actual study, we may want to find the optimal sample size to measure the variations with a certain precision.

We are interested in a CI for the population variance σ^2. The precision of a CI is commonly measured by its width w. For example, a two-sided one-variance CI is formed as $[\widehat{\sigma}^2 - w_{\rm lower}, \widehat{\sigma}^2 + w_{\rm upper}]$, where $\widehat{\sigma}^2$ is the variance point estimate. The CI width, the distance between the upper and lower limits, is $w = w_{\rm lower} + w_{\rm upper}$. The smaller the w the more precise the CI.

In PrSS analysis, it is usually of interest to determine the sample size that would be sufficient for a CI to have a prespecified width in a future study. Generally, larger sample sizes lead to more precise CIs. To compute the required sample size, we need to know the expression for w.

Just like with a one-mean CI, the CI width w for a one-variance CI depends on the sample estimate s^2 of the variance and thus will vary from one sample to another. To ensure that, in a future study, a CI has the desired width, this sampling variability of w must be accounted for when computing the required sample size. Kupper and Hafner (1989) introduce what we call the probability of CI width, which specifies the probability of a future CI to have the width of no larger than some prespecified CI width for a given sample size. This probability is defined based on the assumption of a χ^2 distribution for the sample variance s^2; see *Methods and formulas* for details.

You can use `ciwidth onevariance` to perform PrSS analysis for a CI for a population variance or standard deviation. We discuss the command details in the next section.

Using ciwidth onevariance

`ciwidth onevariance` computes sample size, CI width, or probability of CI width for a one-variance CI. By default, a two-sided CI is assumed, and the confidence level is set to 95%. You may change the confidence level by specifying the `level()` option. Alternatively, you can specify the significance level in the `alpha()` option. You can specify the `upper` and `lower` options to request upper and lower one-sided CIs.

To compute sample size, you must specify the CI width in the `width()` option and the probability of CI width in the `probwidth()` option. To compute CI width, you must specify the sample size in the `n()` option and the probability of CI width in the `probwidth()` option. You can also compute the probability of CI width given the sample size in `n()` and CI width in `width()`. In each case, you must also specify the variance v or standard deviation s as the command argument.

By default, the computation is performed for the variance parameter. You can use the `sd` option to specify the computation for the standard deviation.

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 `ciwidth onevariance`'s computations require iteration. For example, the sample-size computation requires iteration. The default initial value of the estimated sample size is obtained by using a closed-form normal approximation. It may be changed by specifying the `init()` option. See [PSS-3] **ciwidth** for the descriptions of other options that control the iteration procedure.

In the following sections, we describe the use of `ciwidth onevariance` accompanied by examples for computing sample size, CI width, and probability of CI width.

Computing sample size

To compute the sample size required for a one-variance CI to have the width no larger than a target width, you must specify the target CI width in the `width()` option and the desired probability of achieving the target CI width in the `probwidth()` option. You must also specify the variance v or standard deviation s as the command argument.

▷ Example 1: Sample size for a one-variance CI

Consider a study where interest lies in measuring the variability in mileage (measured in miles per gallon) of automobiles of a certain car manufacturer. Industry-wide standards maintain that a variation of at most two miles per gallon (mpg) from an average value is acceptable for commercial production.

We want to compute the required sample size such that the width of a two-sided 95% CI for the variance will not exceed 2 mpg with a 96% certainty. Suppose the variance is 4. We specify the variance $v = 4$ after the command name, the CI width of 2 in the `width()` option, and the probability of obtaining the target CI width in the `probwidth()` option:

```
. ciwidth onevariance 4, probwidth(0.96) width(2)
Performing iteration ...
Estimated sample size for a one-variance CI
Chi-squared two-sided CI
Study parameters:
        level =     95.00
     Pr_width =    0.9600
        width =    2.0000
            v =    4.0000
Estimated sample size:
            N =       183
```

We find that a sample of 183 cars is required for this study.

As we mentioned in the previous section, sample-size computation requires iteration. By default, `ciwidth onevariance` suppresses the iteration log, but it can be displayed by specifying the `log` option.

◁

Computing CI width

To compute the CI width, you must specify the sample size in the n() option and the desired probability of achieving the target CI width in the probwidth() option. You must also specify the variance v or standard deviation s as the command argument.

▷ Example 2: CI width for a one-variance CI

Continuing with example 1, suppose that we anticipate obtaining a sample of 150 cars and want to compute the CI width corresponding to this sample size. To compute the CI width, we use the study parameters from example 1, but we now specify the sample size of 150 in the n() option instead of the width() option:

```
. ciwidth onevariance 4, probwidth(0.96) n(150)
Estimated width for a one-variance CI
Chi-squared two-sided CI
Study parameters:
        level =     95.00
            N =       150
     Pr_width =    0.9600
            v =    4.0000
Estimated width:
        width =    2.2571
```

With a sample size smaller than the one we estimated in example 1, the width of the variance CI increases to about 2.3.

◁

▷ Example 3: Standard deviation scale

Continuing with example 2, suppose that we would like to compute the CI width using the standard-deviation scale instead. Above we used a variance of 4; taking its square root, we specify a standard deviation of 2 as the command argument and the sd option. The other parameters remain unchanged from the example above.

```
. ciwidth onevariance 2, probwidth(0.96) n(150) sd
Estimated width for a one-standard-deviation CI
Chi-squared two-sided CI
Study parameters:
        level =     95.00
            N =       150
     Pr_width =    0.9600
            s =    2.0000
Estimated width:
        width =    0.5060
```

For a sample size of 150, probability of CI width of 0.96, and standard deviation of 2, the estimated largest width for a standard-deviation CI is 0.51 mpg.

◁

Computing probability of CI width

To compute the probability that the width of a future CI will be no larger than the specified width, you must specify the sample size in the `n()` option and the target CI width in the `width()` option.

▷ Example 4: Computing probability of CI width for a one-variance CI

Because CI width may vary across samples, we may want to estimate the probability that the width of a future CI will not exceed a target value. Continuing with example 1, suppose that we have only enough resources to test the mileage of 150 automobiles. We can estimate the probability that the CI width will not exceed a target width of 2, given this sample size and a variance of 4:

```
. ciwidth onevariance 4, width(2) n(150)
Estimated probability of width for a one-variance CI
Chi-squared two-sided CI
Study parameters:
        level =    95.00
            N =      150
        width =   2.0000
            v =   4.0000
Estimated probability of width:
     Pr_width =   0.7453
```

For this sample size, we can be 75% certain that the CI width will be no more than 2 mpg for a 95% CI for the variance.

◁

▷ Example 5: Multiple values of study parameters

As a variation of example 4, we would like to see the effect of an increasing variance on the estimated probability of achieving a target CI width of 2. We compute the probability of CI width for a range of variances between 3 and 5, with the step size of 0.5, by specifying the corresponding numlist as the argument for `ciwidth onevariance`.

```
. ciwidth onevariance (3(0.5)5), width(2) n(150)
Estimated probability of width for a one-variance CI
Chi-squared two-sided CI
```

level	N	Pr_width	width	v
95	150	.9996	2	3
95	150	.969	2	3.5
95	150	.7453	2	4
95	150	.3591	2	4.5
95	150	.1074	2	5

The output shows that the probability of achieving the target CI width decreases rapidly as we increase the variance.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS-3] **ciwidth, table**. If you wish to produce sample-size and other curves, see [PSS-3] **ciwidth, graph**.

◁

Stored results

`ciwidth onevariance` stores the following in `r()`:

Scalars

`r(level)`	confidence level
`r(alpha)`	significance level
`r(N)`	sample size
`r(nfractional)`	`1` if `nfractional` is specified, `0` otherwise
`r(onesided)`	`1` for a one-sided CI, `0` otherwise
`r(Pr_width)`	probability of CI width
`r(Pr_width_a)`	actual probability of CI width (for sample-size determination when `probwidth()` specified)
`r(width)`	CI width
`r(v)`	variance
`r(s)`	standard deviation
`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
`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)`	`ci`
`r(method)`	`onevariance`
`r(scale)`	`variance` or `standard deviation`
`r(onesidedci)`	`upper` or `lower` (for a one-sided CI)
`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

See *Methods and formulas* in [R] **ci** for a general description of CIs for variances.

Consider a random sample $\mathbf{x} = (x_1, \ldots, x_n)$ of size n from a normal population with mean μ and variance σ^2. We are interested in a CI for the population variance σ^2.

A general two-sided CI is defined as $[ll(\mathbf{x}), ul(\mathbf{x})]$, a lower one-sided CI as $[ll(\mathbf{x}), \infty)$, and an upper one-sided CI as $(0, ul(\mathbf{x})]$, where $ll(\mathbf{x}) = ll$ and $ul(\mathbf{x}) = ul$ are the respective lower and upper confidence limits. Let w be the CI width.

Let

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{and} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

be the sample mean and the sample variance, respectively.

A two-sided CI for the variance σ^2 is constructed as

$$[s^2 - w_{\text{lower}}, s^2 + w_{\text{upper}}]$$

where w_{lower} and w_{upper} are such that $w_{\text{upper}} + w_{\text{lower}} = w$.

Lower and upper one-sided CIs are constructed as

$$[s^2 - w_{\text{lower}}, \infty)$$

$$(0, s^2 + w_{\text{upper}}]$$

We define $w = w_{\text{lower}}$ for lower one-sided CIs and $w = w_{\text{upper}}$ for upper one-sided CIs.

We use the CI width w as our measure of CI precision. Let $100(1-\alpha)\%$ denote the confidence level, where $0 \le \alpha \le 1$ is the corresponding significance level.

The following formulas are based on Dixon and Massey (1983). The sampling distribution of the statistic $\chi^2 = (n-1)s^2/\sigma^2$ follows a χ^2 distribution with $n-1$ degrees of freedom. Let $\chi^2_{n-1,p}$ be the pth quantile of the χ^2 distribution with $n-1$ degrees of freedom.

Based on the χ^2 distribution, the constructed CIs are:

$$\begin{cases} \left[\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}\right] & \text{for a two-sided CI} \\ \left[\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha}}, \infty\right) & \text{for a lower CI} \\ \left(0, \frac{(n-1)s^2}{\chi^2_{n-1,\alpha}}\right] & \text{for an upper CI} \end{cases}$$

Similarly to the case of an unknown standard deviation for a one-mean CI, the CI width depends on the sample standard deviation. Again, using the fact that $(n-1)s^2/\sigma^2$ follows a χ^2 distribution with $n-1$ degrees of freedom, we can compute the probability that the CI width is no larger than a specified value w.

Let $\chi^2_{n-1}(\cdot)$ be the c.d.f. of the χ^2 distribution with $n-1$ degrees of freedom. The probability of CI width, $\Pr(w)$, is

$$\Pr(w) = \begin{cases} \chi^2_{n-1}\left(\frac{w^2}{\sigma^2(1/\chi^2_{n-1,1-\alpha/2} - 1/\chi^2_{n-1,\alpha/2})}\right) & \text{for a two-sided CI} \\ \chi^2_{n-1}\left(\frac{w^2}{\sigma^2\{1/(n-1) - 1/\chi^2_{n-1,\alpha}\}}\right) & \text{for a lower one-sided CI} \\ \chi^2_{n-1}\left(\frac{w^2}{\sigma^2\{1/\chi^2_{n-1,1-\alpha} - 1/(n-1)\}}\right) & \text{for an upper one-sided CI} \end{cases} \tag{1}$$

We can compute the desired CI width from (1).

$$w = \begin{cases} \sigma^2\chi^2_{n-1,\Pr(w)}\left(1/\chi^2_{n-1,1-\alpha/2} - 1/\chi^2_{n-1,\alpha/2}\right) & \text{for a two-sided CI} \\ \sigma^2\chi^2_{n-1,\Pr(w)}\{1/(n-1) - 1/\chi^2_{n-1,\alpha}\} & \text{for a lower one-sided CI} \\ \sigma^2\chi^2_{n-1,\Pr(w)}\{1/\chi^2_{n-1,1-\alpha} - 1/(n-1)\} & \text{for an upper one-sided CI} \end{cases} \tag{2}$$

where $\chi^2_{n-1,p}$ is the pth quantile of a χ^2 distribution with $n-1$ degrees of freedom.

We can solve for the sample size iteratively using (2). The default initial value for the sample size is computed using the closed-form formula based on a large-sample normal approximation. Specifically, for a large n, the log-transformed sample variance is approximately normally distributed with mean $2\ln(\sigma)$ and standard deviation $\sqrt{2/n}$.

If the nfractional option is not specified, the computed sample size is rounded up.

References

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

Kupper, L. L., and K. B. Hafner. 1989. How appropriate are popular sample size formulas? *American Statistician* 43: 101–105. https://doi.org/10.2307/2684511.

Also see

[PSS-3] **ciwidth** — Precision and sample-size analysis for CIs

[PSS-3] **ciwidth, graph** — Graph results from the ciwidth command

[PSS-3] **ciwidth, table** — Produce table of results from the ciwidth command

[PSS-2] **power onevariance** — Power analysis for a one-sample variance test

[PSS-5] **Glossary**

[R] **ci** — Confidence intervals for means, proportions, and variances