

ciwidth onemean — Precision analysis for a one-mean CI

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

Description

`ciwidth onemean` computes sample size, CI width, and probability of CI width for a CI for a population mean. 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. Also see [PSS-3] [ciwidth](#) for PrSS analysis for other CI methods.

For power and sample-size analysis for a one-sample mean test, see [PSS-2] [power onemean](#).

Quick start

Sample size for a two-sided 95% CI for a population mean given a CI width of 5 and a standard deviation of 12, with the probability of CI width of 0.9

```
ciwidth onemean, width(5) probwidth(0.9) sd(12)
```

As above, but for an upper one-sided CI with a 90% confidence level

```
ciwidth onemean, width(5) probwidth(0.9) sd(12) level(90) upper
```

Sample size for a two-sided 95% CI for a population mean given a CI width of 5, assuming a known population standard deviation of 12

```
ciwidth onemean, width(5) sd(12) knownsd
```

CI width for sample sizes of 50, 60, 70, and 80, given a probability of CI width of 0.9

```
ciwidth onemean, n(50(10)80) probwidth(0.9) sd(12)
```

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

```
ciwidth onemean, n(50(10)80) probwidth(0.9) sd(12) graph
```

Probability that the width of a two-sided 95% CI is no larger than 5 for a sample size of 100

```
ciwidth onemean, n(100) width(5) sd(12)
```

Menu

Statistics > Power, precision, and sample size

Syntax

Compute sample size

```
ciwidth onemean, width(numlist) probwidth(numlist) [options]
```

Compute CI width

```
ciwidth onemean, probwidth(numlist) n(numlist) [options]
```

Compute probability of CI width

```
ciwidth onemean, width(numlist) n(numlist) [options]
```

<i>options</i>	Description
Main	
* <u>l</u> evel(<i>numlist</i>)	confidence level; default is level(95)
* <u>a</u> lpha(<i>numlist</i>)	significance level; default is alpha(0.05)
* <u>p</u> robwidth(<i>numlist</i>)	probability of CI width; required to compute sample size and CI width
* <u>w</u> idth(<i>numlist</i>)	CI width; required to compute sample size and probability of CI width
* <u>n</u> (<i>numlist</i>)	sample size; required to compute CI width and probability of CI width
<u>n</u> fractional	allow fractional sample sizes
* <u>s</u> d(<i>numlist</i>)	standard deviation; default is sd(1)
<u>k</u> nownsd	request computation assuming a known standard deviation; default is to assume an unknown standard deviation
* <u>f</u> pc(<i>numlist</i>)	finite population correction (FPC) as a sampling rate or as a population size
<u>l</u> ower	lower one-sided CI; default is two-sided CI
<u>u</u> pper	upper one-sided CI; default is two-sided CI
<u>o</u> nesided	synonym for option upper
<u>p</u> arallel	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	
<code>[no]table [(tablespec)]</code>	suppress table or display results as a table; see [PSS-3] ciwidth, table
<code>saving(filename [, replace])</code>	save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>
Graph	
<code>graph [(graphopts)]</code>	graph results; see [PSS-3] ciwidth, graph
Iteration	
<code>init(#)</code>	initial value for sample size; default is to use a closed-form normal approximation
<code>iterate(#)</code>	maximum number of iterations; default is <code>iterate(500)</code>
<code>tolerance(#)</code>	parameter tolerance; default is <code>tolerance(1e-12)</code>
<code>ftolerance(#)</code>	function tolerance; default is <code>ftolerance(1e-12)</code>
<code>[no]log</code>	suppress or display iteration log
<code>[no]dots</code>	suppress or display iterations as dots
<code>notitle</code>	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>level</code>	confidence level	$100(1 - \alpha)$
<code>alpha</code>	significance level	α
<code>N</code>	number of subjects	N
<code>Pr_width</code>	probability of CI width	p_{width}
<code>width</code>	CI width	w
<code>sd</code>	standard deviation	σ
<code>fpc</code>	FPC as population size	N_{pop}
	FPC as sampling rate	γ
<code>_all</code>	display all supported columns	

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

Column `fpc` is shown in the default table if `fpc()` is specified.

Options

Main

`level()`, `alpha()`, `probwidth()`, `width()`, `n()`, `nfractional`; see [PSS-3] **ciwidth**. `probwidth()` may not be combined with `knownsd`. The `nfractional` option is allowed only for sample-size determination.

`sd(numlist)` specifies the population standard deviation or its estimate. The default is `sd(1)`. By default, `sd()` specifies an estimate for the unknown population standard deviation. If `knownsd` is specified, `sd()` specifies the known value for the population standard deviation.

`knownsd` requests that the standard deviation be treated as known in the computation. By default, the standard deviation is treated as unknown, and the computation is performed for a Student's *t*-based CI. If `knownsd` is specified, the computation is performed for a normal-based CI. `knownsd` may not be combined with `probwidth()` and is not allowed when computing the probability of CI width.

`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.

`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 onemean` but is not shown in the dialog box:

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

Remarks and examples

[stata.com](http://www.stata.com)

Remarks are presented under the following headings:

Introduction
Using ciwidth onemean
Computing sample size
Computing CI width
Computing probability of CI width

This entry describes the `ciwidth onemean` command and the methodology for PrSS analysis for a CI for a population mean. 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

Mean estimation is one of the most popular statistical analyses. Schools may measure the mean test score for their students. Factories may measure the mean strength of new alloys or the mean lifetime of bulbs. CIs are often used for inference about the population mean μ . They provide the ranges for plausible values for the mean based on a random sample from a population of interest. The wider the ranges, the less precise the CI.

The precision of a CI is commonly measured by its width w or, for a symmetric CI such as the CI for a population mean, by its half-width d , also known as the margin of error. For example, a two-sided one-mean CI is formed as $[\hat{\mu} - d, \hat{\mu} + d]$, where $\hat{\mu}$ is the mean point estimate. The CI width, the distance between the upper and lower limits, is $w = 2d$; it does not depend on the mean estimate. The smaller the d or 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 . The expression for w depends on various assumptions.

Suppose that we have a random sample of n i.i.d. observations from a normal distribution with mean μ and standard deviation σ . If we know a population standard deviation σ , $w = 2z_{1-\alpha/2}\sigma/\sqrt{n}$, where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ th quantile of the standard normal distribution. If we do not know a σ and estimate it from the sample using the sample standard deviation s , $w = 2t_{n-1,1-\alpha/2} s/\sqrt{n}$, where $t_{n-1,1-\alpha/2}$ is the $(1 - \alpha/2)$ th quantile of the Student's t distribution with $n - 1$ degrees of freedom. $100(1 - \alpha)\%$ is the confidence level of a CI. Using the relationship between CIs and hypothesis tests, a $100(1 - \alpha)\%$ one-mean CI can be viewed as a set of hypothetical values of a mean that cannot be rejected by the corresponding one-sample mean test at the significance level α .

In the case of an unknown standard deviation, the CI width w depends on the sample estimate s of the standard deviation 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 that 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.

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.

You can use `ciwidth onemean` to perform PrSS analysis for a CI for a population mean. We discuss the command details in the next section.

Using ciwidth onemean

`ciwidth onemean` computes sample size, CI width, or probability of CI width for a one-mean 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 `prowidth()` option. To compute CI width, you must specify the sample size in the `n()` option and the probability of CI width in the `prowidth()` option. You can also compute the probability of CI width given the sample size in `n()` and CI width in `width()`.

For CIs for means, the CI width does not depend on the mean point estimate, the sample mean, so it is not needed in the computations.

By default, all computations assume an unknown standard deviation and use a default value of 1 as the estimate of the standard deviation. You may specify other values for the standard deviation in the `sd()` option. For a known standard deviation, you can specify the `knownsd` option to request a normal-based CI instead of the default Student's *t*-based CI. In this case, the `sd()` option specifies the actual population 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 onemean`'s computations require iteration. For example, when the standard deviation is unknown, 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.

All computations assume an infinite population. For a finite population, use the `fpc()` option to specify a sampling rate or a population size.

In the following sections, we describe the use of `ciwidth onemean` 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-mean 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 `prowidth()` option.

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

Consider an example from [Meeker, Hahn, and Escobar \(2017, 152\)](#) of an experiment that measures the mean tensile strength of a new alloy. The number of specimens needed to be tested is of interest. We assume that the strength follows a normal distribution and the true population standard deviation is unknown. We use a conservative estimate of 2,500 kg for it in this example. We want to compute the sample size required for the two-sided 95% CI for the mean tensile strength to have the width of no larger than 3,000 kg. In addition to the sample size, the CI width depends on the standard deviation. Because we assume an unknown standard deviation, the CI width may vary from one study to another. We must account for this sampling variability when estimating the sample size. So, to ensure that the CI will have the width of at most 3,000 kg, we must also specify the probability of achieving the target CI width. If we do not, our results will be based on the assumption that a future sample will have the standard deviation of 2,500 kg, which may not be a realistic assumption given an unknown standard deviation. In our example, we will use the probability of CI width of 0.96. To compute the sample size, we type

```
. ciwidth onemean, sd(2500) probwidth(0.96) width(3000)
Performing iteration ...
Estimated sample size for a one-mean CI
Student's t two-sided CI
Study parameters:
    level =    95.00
   Pr_width =   0.9600
    width = 3000.0000
     sd = 2500.0000
Estimated sample size:
    N =      20
```

We find that a sample of 20 specimens is required to obtain a two-sided 95% CI for the mean that, with a probability of 0.96, will have the width of no more than 3,000 kg.

As we mentioned in *Using ciwidth onemean* and as is also indicated in the output, sample-size computation requires iteration when the standard deviation is unknown. The iteration log is suppressed by default, but you can display it by specifying the `log` option.

◀

▷ Example 2: Known variance

If we know the population standard deviation, we can use the `knownsd` option to compute results for a normal z -based CI. Because we assume a known standard deviation, the `probwidth()` option is not needed. In fact, it is not allowed in combination with `knownsd`.

```
. ciwidth onemean, sd(2500) width(3000) knownsd
Estimated sample size for a one-mean CI
Normal two-sided CI
Study parameters:
    level =    95.00
   width = 3000.0000
     sd = 2500.0000
Estimated sample size:
    N =      11
```

The output now indicates that the computation is based on a normal CI instead of the default Student's t CI. We find that a smaller sample of 11 subjects is required for the mean CI to achieve the same CI width as in [example 1](#) when the standard deviation is known.

◀

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.

▷ Example 3: Precision of a one-mean CI

Continuing with [example 1](#), we can compute the CI width for a given sample size. Suppose we want to compute the CI width corresponding to the sample size of 20 from [example 1](#). We replace the `width(3000)` option with the `n(20)` option:

```
. ciwidth onemean, sd(2500) probwidth(0.96) n(20)
Estimated width for a one-mean CI
Student's t two-sided CI
Study parameters:
    level =    95.00
      N =     20
Pr_width =    0.9600
      sd = 2500.0000
Estimated width:
    width = 2990.8196
```

The estimated CI width is about 2,991 kg and is smaller than the width of 3,000 kg used in [example 1](#). The actual sample size corresponding to the CI width of 3,000 is slightly smaller than 20. You can see this by specifying the `nfractional` option in [example 1](#) to report a fractional sample size. By default, `ciwidth` methods round the actual sample size up to the nearest integer. The CI width corresponding to a larger sample size will be smaller.

◀

► Example 4: One-sided CI

You can specify the `upper` or `lower` option to request an upper or lower one-sided CI. Continuing with [example 3](#), we estimate the CI width for an upper 95% CI.

```
. ciwidth onemean, sd(2500) probwidth(0.96) n(20) upper
Estimated width for a one-mean CI
Student's t upper CI
Study parameters:
    level =    95.00
      N =     20
Pr_width =    0.9600
      sd = 2500.0000
Estimated width:
    width = 1235.4192
```

The upper bound for the mean tensile strength of a new alloy is about 1,235 kg from its point estimate.

◀

► Example 5: Multiple values of study parameters

To investigate the effect of the sample size on the CI width, we can specify a list of sample sizes in the `n()` option:

```
. ciwidth onemean, sd(2500) probwidth(0.96) n(10(10)50)
Estimated width for a one-mean CI
Student's t two-sided CI
```

level	N	Pr_width	width	sd
95	10	.96	5003	2500
95	20	.96	2991	2500
95	30	.96	2289	2500
95	40	.96	1912	2500
95	50	.96	1669	2500

◀

As expected, when the sample size increases, the CI width decreases. The decrease is particularly prominent when the sample size increases from 10 to 20.

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](#).

◀

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 6: Computing probability of CI width for a one-mean CI

Continuing with [example 1](#), we may also want to know the probability that the CI width in a future study will be no larger than a prespecified value for a given sample size. To compute the probability of CI width, we specify the sample size of 20 in `n()` and the CI width of 3,000 in `width()`. We use the same estimate of 2,500 for the standard deviation.

```
. ciwidth onemean, sd(2500) n(20) width(3000)
Estimated probability of width for a one-mean CI
Student's t two-sided CI
Study parameters:
      level =      95.00
         N  =        20
      width = 3000.0000
         sd = 2500.0000
Estimated probability of width:
      Pr_width =      0.9619
```

The estimated probability is about 96%, which is consistent with what we used in [example 1](#).

◀

Stored results

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

Scalars

<code>r(level)</code>	confidence level
<code>r(alpha)</code>	significance level
<code>r(N)</code>	sample size
<code>r(nfractional)</code>	1 if <code>nfractional</code> is specified, 0 otherwise
<code>r(onesided)</code>	1 for a one-sided CI, 0 otherwise
<code>r(Pr_width)</code>	probability of CI width
<code>r(Pr_width_a)</code>	actual probability of CI width (for sample-size determination when <code>prowidth()</code> specified)
<code>r(width)</code>	CI width
<code>r(width_a)</code>	actual CI width (for sample-size determination when <code>knownsd</code> specified)
<code>r(sd)</code>	standard deviation
<code>r(knownsd)</code>	1 if option <code>knownsd</code> is specified, 0 otherwise
<code>r(fpc)</code>	finite population correction (if specified)
<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

<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>	ci
<code>r(method)</code>	onemean
<code>r(onesidedci)</code>	upper or lower (for a one-sided CI)
<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

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a random sample of i.i.d. observations from a normal population with mean μ and 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 $(-\infty, ul(\mathbf{x})]$, 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

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

be the sample mean and the sample variance, respectively.

A two-sided CI for the population mean μ is constructed as

$$[\bar{x} - w/2, \bar{x} + w/2]$$

where $w/2$ is the half-width or margin of error.

Lower and upper one-sided CIs are constructed as

$$\begin{aligned} &[\bar{x} - w, \infty) \\ &(-\infty, \bar{x} + w] \end{aligned}$$

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

The formulas below are based on [Kupper and Hafner \(1989\)](#), [Ryan \(2013\)](#), [Dixon and Massey \(1983\)](#), [Zar \(2010\)](#), and [Chow et al. \(2018\)](#).

Methods and formulas are presented under the following headings:

Known standard deviation
Unknown standard deviation
Finite population size

Known standard deviation

In the case of a known standard deviation, the sampling distribution of the statistic $z = \sqrt{n}(\bar{x} - \mu)/\sigma$ follows the standard normal distribution. Let $z_{1-\alpha}$ be the $(1 - \alpha)$ th quantile of the standard normal distribution.

Based on the normal distribution of z , the corresponding two-sided, lower, and upper CIs are

$$\begin{aligned} & \left[\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \\ & \left[\bar{x} - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}, \infty \right) \\ & \left(-\infty, \bar{x} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right] \end{aligned}$$

The corresponding CI width is

$$w = \begin{cases} \frac{2z_{1-\alpha/2}\sigma}{\sqrt{n}} & \text{for a two-sided CI} \\ \frac{z_{1-\alpha}\sigma}{\sqrt{n}} & \text{for lower and upper one-sided CIs} \end{cases} \quad (1)$$

The sample size n is computed by inverting (1):

$$n = \begin{cases} \left(\frac{2z_{1-\alpha/2}\sigma}{w} \right)^2 & \text{for a two-sided CI} \\ \left(\frac{z_{1-\alpha}\sigma}{w} \right)^2 & \text{for lower and upper one-sided CIs} \end{cases} \quad (2)$$

Unknown standard deviation

In the case of an unknown standard deviation, an unbiased estimator s is used in place of σ in the definition of a z statistic from [Known standard deviation](#). The sampling distribution of the corresponding statistic $t = \sqrt{n}(\bar{x} - \mu)/s$ follows a Student's t distribution with $n - 1$ degrees of freedom. Let $t_{n-1,1-\alpha}$ be the $(1 - \alpha)$ th quantile of the Student's t distribution with $n - 1$ degrees of freedom.

The corresponding two-sided, lower, and upper CIs are

$$\begin{aligned} & \left[\bar{x} - t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}} \right] \\ & \left[\bar{x} - t_{n-1,1-\alpha} \frac{s}{\sqrt{n}}, \infty \right) \\ & \left(-\infty, \bar{x} + t_{n-1,1-\alpha} \frac{s}{\sqrt{n}} \right] \end{aligned}$$

The CI width depends on the sample standard deviation s , which will vary between random samples. To account for the sampling variability of s and consequently of the CI width, [Kupper and Hafner \(1989\)](#) consider the probability of CI width, $\Pr(w)$, such that

$$\Pr\left(2t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}} \leq w\right) \geq \Pr(w) \quad (3)$$

The above is for a two-sided CI, but a similar expression can be constructed for a one-sided CI. $\Pr(w)$ is the probability that the width of a future CI will not exceed some prespecified width w .

Formula (3) can be rewritten as

$$\Pr\left\{\frac{(n-1)s^2}{\sigma^2} \geq \frac{n(n-1)w^2}{4\sigma^2 t_{n-1,1-\alpha/2}^2}\right\} \geq \Pr(w) \quad (4)$$

When data \mathbf{x} are normally distributed, the sampling distribution of the statistic $(n-1)s^2/\sigma^2$ is a χ^2 distribution with $n-1$ degrees of freedom. Then, using (4), we can compute the probability of CI width as follows:

$$\Pr(w) = \begin{cases} \chi_{n-1}^2 \left\{ \frac{n(n-1)w^2}{4\sigma^2 t_{n-1,1-\alpha/2}^2} \right\} & \text{for a two-sided CI} \\ \chi_{n-1}^2 \left\{ \frac{n(n-1)w^2}{\sigma^2 t_{n-1,1-\alpha}^2} \right\} & \text{for lower and upper one-sided CIs} \end{cases} \quad (5)$$

where $\chi_{n-1}^2(\cdot)$ is the c.d.f. of a χ^2 distribution with $n-1$ degrees of freedom.

We can compute the CI width from (5) as follows:

$$w = \begin{cases} 2t_{n-1,1-\alpha/2}\sigma\sqrt{\frac{\chi_{n-1}^2, \Pr(w)}{n(n-1)}} & \text{for a two-sided CI} \\ t_{n-1,1-\alpha}\sigma\sqrt{\frac{\chi_{n-1}^2, \Pr(w)}{n(n-1)}} & \text{for lower and upper one-sided CIs} \end{cases} \quad (6)$$

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

We solve for the sample size iteratively using (6) with initial values obtained from (2).

Finite population size

The above formulas assume that the random sample is drawn from an infinite population. In cases when the size of the population is known, we need to make the following adjustment to the standard deviation,

$$\sigma_{\text{fpc}} = \sigma\sqrt{\left(1 - \frac{n}{N}\right)}$$

where σ_{fpc} is the population standard deviation adjusted for finite population size.

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

References

- Chow, S.-C., J. Shao, H. Wang, and Y. Lohknygina. 2018. *Sample Size Calculations in Clinical Research*. 3rd ed. Boca Raton, FL: CRC Press.
- 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>.
- Meeker, W. Q., G. J. Hahn, and L. A. Escobar. 2017. *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. Hoboken, NJ: Wiley.
- Ryan, T. P. 2013. *Sample Size Determination and Power*. Hoboken, NJ: Wiley.
- Zar, J. H. 2010. *Biostatistical Analysis*. 5th ed. Upper Saddle River, NJ: Pearson.

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 onemean** — Power analysis for a one-sample mean test
- [PSS-5] **Glossary**
- [R] **ci** — Confidence intervals for means, proportions, and variances