

Chapter 421

Confidence Intervals for One Mean with Tolerance Probability

Introduction

This procedure calculates the sample size necessary to achieve a specified distance from the mean to the confidence limit(s) with a given tolerance probability at a stated confidence level for a confidence interval about a single mean when the underlying data distribution is normal.

Technical Details

For a single mean from a normal distribution with unknown variance, a two-sided, $100(1 - \alpha)\%$ confidence interval is calculated by

$$\bar{X} \pm \frac{t_{1-\alpha/2, n-1} \hat{\sigma}}{\sqrt{n}}$$

A one-sided $100(1 - \alpha)\%$ upper confidence limit is calculated by

$$\bar{X} + \frac{t_{1-\alpha, n-1} \hat{\sigma}}{\sqrt{n}}$$

Similarly, the one-sided $100(1 - \alpha)\%$ lower confidence limit is

$$\bar{X} - \frac{t_{1-\alpha, n-1} \hat{\sigma}}{\sqrt{n}}$$

Each confidence interval is calculated using an estimate of the mean plus and/or minus a quantity that represents the distance from the mean to the edge of the interval. For two-sided confidence intervals, this distance is sometimes called the precision, margin of error, or half-width. We will label this distance, D .

The basic equation for determining sample size when D has been specified is

$$D = \frac{t_{1-\alpha/2, n-1} \hat{\sigma}}{\sqrt{n}}$$

Confidence Intervals for One Mean with Tolerance Probability

Solving for n , we obtain

$$n = \left(\frac{t_{1-\alpha/2, n-1} \hat{\sigma}}{D} \right)^2$$

This equation can be solved for any of the unknown quantities in terms of the others. The value $\alpha/2$ is replaced by α when a one-sided interval is used.

There is an additional subtlety that arises when the standard deviation is to be chosen for estimating sample size. The sample sizes determined from the formula above produce confidence intervals with the specified widths only when the future sample has a sample standard deviation that is no greater than the value specified.

As an example, suppose that 15 individuals are sampled in a pilot study, and a standard deviation estimate of 3.5 is obtained from the sample. The purpose of a later study is to estimate the mean within 10 units. Suppose further that the sample size needed is calculated to be 57 using the formula above with 3.5 as the estimate for the standard deviation. The sample of size 57 is then obtained from the population, but the standard deviation of the 57 individuals turns out to be 3.9 rather than 3.5. The confidence interval is computed and the distance from the mean to the confidence limits is greater than 10 units.

This example illustrates the need for an adjustment to adjust the sample size such that the distance from the mean to the confidence limits will be below the specified value with known probability.

Such an adjustment for situations where a previous sample is used to estimate the standard deviation is derived by Harris, Horvitz, and Mood (1948) and discussed in Zar (1984) and Hahn and Meeker (1991). The adjustment is

$$n = \left(\frac{t_{1-\alpha/2, n-1} \hat{\sigma}}{D} \right)^2 F_{1-\gamma; n-1, m-1}$$

where $1 - \gamma$ is the probability that the distance from the mean to the confidence limit(s) will be below the specified value, and m is the sample size in the previous sample that was used to estimate the standard deviation.

The corresponding adjustment when no previous sample is available is discussed in Kupper and Hafner (1989) and Hahn and Meeker (1991). The adjustment in this case is

$$n = \left(\frac{t_{1-\alpha/2, n-1} \hat{\sigma}}{D} \right)^2 \left(\frac{\chi_{1-\gamma, n-1}^2}{n-1} \right)$$

where, again, $1 - \gamma$ is the probability that the distance from the mean to the confidence limit(s) will be below the specified value.

Each of these adjustments accounts for the variability in a future estimate of the standard deviation. In the first adjustment formula (Harris, Horvitz, and Mood, 1948), the distribution of the standard deviation is based on the estimate from a previous sample. In the second adjustment formula, the distribution of the standard deviation is based on a specified value that is assumed to be the population standard deviation.

Finite Population Size

The above calculations assume that samples are being drawn from a large (infinite) population. When the population is of finite size (N), an adjustment must be made. The adjustment reduces the standard deviation as follows:

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

This new standard deviation replaces the regular standard deviation in the above formulas.

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean is $1 - \alpha$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters.

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

Population

Population Size

This is the number of individuals in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made. This option sets the population size.

Confidence Intervals for One Mean with Tolerance Probability

Confidence and Tolerance

Confidence Level ($1 - \alpha$)

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90*, *0.95* or *0.90 to 0.99 by 0.01*.

Tolerance Probability

This is the probability that a future interval with sample size N and the specified confidence level will have a distance from the mean to the limit(s) that is less than or equal to the distance specified.

If a tolerance probability is not used, as in the 'Confidence Intervals for One Mean' procedure, the sample size is calculated for the expected distance from the mean to the limit(s), which assumes that the future standard deviation will also be the one specified.

Using a tolerance probability implies that the standard deviation of the future sample will not be known in advance, and therefore, an adjustment is made to the sample size formula to account for the variability in the standard deviation. Use of a tolerance probability is similar to using an upper bound for the standard deviation in the 'Confidence Intervals for One Mean' procedure.

Values between 0 and 1 can be entered. The choice of the tolerance probability depends upon how important it is that the distance from the interval limit(s) to the mean is at most the value specified.

You can enter a range of values such as *0.70 0.80 0.90* or *0.70 to 0.95 by 0.05*.

Sample Size

N (Sample Size)

Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

You can enter a single value or a range of values.

Precision

Distance from Mean to Limit(s)

This is the distance from the confidence limit(s) to the mean. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Confidence Intervals for One Mean with Tolerance Probability

Standard Deviation

Standard Deviation Source

This procedure permits two sources for estimates of the standard deviation:

- **S is a Population Standard Deviation**

This option should be selected if there is no previous sample that can be used to obtain an estimate of the standard deviation. In this case, the algorithm assumes that future sample obtained will be from a population with standard deviation S .

- **S from a Previous Sample**

This option should be selected if the estimate of the standard deviation is obtained from a previous random sample from the same distribution as the one to be sampled. The sample size of the previous sample must also be entered under 'Sample Size of Previous Sample'.

Standard Deviation – Population Standard Deviation

S (Standard Deviation)

Enter an estimate of the standard deviation (must be positive). In this case, the algorithm assumes that future samples obtained will be from a population with standard deviation S .

One common method for estimating the standard deviation is the range divided by 4, 5, or 6.

You can enter a range of values such as *1 2 3* or *1 to 10 by 1*.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

Standard Deviation – Standard Deviation from Previous Sample

S (SD Estimated from a Previous Sample)

Enter an estimate of the standard deviation from a previous (or pilot) study. This value must be positive.

A range of values may be entered.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

Sample Size of Previous Sample

Enter the sample size that was used to estimate the standard deviation entered in S (SD Estimated from a Previous Sample).

This value is entered only when 'Standard Deviation Source:' is set to 'S from a Previous Sample'.

Confidence Intervals for One Mean with Tolerance Probability

Example 1 – Calculating Sample Size

A researcher would like to estimate the mean weight of a population with 95% confidence. It is very important that the mean weight is estimated within 15 grams. Data available from a previous study are used to provide an estimate of the standard deviation. The estimate of the standard deviation is 45.1 grams, from a sample of size 14. The goal is to determine the sample size necessary to obtain a two-sided confidence interval such that the mean weight is estimated within 15 grams. Tolerance probabilities of 0.70 to 0.95 will be examined.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Mean with Tolerance Probability** procedure window by expanding **Means**, then **One Mean**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Mean with Tolerance Probability**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Interval Type	Two-Sided
Population Size.....	Infinite
Confidence Level.....	0.95
Tolerance Probability	0.70 to 0.95 by 0.05
Distance from Mean to Limit(s).....	15
Standard Deviation Source.....	S from a Previous Sample
S.....	45.1
Sample Size of Previous Sample	14

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Confidence Intervals

Confidence Level	Sample Size (N)	Target Distance from Mean to Limits	Actual Distance from Mean to Limits	Standard Deviation (S)	Tolerance Probability
0.950	49	15.000	14.899	45.100	0.700
0.950	52	15.000	14.959	45.100	0.750
0.950	56	15.000	14.969	45.100	0.800
0.950	61	15.000	14.997	45.100	0.850
0.950	69	15.000	14.925	45.100	0.900
0.950	82	15.000	14.943	45.100	0.950

Sample size for estimate of S from previous sample = 14.

References

- Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.
 Zar, J. H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.
 Harris, M., Horvitz, D. J., and Mood, A. M. 1948. 'On the Determination of Sample Sizes in Designing Experiments', Journal of the American Statistical Association, Volume 43, No. 243, pp. 391-402.

Confidence Intervals for One Mean with Tolerance Probability

Report Definitions

Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population mean.

N is the size of the sample drawn from the population.

Distance from Mean to Limit is the distance from the confidence limit(s) to the mean. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

Target Distance from Mean to Limit is the value of the distance that is entered into the procedure.

Actual Distance from Mean to Limit is the value of the distance that is obtained from the procedure.

The standard deviation of the population measures the variability in the population.

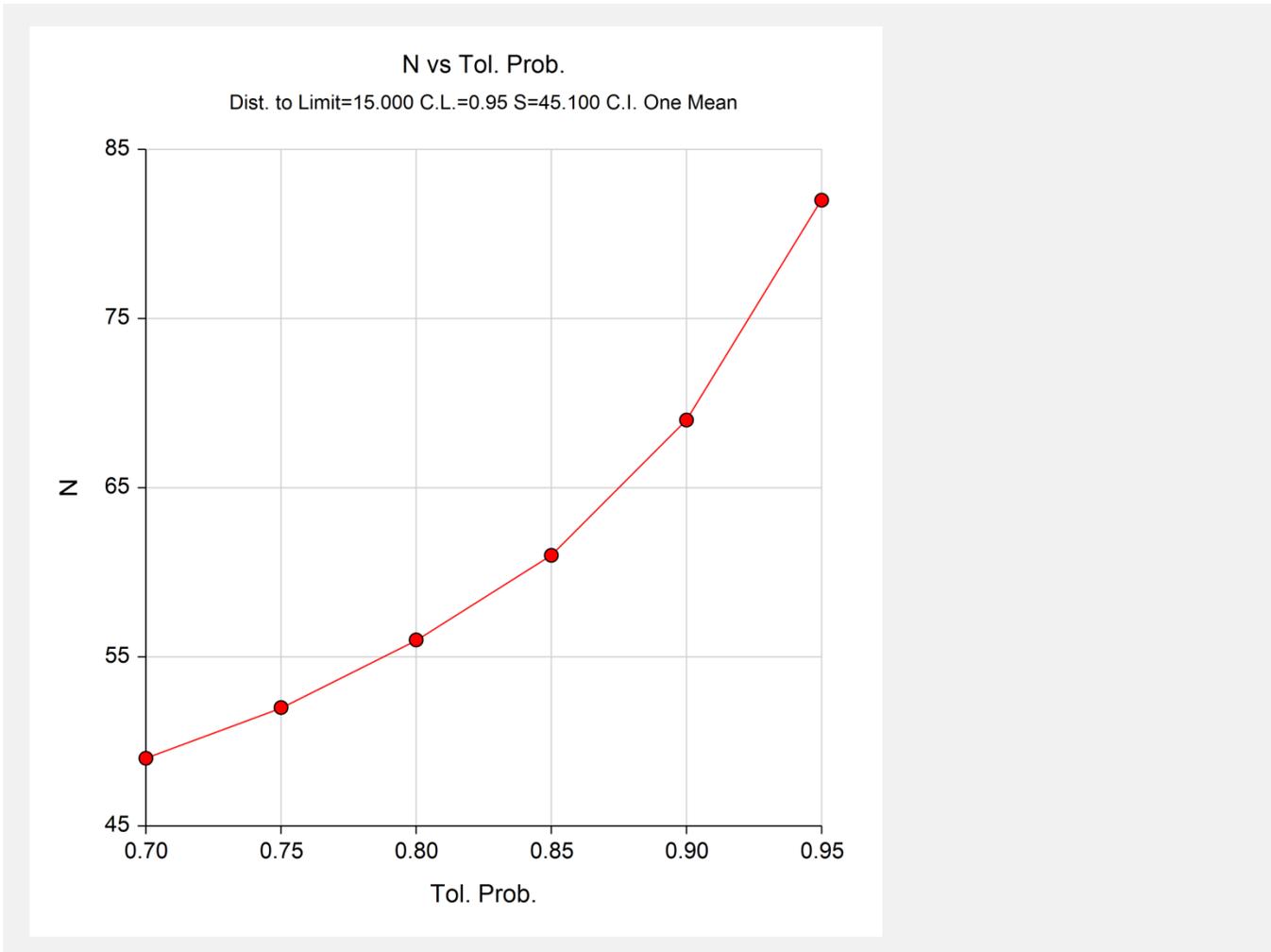
Tolerance Probability is the probability that a future interval with sample size N and corresponding confidence level will have a distance from the mean to the limit(s) that is less than or equal to the specified distance.

Summary Statements

The probability is 0.700 that a sample size of 49 will produce a two-sided 95% confidence interval with a distance from the mean to the limits that is less than or equal to 14.899 if the population standard deviation is estimated to be 45.100 by a previous sample of size 14.

This report shows the calculated sample size for each of the scenarios.

Plots Section



This plot shows the sample size versus the tolerance probability.

Example 2 – Validation using Hahn and Meeker

Hahn and Meeker (1991) page 139 give an example of a sample size calculation for a two-sided confidence interval on the mean when the confidence level is 95%, the population standard deviation is assumed to be 2500, the distance from the mean to the limit is 1500, and the tolerance probability is 0.90. The necessary sample size is 19.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Mean with Tolerance Probability** procedure window by expanding **Means**, then **One Mean**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Mean with Tolerance Probability**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Interval Type	Two-Sided
Population Size	Infinite
Confidence Level	0.95
Tolerance Probability	0.90
Distance from Mean to Limit(s)	1500
Standard Deviation Source	S is a Population Standard Deviation
S	2500

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Confidence Intervals					
Confidence Level	Sample Size (N)	Target Distance from Mean to Limits	Actual Distance from Mean to Limits	Standard Deviation (S)	Tolerance Probability
0.95	19	1500.000	1447.889	2500.000	0.90

PASS also calculated the necessary sample size to be 19.

Example 3 – Validation using Zar

Zar (1984) pages 109-110 give an example of a sample size calculation for a two-sided confidence interval on the mean when the confidence level is 95%, the standard deviation is estimated to be 4.247211 by a previous sample of size 25, the distance from the mean to the limit is 1.5, and the tolerance probability is 0.90. The necessary sample size is 53.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Mean with Tolerance Probability** procedure window by expanding **Means**, then **One Mean**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Mean with Tolerance Probability**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Interval Type	Two-Sided
Population Size	Infinite
Confidence Level	0.95
Tolerance Probability	0.90
Distance from Mean to Limit(s)	1.5
Standard Deviation Source	S from a Previous Sample
S	4.247211
Sample Size of Previous Sample	25

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Confidence Intervals					
Confidence Level	Sample Size (N)	Target Distance from Mean to Limits	Actual Distance from Mean to Limits	Standard Deviation (S)	Tolerance Probability
0.95	53	1.500	1.489	4.247	0.90

PASS also calculated the necessary sample size to be 53.

Example 4 – Validation using Harris, Horvitz, and Mood

Harris, Horvitz, and Mood (1948) pages 392-393 give an example of a sample size calculation for a two-sided confidence interval on the mean when the confidence level is 99%, the standard deviation is estimated to be 3 by a previous sample of size 9, the distance from the mean to the limit is 2, and the tolerance probability is 0.95. The necessary sample size is 49.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Mean with Tolerance Probability** procedure window by expanding **Means**, then **One Mean**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Mean with Tolerance Probability**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Interval Type	Two-Sided
Population Size	Infinite
Confidence Level	0.99
Tolerance Probability	0.95
Distance from Mean to Limit(s)	2
Standard Deviation Source	S from a Previous Sample
S	3
Sample Size of Previous Sample	9

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Confidence Intervals					
Confidence Level	Sample Size (N)	Target Distance from Mean to Limits	Actual Distance from Mean to Limits	Standard Deviation (S)	Tolerance Probability
0.99	49	2.000	1.999	3.000	0.95

PASS also calculated the necessary sample size to be 49.