Chapter 641

Confidence Intervals for One Standard Deviation with Tolerance Probability

Introduction

This procedure calculates the sample size necessary to achieve a specified width (or in the case of one-sided intervals, the distance from the standard deviation to the confidence limit) with a given tolerance probability at a stated confidence level for a confidence interval about a single standard deviation when the underlying data distribution is normal.

Technical Details

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

$$\left[s \left\{ \frac{n-1}{\chi^2_{1-\alpha/2,n-1}} \right\}^{1/2}, s \left\{ \frac{n-1}{\chi^2_{\alpha/2,n-1}} \right\}^{1/2} \right]$$

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

$$s \left\{ \frac{n-1}{\chi^2_{\alpha,n-1}} \right\}^{1/2}$$

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

$$s \left\{ \frac{n-1}{\chi^2_{1-\alpha,n-1}} \right\}^{1/2}$$

For two-sided intervals, the distance from the standard deviation to each of the limits is different. Thus, instead of specifying the distance to the limits we specify the width of the interval, W.

The basic equation for determining sample size for a two-sided interval when W has been specified is

$$W = s \left\{ \frac{n-1}{\chi^2_{\alpha/2,n-1}} \right\}^{1/2} - s \left\{ \frac{n-1}{\chi^2_{1-\alpha/2,n-1}} \right\}^{1/2}$$

For one-sided intervals, the distance from the standard deviation to limits, D, is specified.

The basic equation for determining sample size for a one-sided upper limit when D has been specified is

$$D = s \left\{ \frac{n-1}{\chi^2_{\alpha/2,n-1}} \right\}^{1/2} - s$$

The basic equation for determining sample size for a one-sided lower limit when D has been specified is

$$D = s - s \left\{ \frac{n-1}{\chi^2_{1-\alpha/2,n-1}} \right\}^{1/2}$$

These equations can be solved for any of the unknown quantities in terms of the others.

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 formulas 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 standard deviation with a confidence interval with width no greater than 1.3 units. Suppose further that the sample size needed is calculated to be 105 using the formula above with 3.5 as the estimate for the standard deviation. The sample of size 105 is then obtained from the population, but the standard deviation of the 105 individuals turns out to be 3.9 rather than 3.5. The confidence interval is computed, and the width of the interval is greater than 1.3 units.

This example illustrates the need for an adjustment to adjust the sample size such that the width or distance from the standard deviation 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 for the case of confidence intervals for a mean by Harris, Horvitz, and Mood (1948) and discussed in Zar (1984) and Hahn and Meeker (1991). The adjustment is made by replacing s with

$$s = \sigma \sqrt{F_{1-\gamma;n-1,m-1}}$$

where $1 - \gamma$ is the probability that the width or distance from the standard deviation to the confidence limit 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

$$s = \sigma \left(\frac{\chi^2_{1-\gamma,n-1}}{n-1} \right)^{1/2}$$

where, again, $1 - \gamma$ is the probability that the width or distance from the standard deviation to the confidence limit 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.

For this procedure, both adjustments are adapted from the case of a one-sample confidence interval for a single mean to the case of a one-sample confidence interval for a single standard deviation.

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 standard deviation is $1 - \alpha$.

Example 1 – Calculating Sample Size

A researcher would like to estimate the standard deviation of a population with 95% confidence. It is very important that the interval width is less than 5 grams. Data available from a previous study are used to provide an estimate of the standard deviation. The estimate of the standard deviation is 25.4 grams, from a sample of size 12.

The goal is to determine the sample size necessary to obtain a two-sided confidence interval such that the width of the interval is less than 5 grams. Tolerance probabilities of 0.70 to 0.95 will be examined.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Sample Size**
Interval Type ....**Two-Sided**
Confidence Level (1 - Alpha) **0.95**
Tolerance Probability ....**0.70 to 0.95 by 0.05**
Confidence Interval Width (Two-Sided) **5**
Standard Deviation Source ....**S from a Previous Sample**
S (SD Estimated from a Previous Sample) **25.4**
Sample Size of Previous Sample...................**12**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Interval Type: Two-Sided

Confidence Level	Sample Size N	Confidence Interval Width Target	Confidence Interval Width Actual	Standard Deviation S	Tolerance Probability
0.95	273	5	4.997	25.4	0.70
0.95	293	5	4.998	25.4	0.75
0.95	318	5	4.997	25.4	0.80
0.95	351	5	4.993	25.4	0.85
0.95	398	5	4.995	25.4	0.90
0.95	484	5	4.997	25.4	0.95

Sample size for estimate of S from a previous sample = 12.

Confidence Level	The proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population standard deviation.
N	The size of the sample drawn from the population.
Confidence Interval Width	The distance from the lower limit to the upper limit.
Target Width	The value of the width that is entered into the procedure.
Actual Width	The value of the width that is obtained from the procedure.
S	The estimated standard deviation based on a previous sample.
Tolerance Probability	The probability that a future interval with sample size N and corresponding confidence level will have a width that is less than or equal to the specified width.

Summary Statements

A single-group design will be used to obtain a two-sided 95% confidence interval for a single standard deviation. The standard Chi-square-based formula will be used to calculate the confidence interval. The population standard deviation is estimated to be 25.4 by a previous sample of size 12. To produce a confidence interval with 0.7 probability that the width will be no more than 5, 273 subjects will be needed.

Confidence Intervals for One Standard Deviation with Tolerance Probability

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	273	342	69
20%	293	367	74
20%	318	398	80
20%	351	439	88
20%	398	498	100
20%	484	605	121

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which the confidence interval is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated confidence interval.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. After solving for N, N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 342 subjects should be enrolled to obtain a final sample size of 273 subjects.

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.

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

Plots Section
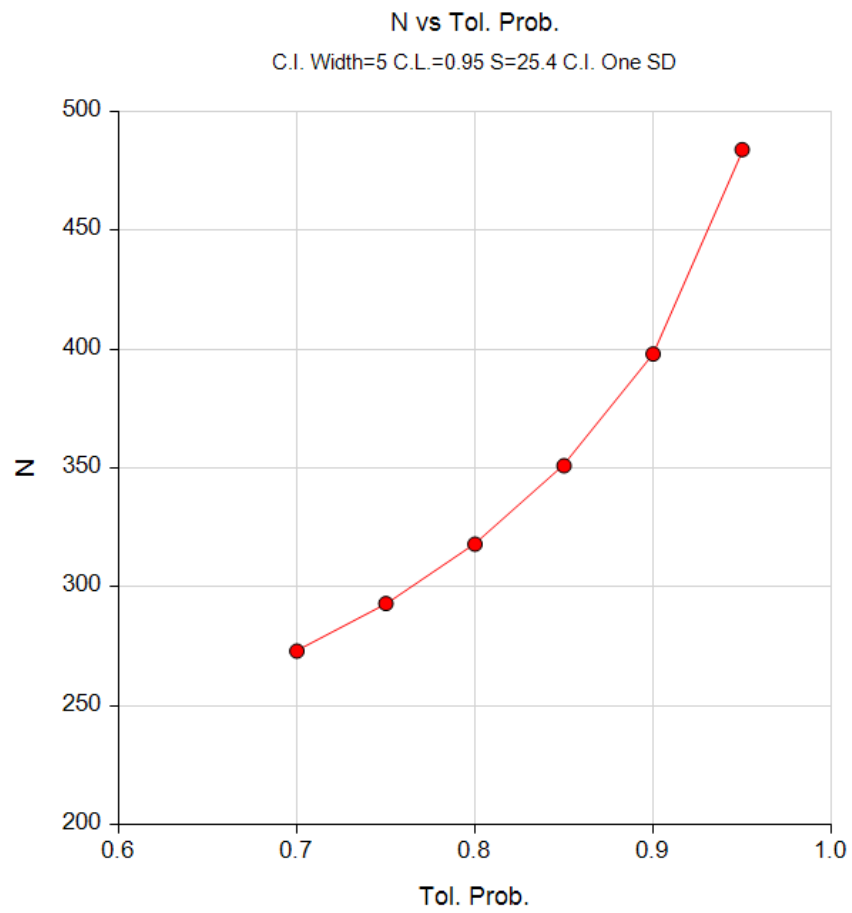
Plots
_____

This plot shows the sample size versus the tolerance probability.

Example 2 – Validation using Simulation

We could not find a published result for confidence intervals for a standard deviation with tolerance probability. This procedure is validated using a simulation. A simulation was run with a confidence level of 95%, sample size of 100, a specified confidence interval width of 6, and assumed population standard deviation of 20.

The number of simulations was 100,000. The proportion of intervals with width less than 6 (tolerance probability) was 0.80168.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Tolerance Probability**
Interval Type...**Two-Sided**
Confidence Level (1 - Alpha) **0.95**
N (Sample Size)..**100**
Confidence Interval Width (Two-Sided) **6**
Standard Deviation Source ....**S is a Population Standard Deviation**
S (Standard Deviation)..................................**20**

Output

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

Numeric Results
───
Solve For: Tolerance Probability
Interval Type: Two-Sided
───

Confidence Level	Sample Size N	Confidence Interval Width Target	Actual	Standard Deviation S	Tolerance Probability
0.95	100	6	6	20	0.802

PASS calculated the tolerance probability to be 0.802, which is well within the simulation error of the simulation estimate of 0.80168.