Chapter 404

Confidence Intervals for a Percentile of a Normal Distribution

Introduction

This procedure calculates the sample size necessary to achieve a specified interval width or distance from a sample percentile to the confidence limit at a stated confidence level when the underlying data distribution is Normal.

Caution: This procedure assumes that the percentile and standard deviation of the future sample will be the same as the percentile and standard deviation that is specified. If the sample percentile or standard deviation is different from those specified when running this procedure, the interval width (or distance to limit) may be narrower or wider than specified.

Technical Details

Chakraborti and Li (2007) compare several methods of confidence interval estimation of a Normal percentile. Their simulation studies showed four of the methods to behave almost identically. One of those methods, which they called the Lawless method (Lawless, 2003, p. 231), is the method used in this procedure. Hahn and Meeker (1991) and Odeh and Owen (1980) provide additional insights into this method.

For a single sample percentile $\hat{Y}_p$ from a Normal distribution with sample standard deviation $\hat{\sigma}$, a two-sided, 100(1 – α)% confidence interval is calculated by

$$\left(\bar{X} - t'_{1-\alpha/2,[n-1,(-\sqrt{n}z_p)]}\frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} - t'_{\alpha/2,[n-1,(-\sqrt{n}z_p)]}\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

with one-sided limits calculated by replacing $\alpha/2$ with α. t' is the non-central t distribution, with corresponding degrees of freedom and noncentrality parameter.

The width of the two-sided confidence interval is calculated as the difference between the upper and lower limits, or

$$W = \bar{X} - t'_{\alpha/2,[n-1,(-\sqrt{n}z_p)]}\frac{\hat{\sigma}}{\sqrt{n}} - \left(\bar{X} - t'_{1-\alpha/2,[n-1,(-\sqrt{n}z_p)]}\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

$$= -t'_{\alpha/2,[n-1,(-\sqrt{n}z_p)]}\frac{\hat{\sigma}}{\sqrt{n}} + t'_{1-\alpha/2,[n-1,(-\sqrt{n}z_p)]}\frac{\hat{\sigma}}{\sqrt{n}}$$

With the sample percentile $\hat{Y}_p$ calculated as

$$\hat{Y}_p = \bar{X} + Z_p\hat{\sigma}$$

the distance from $\hat{Y}_p$ to the lower and upper limits is given as

$$D_{Lower} = \bar{X} + Z_p\hat{\sigma} - \left(\bar{X} - t'_{1-\alpha/2,[n-1,(-\sqrt{n}Z_p)]}\frac{\hat{\sigma}}{\sqrt{n}}\right) = Z_p\hat{\sigma} + t'_{1-\alpha/2,[n-1,(-\sqrt{n}Z_p)]}\frac{\hat{\sigma}}{\sqrt{n}}$$

and

$$D_{Upper} = \bar{X} - t'_{\alpha/2,[n-1,(-\sqrt{n}Z_p)]}\frac{\hat{\sigma}}{\sqrt{n}} - \left(\bar{X} + Z_p\hat{\sigma}\right) = -t'_{\alpha/2,[n-1,(-\sqrt{n}Z_p)]}\frac{\hat{\sigma}}{\sqrt{n}} - Z_p\hat{\sigma}$$

respectively.

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the 10th Percentile such that the width of the interval is no wider than 6 units. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The standard deviation estimate, based on the range of data values, is 22.4. Instead of examining only the interval width of 6, a series of widths from 4 to 8 will also be considered.

The goal is to determine the necessary sample size.

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 0.99**
Confidence Interval Width	**4 to 8 by 0.5**
Percentile Percentage	**10**
Sample Standard Deviation	**22.4**

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		Percentile Percentage	Sample Standard Deviation
		Target	Actual		
0.95	881	4.0	4.000	10	22.4
0.99	1521	4.0	3.999	10	22.4
0.95	697	4.5	4.499	10	22.4
0.99	1203	4.5	4.499	10	22.4
0.95	565	5.0	5.000	10	22.4
0.99	975	5.0	5.000	10	22.4
0.95	468	5.5	5.497	10	22.4
0.99	807	5.5	5.498	10	22.4
0.95	394	6.0	5.995	10	22.4
0.99	679	6.0	5.997	10	22.4
0.95	336	6.5	6.496	10	22.4
0.99	579	6.5	6.499	10	22.4
0.95	290	7.0	6.998	10	22.4
0.99	500	7.0	6.998	10	22.4
0.95	253	7.5	7.499	10	22.4
0.99	436	7.5	7.499	10	22.4
0.95	223	8.0	7.994	10	22.4
0.99	384	8.0	7.997	10	22.4

Confidence Level	The proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true percentile.
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.
Percentile Percentage	The percent of the observations that falls below the percentile value. For example, a value of 70 indicates the 70th percentile is of interest.
Sample Standard Deviation	The anticipated sample standard deviation.

Summary Statements

A single-group design will be used to obtain a two-sided 95% confidence interval for a Normal distribution 10th percentile. The Lawless (2003) method will be used in the confidence interval calculations. The sample standard deviation is assumed to be 22.4. To produce a confidence interval with a width of no more than 4, 881 subjects will be needed.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	881	1102	221
20%	1521	1902	381
20%	697	872	175
20%	1203	1504	301
20%	565	707	142
20%	975	1219	244
20%	468	585	117
20%	807	1009	202
20%	394	493	99
20%	679	849	170
20%	336	420	84
20%	579	724	145
20%	290	363	73
20%	500	625	125
20%	253	317	64
20%	436	545	109
20%	223	279	56
20%	384	480	96

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, 1102 subjects should be enrolled to obtain a final sample size of 881 subjects.

References

Chakraborti, S., and Li, J. 2007. 'Confidence Interval Estimation of a Normal Percentile.' The American Statistician, Volume 61, No. 4, pages 1-6.

Lawless, Jerald F. 2003. Statistical Models and Methods for Lifetime Data, 2nd Edition. John Wiley, New York.

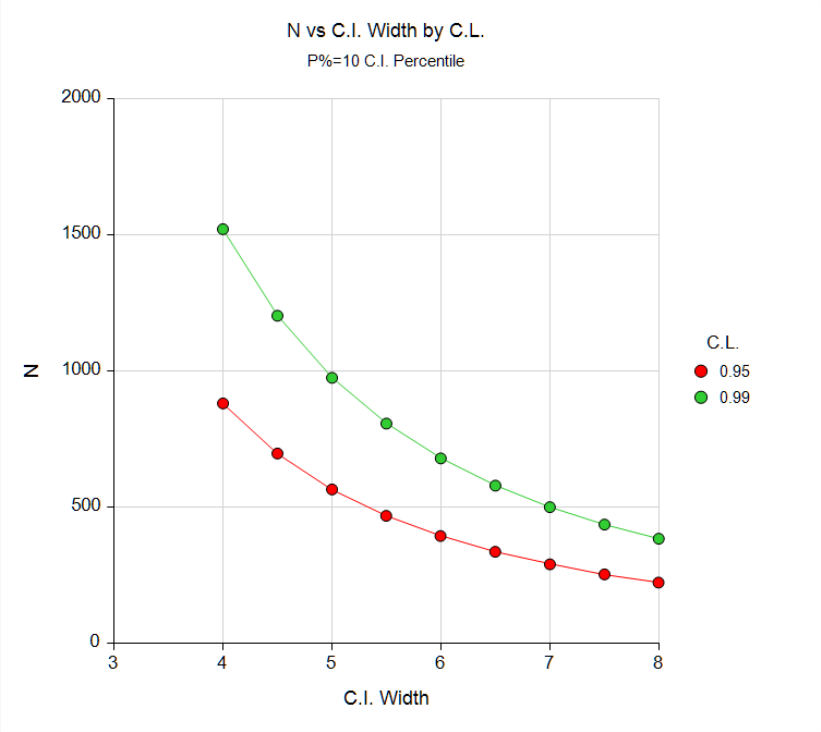
Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.

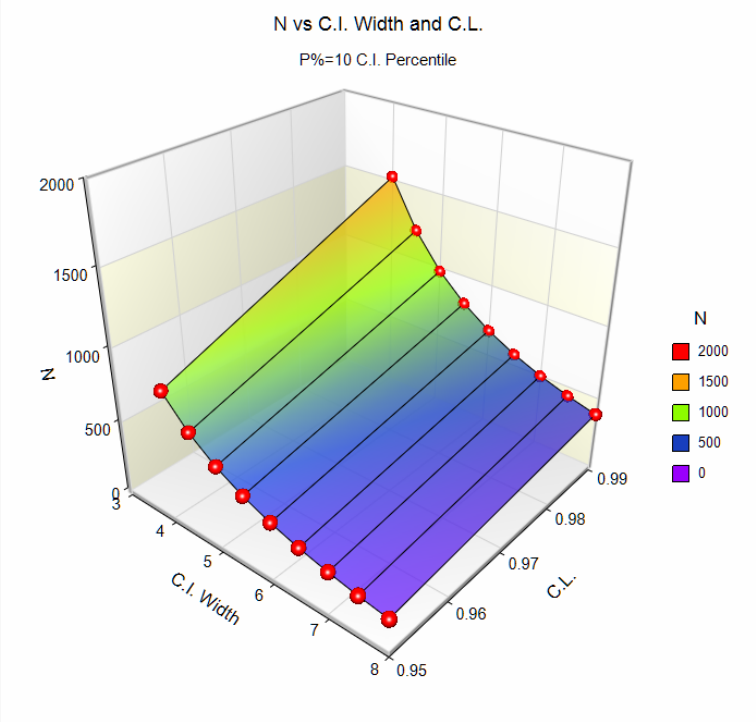
Odeh, R.E. and Owen, D.B. 1980. Tables for Normal Tolerance Limits, Sampling Plans, and Screening. Marcel Dekker, Inc. New York, NY.

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

Plots Section

Plots





These plots show the sample size versus the confidence interval widths for the two confidence levels.

Example 2 – Validation using Hahn and Meeker (1991)

Hahn and Meeker (1991) page 57 give an example of an interval calculation for a confidence interval for the 10^{th} percentile when the confidence coefficient is 95%, the standard deviation is 1.31, and the confidence interval width is 4.95. The sample size given is 5.

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**
Confidence Interval Width	**4.95**
Percentile Percentage	**10**
Sample Standard Deviation	**1.31**

Output

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

Numeric Results

Solve For: Sample Size
Interval Type: Two-Sided

Confidence Level	Sample Size N	Confidence Interval Width		Percentile Percentage	Sample Standard Deviation
		Target	Actual		
0.95	5	4.95	4.949	10	1.31

PASS also calculated the sample size to be 5.