Chapter 581

Confidence Intervals for a Percentile of a Normal Distribution using Expected Width

Introduction

This procedure calculates the sample size necessary to achieve an exact, equal-error-per-tail confidence interval width for a percentile (often called the reference limit) at a stated confidence level when the underlying data distribution is normal. It bases the optimal sample size on the assurance probability.

Technical Details

The methods used in this procedure come from Shieh (2018). This article provides useful insights into why one should use an alternative to the classic Bland-Altman interval.

Confidence Interval

Assume that you are planning a study of a population whose distribution is reasonably close to $N(\mu, \sigma^2)$ to create a confidence interval about the 100*p*th percentile of this population.

Let *X* represent the variable of interest. Suppose that a random sample of *N* observations will be obtained and the usual mean and standard deviation $(\bar{X}, S)$ will be used to estimate the 100*p*th percentile of *X* which is denoted by θ, where

$$\theta = \mu + z_p \sigma$$

Here, z_p is the 100*p*th percentile of the standard normal distribution $N(0,1)$.

An exact two-sided, 100(1 – *α*)% confidence interval is calculated using

$$\left(\bar{X} + t'_{\frac{\alpha}{2},[N-1,(\sqrt{N}z_p)]}\frac{S}{\sqrt{N}}, \bar{X} + t'_{1-\frac{\alpha}{2},[N-1,(\sqrt{N}z_p)]}\frac{S}{\sqrt{N}}\right)$$

with one-sided limits calculated by replacing $\frac{\alpha}{2}$ 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

$$W = \left(t'_{1-\frac{\alpha}{2},[N-1,(\sqrt{N}z_p)]}\frac{S}{\sqrt{N}} - t'_{\frac{\alpha}{2},[N-1,(\sqrt{N}z_p)]}\frac{S}{\sqrt{N}}\right)$$

Sample Size Based on a Bounded Expected Width

The sample size determination method used by this procedure is to determine an *N* that guarantees that the expected width of the confidence interval is less than a boundary value, δ. That is, we select *N* so that $\mathrm{E}(W) \leq \delta$. This leads to the following expression from which the desired *N* can be determined by a simple binary search.

$$\frac{t'_{1-\frac{\alpha}{2},[N-1,(\sqrt{N}Z_p)]} - t'_{\frac{\alpha}{2},[N-1,(\sqrt{N}Z_p)]}}{c\sqrt{N}} \leq \frac{\delta}{\sigma}$$

where

$$c = \frac{\Gamma\left(\frac{N-1}{2}\right)\sqrt{\frac{N-1}{2}}}{\Gamma\left(\frac{N}{2}\right)}$$

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct an exact, two-sided 95% confidence interval for the 95th percentile such that the interval width is guaranteed to be no wider than 1, 2, or 3 units. The standard deviation is between 5 and 7.

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 (N)**
Confidence Level (1 - α)	**0.95**
δ (Upper Bound of Expected C.I. Width)	**1 2 3**
p (Percentile Proportion)	**0.9**
σ (Standard Deviation)	**5 6 7**

Output

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

Numeric Reports

Numeric Results

Solve For:	Sample Size (N)
E(W) Requirement:	E(W) ≤ δ
Interval Type:	Two-Sided Confidence Interval

Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of Expected C.I. Width: Target δT	Upper Bound of Expected C.I. Width: Actual δA	Standard Deviation σ
0.95	703	0.9	1	1.000	5
0.95	1011	0.9	1	1.000	6
0.95	1375	0.9	1	1.000	7
0.95	178	0.9	2	1.998	5
0.95	255	0.9	2	1.999	6
0.95	346	0.9	2	1.999	7
0.95	81	0.9	3	2.991	5
0.95	115	0.9	3	2.997	6
0.95	156	0.9	3	2.992	7

1 - α	The confidence level of the confidence interval of the percentile.
N	The sample size of the study.
p	The percentile proportion. It is the proportion of observations that fall at or below the 100pth percentile value. For example, a value of 0.7 indicates the 70th percentile.
δT	The target upper bound of the expected width of the confidence interval.
δA	The actual upper bound of the expected width of the confidence interval. It may be different from the target value because of the discrete nature of N.
σ	The standard deviation of the population about which the percentile is being constructed.

Summary Statements

A single-group design will be used to obtain a two-sided 95% confidence interval for a Normal distribution 90th percentile. The t-distribution-based formulas, as discussed in Shieh (2018), will be used in the confidence interval calculations. The population standard deviation is assumed to be 5. To produce a confidence interval with an expected confidence interval width that is no more than 1, 703 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%	703	879	176
20%	1011	1264	253
20%	1375	1719	344
20%	178	223	45
20%	255	319	64
20%	346	433	87
20%	81	102	21
20%	115	144	29
20%	156	195	39

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, 879 subjects should be enrolled to obtain a final sample size of 703 subjects.

References

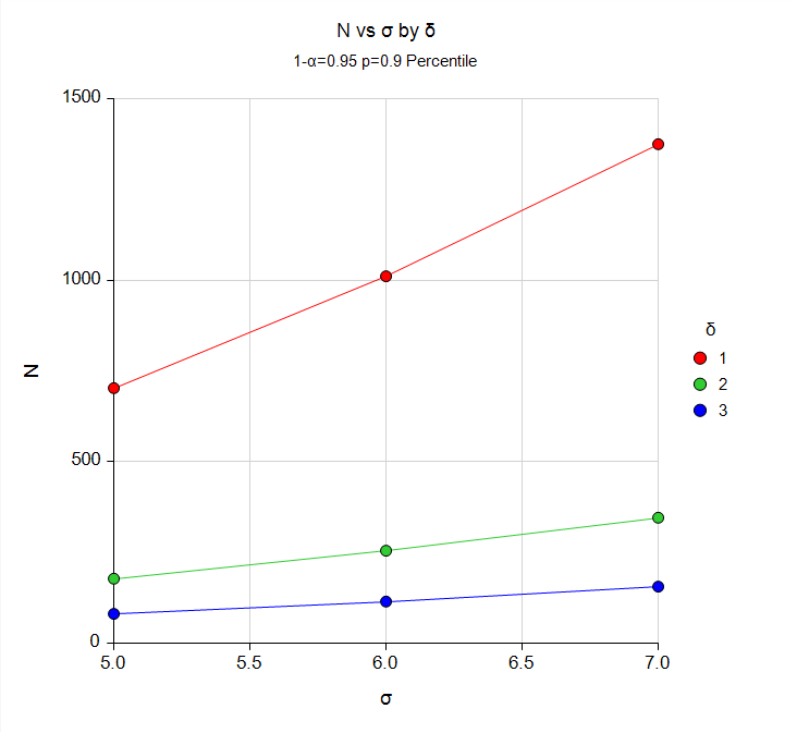
Shieh, Gwowen. 2018. 'The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement.' BMC Medical Research Methodology. 18,45,1. https://doi.org/10.1186/s12874-018-0505-y

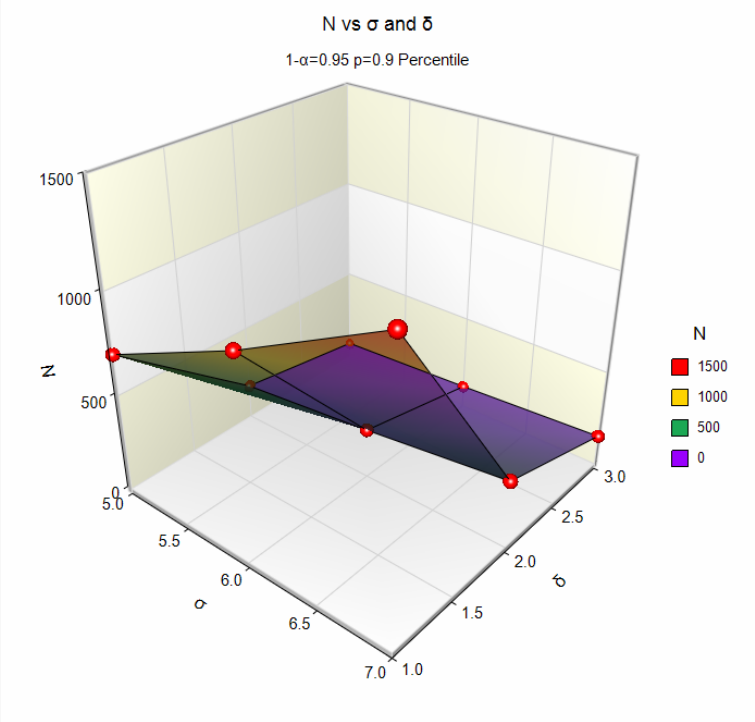
Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.

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

Plots Section

Plots





These plots show the sample sizes required for the various scenarios.

Example 2 – Validation using Shieh (2018)

Shieh (2018) page 7 gives an example of an exact interval calculation for a confidence interval for the 97.5th percentile when the confidence coefficient is 95%, the standard deviation is 19.61, and the expected confidence interval width lower bound is 9.805. The resulting sample size is 183.

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	**Sample Size (N)**
Confidence Level (1 - α)	**0.95**
δ (Upper Bound of Expected C.I. Width)	**9.805**
p (Percentile Proportion)	**0.975**
σ (Standard Deviation)	**19.61**

Output

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

Numeric Results

Solve For:	Sample Size (N)
E(W) Requirement:	E(W) ≤ δ
Interval Type:	Two-Sided Confidence Interval

Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of Expected C.I. Width: Target δT	Upper Bound of Expected C.I. Width: Actual δA	Standard Deviation σ
0.95	183	0.975	9.805	9.799	19.61

The sample size computed by **PASS** is also 183. This validates the procedure.