Chapter 574

Confidence Intervals for a Percentile of a Normal Distribution using Assurance Probability

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 100pth 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 100pth percentile of X which is denoted by θ, where

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

Here, z_p is the 100pth 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 the Assurance Probability

The method of determining a useful sample size used in this procedure is determine an N that guarantees with specified assurance probability no less than $(1 - \gamma)$ that the actual width of the confidence interval no greater than a boundary value, ω. That is, we select N so that $\Pr(W \leq \omega) \geq 1 - \gamma$. This leads to the following expression from which the desired N can be determined by a simple search.

$$k = \frac{N(N-1)\left(\frac{\omega}{\sigma}\right)^2}{\left\{t'_{1-\frac{\alpha}{2},[N-1,(\sqrt{N}Z_p)]} - t'_{\frac{\alpha}{2},[N-1,(\sqrt{N}Z_p)]}\right\}^2}$$

Now, we find the smallest N for which

$$X^2_{N-1}(k) \geq 1 - \gamma$$

where $X^2_{N-1}(k)$ is the cumulative chi-square distribution with degrees of freedom N – 1.

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 90th percentile such that the interval width is guaranteed to be no wider than 6 units with probability 0.90, 0.95, or 0.99. The standard deviation is between 21 and 23.

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**
1 - γ (Lower Bound of Assurance Probability).........**0.9 0.95 0.99**
ω (Upper Bound of the C.I. Width)..........................**6**
p (Percentile Proportion)...**0.9**
σ (Standard Deviation)...**21 22 23**

Output

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

Numeric Reports

Numeric Results
———
Solve For: Sample Size (N)
Assurance Requirement: Pr(W ≤ ω) ≥ 1 - γ
Interval Type: Two-Sided Confidence Interval
———

Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of the Confidence Interval Width ω	Lower Bound of Assurance Probability		Standard Deviation σ
				Target 1 - γт	Actual 1 - γᴀ	
0.95	379	0.9	6	0.90	0.9038	21
0.95	414	0.9	6	0.90	0.9035	22
0.95	451	0.9	6	0.90	0.9059	23
0.95	388	0.9	6	0.95	0.9515	21
0.95	424	0.9	6	0.95	0.9534	22
0.95	461	0.9	6	0.95	0.9531	23
0.95	405	0.9	6	0.99	0.9903	21
0.95	442	0.9	6	0.99	0.9909	22
0.95	479	0.9	6	0.99	0.9901	23

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.
ω   An upper bound on the confidence interval width. The sample size guarantees (assures) that 100 (1 - γ)% of interval
        widths will be less than this value.
1 - γT A lower bound on the target (planned) assurance probability. The assurance probability is the probability that the
        confidence interval width will be less than ω.
1 - γA A lower bound on the actual assurance probability achieved by this sample size. 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 21. To produce a confidence interval with 0.9
probability that the confidence interval width will be no more than 6, 379 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%	379	474	95
20%	414	518	104
20%	451	564	113
20%	388	485	97
20%	424	530	106
20%	461	577	116
20%	405	507	102
20%	442	553	111
20%	479	599	120

───

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, 474 subjects should be enrolled to obtain a final sample size of 379 subjects.
───

References
───
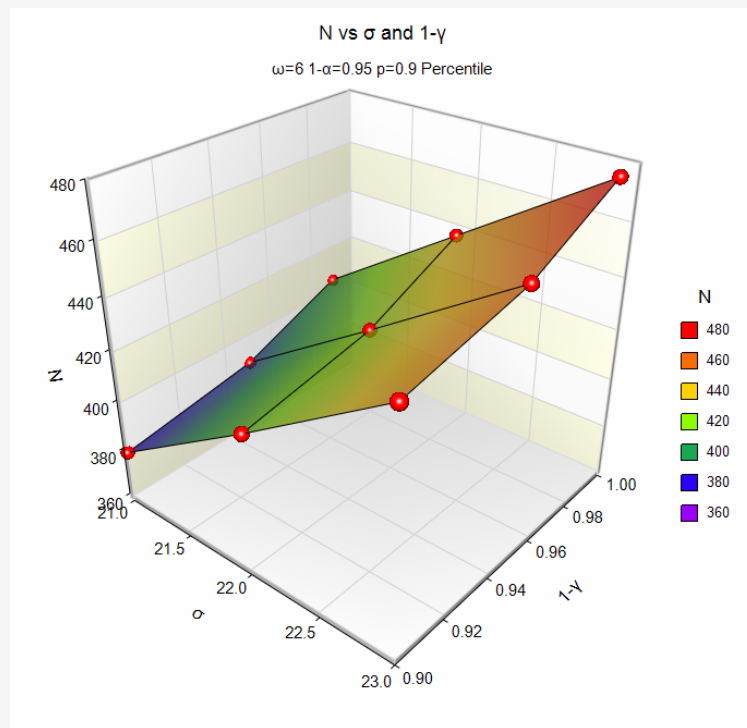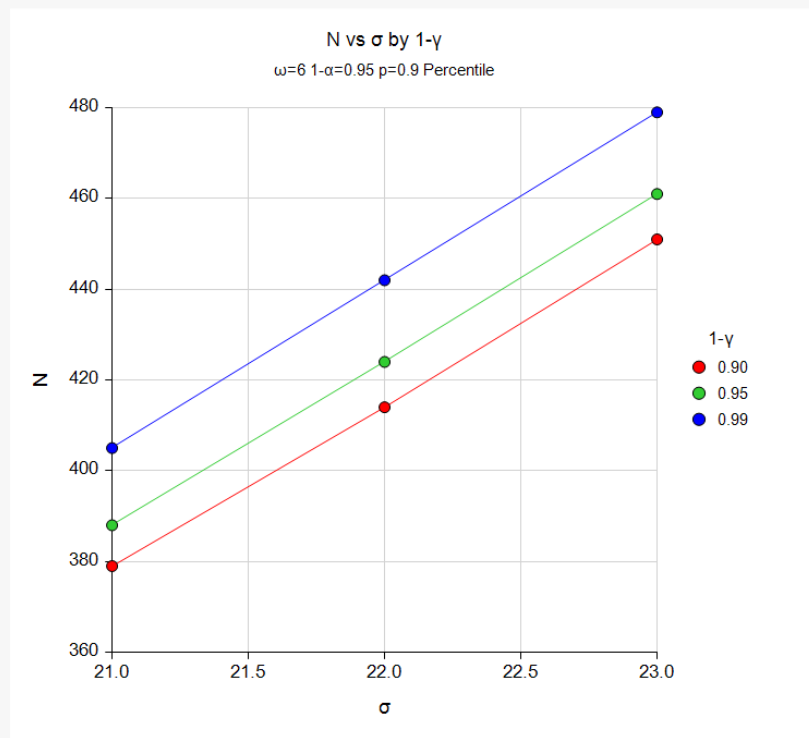
Shieh, Gwowen. 2018. 'The appropriateness of Bland-Altman's approximate confidence intervals for limits of
    agreement.' BMC Medical Research Methodology. 18,45,1. https://doi.org/10.1186/s12874-018-0505-y
Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.
───

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, the assurance level is 0.90, and the confidence interval width boundary is 9.805. The sample size is 207.

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**
1 - γ (Lower Bound of Assurance Probability).........**0.9**
ω (Upper Bound of the 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)
Assurance Requirement: Pr(W ≤ ω) ≥ 1 - γ
Interval Type: Two-Sided Confidence Interval

Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of the Confidence Interval Width ω	Lower Bound of Assurance Probability		Standard Deviation σ
				Target 1 - γ$_T$	Actual 1 - γ$_A$	
0.95	207	0.975	9.805	0.9	0.9065	19.61

The sample size computed by **PASS** is also 207. This validates the procedure.