Chapter 579

Confidence Intervals for Regression-Based Reference Limits using Expected Relative Precision

Introduction

This procedure calculates the sample size necessary to achieve a stated expected relative precision in exact confidence intervals of regression-based reference limits (percentiles). Using the ratio of the widths of the confidence interval and the reference interval as a relative precision index, optimal sample size can be found using a specific tolerance interval. The procedure is based on the work of Shieh (2017) which shows that the proposed exact confidence intervals are much more precise than the approximate procedures that are commonly used.

Technical Details

These methods used in this procedure come from Shieh (2017). This article provides useful insights into why one should use the new procedure proposed by the author.

Confidence Interval Estimation

Assume that you are planning a study that will generate data commonly analyzed by multiple linear regression model consisting of a response variable *Y* and *K* covariate (independent) variables $X_1, X_2, \ldots, X_K$. These variables are measured on *N* subjects. The regression model is of the form

$$Y_i = \beta_0 + \sum_{k=1}^{K} X_{ik}\beta_k + e_i$$

Where Y_i is an observed, continuous response variable on the i^{th} subject; $X_{i1}, \ldots, X_{iK}$ are observed, continuous covariates; $\beta_0, \beta_1, \ldots, \beta_K$ are unknown regression parameters to be estimated; and e_i are $iid\ N(0, \sigma^2)$ random errors for $i = 1, \ldots, N$.

Suppose the mean response at a particular value of X, say X_0, $\mu = \mathrm{E}[Y|\{X_{01}, \ldots, X_{0K}\}] = \beta_0 + \sum_{k=1}^{K} X_{0k}\beta_k$. The estimated value of μ is

$$\hat{\mu} = b_0 + \sum_{k=1}^{K} X_{0k}b_k$$

where the *b's* are the regression estimates of the coefficients.

Suppose you want to estimate the 100pth percentile (reference limit) which is denoted by θ_p, with

$$\theta_p = \mu + z_p \sigma.$$

The minimum variance unbiased estimator is

$$\hat{\theta}_p = \hat{\mu} + c z_p \hat{\sigma}$$

where

$$c = \frac{\sqrt{v/2}\,\Gamma(v/2)}{\Gamma((v+1)/2)}, \quad v = N - K - 1$$

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

The calculation of the exact two-sided, 100(1 – α)% confidence interval for the reference limit $\left(\hat{\theta}_{pL}, \hat{\theta}_{pU}\right)$ is done as follows.

$$X_i = (X_{i1}, \dots, X_{iK})'$$

$$X_0 = (X_{01}, \dots, X_{0K})'$$

$$\bar{X} = \sum_{i=1}^{N} X_i / N$$

$$A = \sum_{i=1}^{N} (X_i - \bar{X})\,(X_i - \bar{X})'$$

$$W = \frac{1}{N} + \sum_{i=1}^{N} (X_0 - \bar{X})' A^{-1}\,(X_0 - \bar{X})$$

$$\hat{\sigma}^2 = \frac{SSE}{v}$$

$$\hat{\theta}_{pL} = \hat{\mu} + \sqrt{W}\, t'_{\frac{\alpha}{2}, v, \frac{z_p}{\sqrt{W}}}\, \hat{\sigma}$$

$$\hat{\theta}_{pU} = \hat{\mu} + \sqrt{W}\, t'_{1-\frac{\alpha}{2}, v, \frac{z_p}{\sqrt{W}}}\, \hat{\sigma}$$

with SSE = error sum of squares from the multiple regression and t' is the percentile of the non-central t distribution with corresponding degrees of freedom and noncentrality parameter. Note that t' is not symmetric so that the confidence interval is not symmetric either.

Sample Size

The sample size calculation is based on the relative precision, *Q*, which is given by

$$Q = C/R$$

where $C = \hat{\theta}_{pU} - \hat{\theta}_{pL}$ and $R = 2z_p c\hat{\sigma}$. Note that $\hat{\mu}$ and $\hat{\sigma}$ cancel out of the Q ratio.

The sample size is than constructed so that

$$\mathrm{E}(Q) \leq \delta.$$

Shieh (2017) provides a numerical integration algorithm for calculating this expected value, so the sample size can be determined by a simple search.

Example 1 – Finding Sample Size

Suppose a study is planned in which the researcher wishes to construct an exact two-sided 95% confidence interval for the 97.5th percentile such that the expected value of the relative precision, E(Q), is guaranteed to be no larger than 0.1, 0.2, or 0.3. The values of λ are 1, 1.2, 1.4 when K = 1.

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**
K (Number of Covariates) **1**
δ (Upper Bound of E(Q))............................... **0.1 0.2 0.3**
p (Percentile Proportion)............................... **0.975**
λ (Effect Size|Xo) **1 1.2 1.4**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size (N)
E(Q) Requirement: E(Q) ≤ δ
Interval Type: Two-Sided Confidence Interval
Number of Covariates: 1

Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of E(Q) Target δT	Upper Bound of E(Q) Actual δA	Effect Size|Xo λ	Confidence Interval Width Factor CIW
0.95	397	0.975	0.1	0.100	1.0	0.392
0.95	441	0.975	0.1	0.100	1.2	0.392
0.95	493	0.975	0.1	0.100	1.4	0.392
0.95	103	0.975	0.2	0.199	1.0	0.783
0.95	114	0.975	0.2	0.199	1.2	0.783
0.95	127	0.975	0.2	0.199	1.4	0.783
0.95	48	0.975	0.3	0.299	1.0	1.180
0.95	53	0.975	0.3	0.299	1.2	1.179
0.95	59	0.975	0.3	0.299	1.4	1.176

Q The relative precision. Q = C / R where C is the width of the exact confidence interval of the reference limit (percentile) and R is the width of the reference range.
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.
δ_T The target upper bound of the expected relative precision, E(Q).
δ_A The actual upper bound of the expected relative precision, E(Q).
λ The effect size given the particular value of Xo. If K = 1, $\lambda = |\mu_x - Xo|/\sigma_x$.
CIW A factor that when multiplied by the standard deviation of the residuals (σ, not σ_x) gives the width of the confidence interval of the reference limit using δ as the value of Q.

Summary Statements

A regression design will be used to obtain a two-sided 95% confidence interval for the 97.5th percentile (regression-based reference limit). The methodology described in Shieh (2017) will be used in the (exact) confidence interval calculations. The number of covariates will be 1. The effect size at the covariate value of interest is assumed to be 1. To produce a confidence interval with an expected relative precision that is no more than 0.1, 397 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%	397	497	100
20%	441	552	111
20%	493	617	124
20%	103	129	26
20%	114	143	29
20%	127	159	32
20%	48	60	12
20%	53	67	14
20%	59	74	15

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, 497 subjects should be enrolled to obtain a final sample size of 397 subjects.

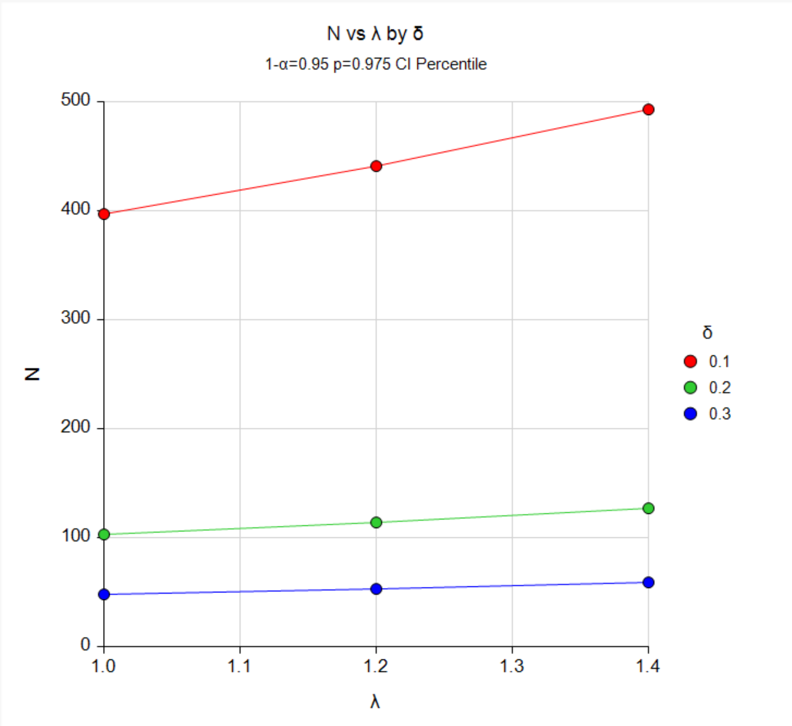
References

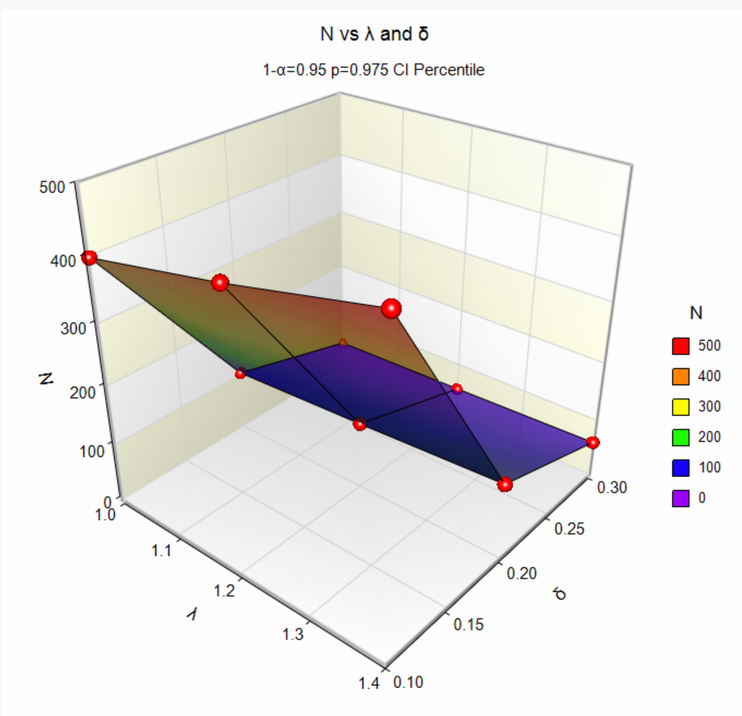
Shieh, Gwowen. 2018. 'The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement.' BMC Medical Research Methodology. 18,45,1. https://doi.org/10.1186/s12874-018-0505-y
Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.

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 (2017)

Shieh (2017) page 196 gives several examples of calculating the necessary sample size for of an exact confidence interval for the 97.5th percentile when the confidence coefficient is 95%, λ is 1, and δ is 0.1. The number of covariates is 3. The sample size is 399.

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**
K (Number of Covariates) **3**
δ (Upper Bound of E(Q))...............................**0.1**
p (Percentile Proportion)...............................**0.975**
λ (Effect Size|Xo) **1**

Output

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

Numeric Results

Solve For: Sample Size (N)
E(Q) Requirement: E(Q) ≤ δ
Interval Type: Two-Sided Confidence Interval
Number of Covariates: 3

Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of E(Q) Target δT	Upper Bound of E(Q) Actual δA	Effect Size|Xo λ	Confidence Interval Width Factor CIW
0.95	399	0.975	0.1	0.1	1	0.392

The sample size computed by **PASS** is also 399. This validates the procedure.