Chapter 427

Bland-Altman Method for Assessing Agreement in Method Comparison Studies

Introduction

This procedure provides sample size and power calculations for a method comparison study that is analyzed with the Bland – Altman (1986) method. The procedure is based on the work of Lu *et al.* (2016).

A method comparison study is used to compare two measurement methods. The data consists of pairs of measurements taken on *N* subjects. The Bland-Altman technique forms two limits of agreement (LoA) from the *N* paired differences. Next, a confidence interval is constructed for each of the two limits. To complete the analysis, the minimum and maximum values of these two confidence intervals are compared to an allowable range which is defined as $-\delta$ to δ. If the minimum and maximum values are both between the allowable range, the two measurements are said to be in agreement. Otherwise, they are not.

Technical Details

Suppose two measurements (X and Y) are obtained using two measurement methods on each of *N* subjects drawn from a population of interest. It is assumed that the differences computed on these data pairs are normally distributed with constant variance and no noticeable trends or patterns in the *N* sums of these two measurements. The differences of these measurements are formed and analyzed as follows. For each subject, form the difference

$$d_i = x_i - y_i.$$

The 100 (1 - γ)% LoAs can be calculated for the usual mean and standard deviation of the differences

$$\overline{D} \pm z_{1-\gamma/2} SD$$

where *z* is the cumulative percentile of a standard normal distribution.

The lower and upper confidence limits of these LoA limits are

$$Lower = \overline{D} - z_{1-\frac{\gamma}{2}} SD - t_{1-\frac{\alpha}{2}, N-1} SD \sqrt{\frac{1}{N} + \frac{z_{1-\gamma/2}^2}{2(N-1)}}$$

$$Upper = \overline{D} + z_{1-\frac{\gamma}{2}} SD + t_{1-\frac{\alpha}{2}, N-1} SD \sqrt{\frac{1}{N} + \frac{z_{1-\gamma/2}^2}{2(N-1)}}$$

Finally, a maximum allowable difference value δ is defined. If $-\delta$ < lower < upper < δ, the two measurements are said to be in agreement. Otherwise, they are not.

Power Analysis

The power of this testing procedure, given by Lu *et al.* (2016), was shown to be

$$power = \text{NCT}\left(t_{1-\frac{\alpha}{2},N-1}, N-1, \lambda_1\right) + \text{NCT}\left(t_{1-\frac{\alpha}{2},N-1}, N-1, \lambda_2\right)$$

where NCT( ) is the cumulative distribution function of a Student's non-central *t* distribution with non-centrality parameter λ.

The values of the non-centrality parameters are

$$\lambda_1 = \frac{\delta - \overline{D} - z_{1-\frac{\gamma}{2}} SD}{SD\sqrt{\frac{1}{N} + \frac{z_{1-\gamma/2}^2}{2(N-1)}}}$$

$$\lambda_2 = \frac{\delta + \overline{D} - z_{1-\frac{\gamma}{2}} SD}{SD\sqrt{\frac{1}{N} + \frac{z_{1-\gamma/2}^2}{2(N-1)}}}$$

If the sample size is required, this power formula can be used in a binary search for *N*.

Example 1 – Finding the Sample Size

Clinicians wish to conduct a method comparison study and analyze it with the Bland-Altman method. They want to estimate the number of subjects that must be measured if the confidence levels are both 0.95, the power is 0.8 or 0.9, δ is 7, D is anticipated to be 0.5, and the SD is expected to be between 2.5 and 2.7.

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 .... **N (Number of Measurement Pairs)**
Power.. **0.8 0.9**
Confidence Level of LoAs **0.95**
Confidence Level of CIs of LoAs................... **0.95**
δ (Max Allowable Difference)......................... **7**
D (Mean of Differences)................................ **0.5**
SD (Std Dev of Differences).......................... **2.5 2.6 2.7**

Output

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

Numeric Reports

Numeric Results

Solve For: N (Number of Measurement Pairs)
Hypotheses: H0: Measurement methods do not agree. H1: Measurement methods are in agreement.
Conclude measurement agreement if -δ < Lower LoA - Lower CI < Upper LoA + Upper CI < δ. Otherwise, don't conclude agreement.

		Confidence Level			Sample Differences	
Power	Number of Measurement Pairs N	Limits of Agreement CL LoA	Confidence Intervals of Limits of Agreement CL CI	Maximum Allowable Difference δ	Mean D	Standard Deviation SD
0.8059	60	0.95	0.95	7	0.5	2.5
0.8019	82	0.95	0.95	7	0.5	2.6
0.8024	118	0.95	0.95	7	0.5	2.7
0.9014	78	0.95	0.95	7	0.5	2.5
0.9003	108	0.95	0.95	7	0.5	2.6
0.9002	156	0.95	0.95	7	0.5	2.7

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The number of subjects. A pair of measurements is made on each subject.
CL LoA The confidence level of the limits of agreement.
CL CI The confidence level of the confidence intervals constructed for each of the limits of agreement.
δ The maximum allowable difference. The boundaries between which the data limits must fall are -δ and δ.
D The anticipated mean of the sample differences.
SD The anticipated standard deviation of the sample differences.

Summary Statements

A paired (two-measurement) design will be used to test agreement of two measurement methods (e.g., assays) using the Bland-Altman range of agreement methods. Each subject will be measure twice, once with each measurement method. Paired measurements are averaged and differenced to produce a Bland-Altman plot. The test will be performed by comparing the maximum allowable difference to the upper confidence limit of the upper limit of agreement and the negative of the maximum allowable difference to the lower confidence limit of the lower limit of agreement. If the maximum maximum allowable difference is greater than the upper confidence limit of the upper limit of agreement, and the negative of the maximum maximum allowable difference is lower than the lower confidence limit of the lower limit of agreement, the null hypothesis is rejected and the conclusion of method agreement is reached. The confidence level for the limits of agreement is assumed to be 0.95. The confidence level for the confidence intervals of the limits of agreement is assumed to be 0.95. The desired maximum allowable difference is 7. The mean of the sample differences is anticipated to be 0.5 and standard deviation of the sample differences is anticipated to be 2.5. With these parameters, to obtain a power (to detect agreement) of 80%, 60 subject pairs will be needed.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	60	75	15
20%	82	103	21
20%	118	148	30
20%	78	98	20
20%	108	135	27
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 power is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
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, 75 subjects should be enrolled to obtain a final sample size of 60 subjects.

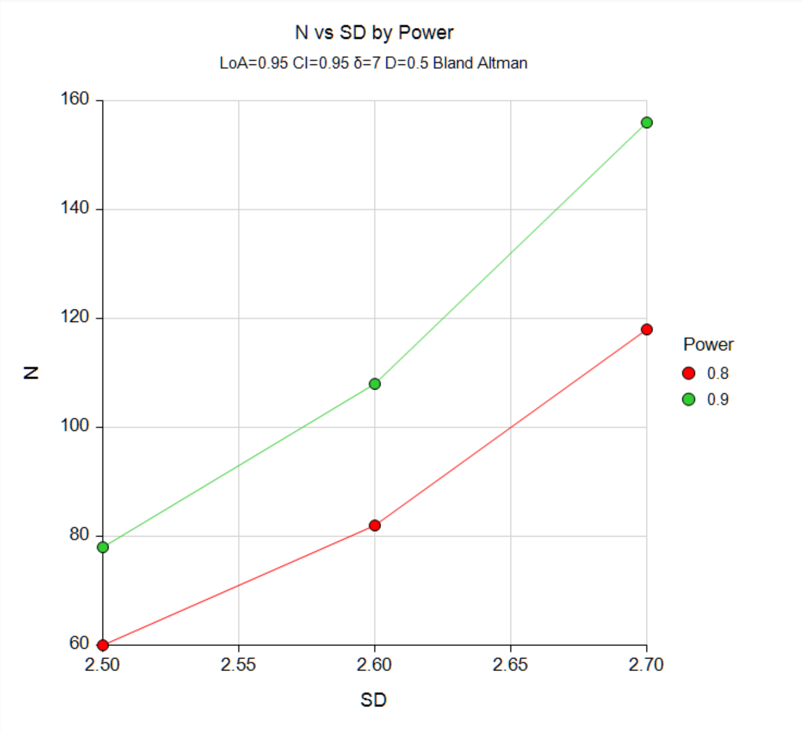
References

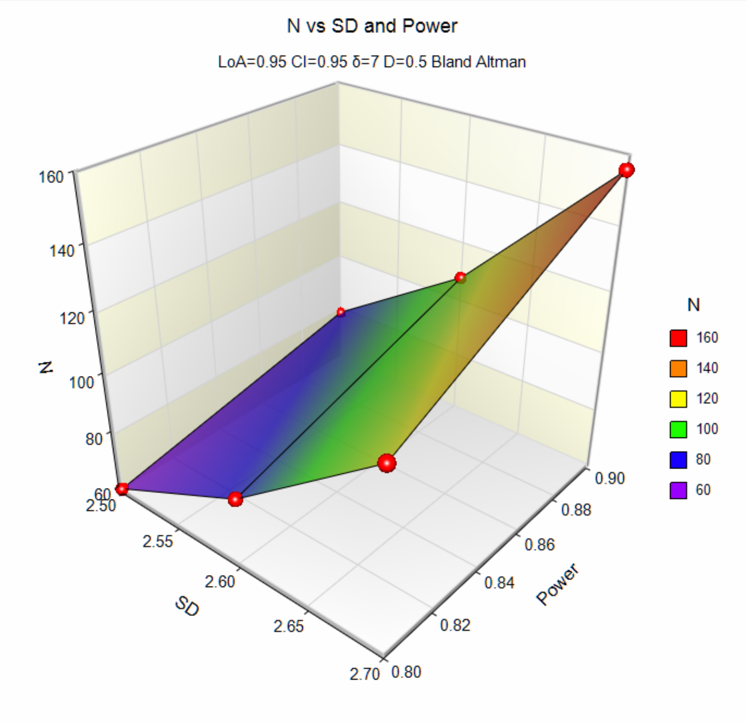
Lu, M.J.,Zhong, W.H., Liu, Y.X., Miao, H.Z., Li, Y.C., Ji, M.H. 2016. 'Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method.' The International Journal of Biostatistics. Article 20150039. (Published online).
Bland, J.M., Altman, D.G. 1986. 'Statistical methods for assessing agreement between two methods of clinical measurement.' The Lancet i:307-310.

These reports show the values of each of the parameters, one scenario per row.

Plots Section

Plots





These plots show the relationship between the standard deviation and sample size for the two alpha levels.

Example 2 – Validation using Lu et al. (2016)

Lu *et al.* (2016) provide several examples in their Table 1. We will validate the program by considering one of the table entries. Set the confidence levels both to 0.95, the power to 0.8, δ is 2.5, D to 0.2, and SD to 1. They report a sample size of 203.

We will compute the power for a range of N's from 200 to 203.

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 .... **Power**
Confidence Level of LoAs............................. **0.95**
Confidence Level of Cis of LoAs................... **0.95**
N (Number of Measurement Pairs)................ **200 201 202 203**
δ (Max Allowable Difference)......................... **2.5**
D (Mean of Differences)................................ **0.2**
SD (Std Dev of Differences)........................... **1**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: Measurement methods do not agree. H1: Measurement methods are in agreement.
Conclude measurement agreement if -δ < Lower LoA - Lower CI < Upper LoA + Upper CI < δ. Otherwise, don't conclude agreement.

		Confidence Level			Sample Differences	
Power	Number of Measurement Pairs N	Limits of Agreement CL LoA	Confidence Intervals of Limits of Agreement CL CI	Maximum Allowable Difference δ	Mean D	Standard Deviation SD
0.7983	200	0.95	0.95	2.5	0.2	1
0.8003	201	0.95	0.95	2.5	0.2	1
0.8023	202	0.95	0.95	2.5	0.2	1
0.8042	203	0.95	0.95	2.5	0.2	1

PASS also calculates the power to be 0.8 (within rounding).

This example shows a problem with Lu *et al.* (2016). That is, their answer of N = 203 has a power of 0.8 within rounding. However, lower values of N (201 and 202) also have powers over 0.8. We traced the problem to their equation (6) which is an approximation. In **PASS**, we use their formula (5) which gives the exact answer. Because of this problem, we found several entries in Table 1 that were slightly off.