

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 $(1 - \gamma)\%$ LoAs can be calculated for the usual mean and standard deviation of the differences

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

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

The lower and upper confidence limits of these LoA limits are

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

Bland-Altman Method for Assessing Agreement in Method Comparison Studies

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

Finally, a maximum allowable difference value δ is defined. If $-\delta < \text{lower} < \text{upper} < \delta$, 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

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

where $NCT(\cdot)$ is the inverse of a Student's non-central t distribution with non-centrality parameter λ .

The values of the non-centrality parameters are

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

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

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

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power* and *Number of Pairs*. The 'Solve For' parameter will be displayed on the vertical axis of any plots that are generated.

Power and Confidence Levels

Power

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1 - Beta, so specifying power implicitly specifies beta. Beta is the probability obtaining a false negative with the statistical test. That is, it is the probability of accepting a false null hypothesis.

Bland-Altman Method for Assessing Agreement in Method Comparison Studies

The valid range is 0 to 1. Different disciplines have different standards for setting power. The most common choice is 0.80, but 0.90 and 0.95 are also popular.

You can enter a single value, such as 0.90, or a series of values, such as 0.70 0.80 0.90, or 0.70 to 0.90 by 0.05.

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

Confidence Level of LoAs

Enter one or more values for the confidence level of the limits of agreement (LoAs). This is the proportion of differences that are expected to be between the limits.

The absolute range is between 0 and 1. The recommended value of 0.95 is almost always used for method comparison studies.

You can enter a single value such as 0.95 or a series of values such as 0.9 0.95 0.99 or 0.9 to 0.99 by 0.01.

Confidence Level CIs of LoAs

Enter one or more values for the confidence level of the confidence intervals for lower and upper LoAs. The confidence level is the proportion of confidence intervals (constructed in like manner) that would contain the true population percentile.

The absolute range is between 0 and 1. The recommended value of 0.95 is almost always used for method comparison studies.

You can enter a single value such as 0.95 or a series of values such as 0.9 0.95 0.99 or 0.9 to 0.99 by 0.01.

Effect Size

δ (Max Allowable Difference)

Enter one or more values of δ : the maximum allowable mean difference between the two measurement methods. This is used to define the boundaries between which the limits computed from the data must fall. The boundaries are $-\delta$ and δ .

Let

A = Lower LoA - Lower CL of Lower LoA CI

B = Upper LoA + Upper CL of Upper LoA CI

If $-\delta < A < B < \delta$, the measurements are considered as in agreement. Otherwise, they are not.

The range is $0 < \delta$.

Enter a single value such as 1.0, a series such as 1 2 3, or a range such as 1 to 5 by 1.

If δ is too small (i.e. less than $D + 2 SD$), no solution will be available.

D (Mean of Differences)

Enter one or more values for D, the anticipated mean of the sample differences.

The value of D is unrestricted.

Enter a single value such as 1.0, a series such as 1 2 3, or a range such as 1 to 5 by 1.

SD (Std Dev of Differences)

Enter one or more values for SD, the anticipated standard deviation of the sample differences.

The value of SD must be greater than zero.

Enter a single value such as 1.0, a series such as 1 2 3, or a range such as 1 to 5 by 1.

Press the small ' σ ' button to the right to obtain options for estimating the standard deviation.

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Bland-Altman Method for Assessing Agreement in Method Comparison Studies** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

| <u>Option</u> | <u>Value</u> |
|--|--|
| 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 |

Annotated Output

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

Numeric Results

| Numeric Results | | | | | | | |
|-----------------|-----------------------------|-------------------|-------------------------|------------------------------|---------------------|------------------------|--|
| | Number of Measurement Pairs | Conf Level of LoA | Conf Level of CI of LoA | Maximum Allowable Difference | Mean of Differences | Std Dev of Differences | |
| Power | N | CL LoA | CL CI | δ | D | SD | |
| 0.8059 | 60 | 0.950 | 0.950 | 7.000 | 0.500 | 2.500 | |
| 0.8019 | 82 | 0.950 | 0.950 | 7.000 | 0.500 | 2.600 | |
| 0.8024 | 118 | 0.950 | 0.950 | 7.000 | 0.500 | 2.700 | |
| 0.9014 | 78 | 0.950 | 0.950 | 7.000 | 0.500 | 2.500 | |
| 0.9003 | 108 | 0.950 | 0.950 | 7.000 | 0.500 | 2.600 | |
| 0.9002 | 156 | 0.950 | 0.950 | 7.000 | 0.500 | 2.700 | |

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.

Bland-Altman Method for Assessing Agreement in Method Comparison Studies

Report Definitions

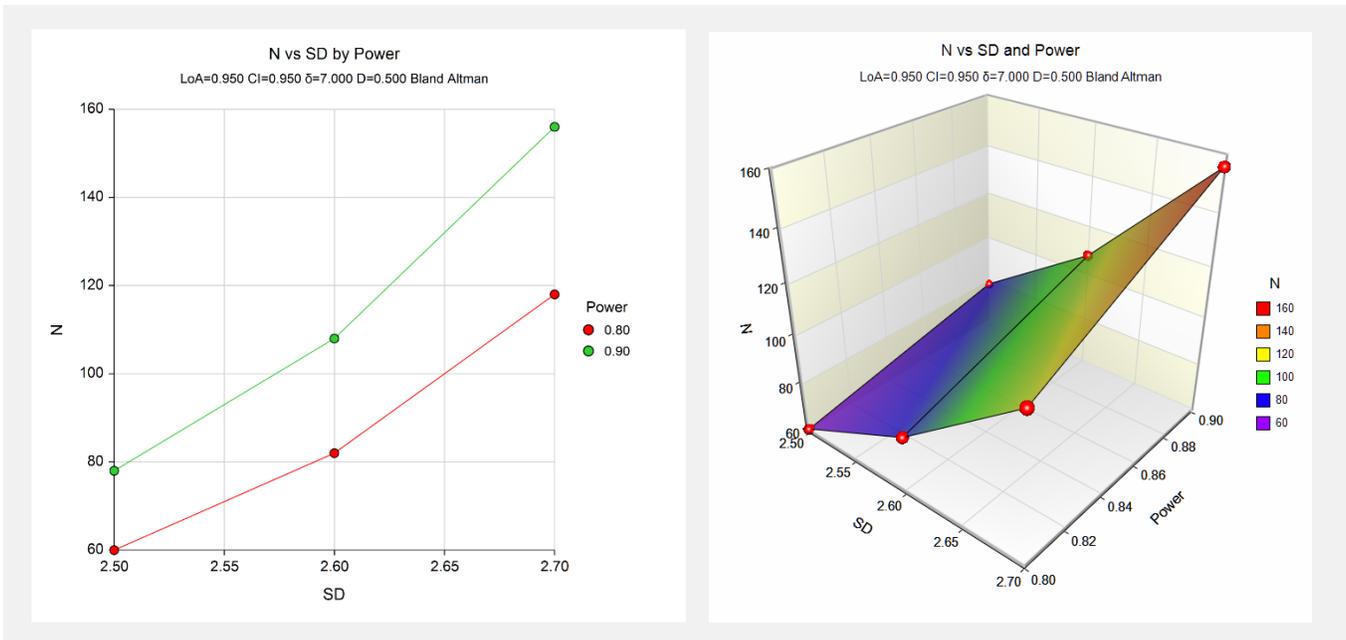
Hypotheses: H0: Measurement methods do not agree. H1: Measurement methods are in agreement.
 Conclude measurement agreement if $-\delta < \text{Lower LoA} - \text{Lower CI} < \text{Upper LoA} + \text{Upper CI} < \delta$. Otherwise, don't conclude agreement.
 Power is the probability of rejecting a false null hypothesis.
 N is the number of subjects. A pair of measurements is made on each subject.
 CL LoA is the confidence level of the limits of agreement.
 CL CI is the confidence level of the confidence intervals constructed for each of the limits of agreement.
 δ is the maximum allowable difference. The boundaries between which the data limits must fall are $-\delta$ and δ .
 D is the anticipated mean of the sample differences.
 SD is the anticipated standard deviation of the sample differences.

Summary Statements

A method comparison study will be analyzed with a Bland-Altman plot. This study will produce limits of agreement (LoA) and associated confidence intervals from the two measurements obtained on each subject. The conclusion that the two measurements are in agreement depends whether the confidence intervals constructed from the data are within the boundaries set from the maximum allowable difference. A sample of 60 subjects achieves 81% power to detect agreement when the confidence level of the LoA is 0.950 and the confidence level of the confidence intervals about the LoAs is 0.950. The maximum allowable difference is 7.000. The mean and standard deviation of the sample differences are anticipated to be 0.500 and 2.500.

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

Plots Section



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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Bland-Altman Method for Assessing Agreement in Method Comparison Studies** procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

| <u>Option</u> | <u>Value</u> |
|--|------------------------|
| 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

| Numeric Results | | | | | | |
|-----------------|-----------------------------------|----------------------|----------------------------|------------------------------------|------------------------|---------------------------|
| | Number of Measurement Pairs | Conf Level of LoA | Conf Level of CI of LoA | Maximum Allowable Difference | Mean of Differences | Std Dev of Differences |
| Power | N | CL LoA | CL CI | δ | D | SD |
| 0.7983 | 200 | 0.950 | 0.950 | 2.500 | 0.200 | 1.000 |
| 0.8003 | 201 | 0.950 | 0.950 | 2.500 | 0.200 | 1.000 |
| 0.8023 | 202 | 0.950 | 0.950 | 2.500 | 0.200 | 1.000 |
| 0.8042 | 203 | 0.950 | 0.950 | 2.500 | 0.200 | 1.000 |

You can see that **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.