

How to Create a Bland-Altman Plot to Compare Two Sets of Data

Authored by
stats writer

December 6, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Create a Bland-Altman Plot to Compare Two Sets of Data*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106450>

The Bland-Altman plot is a powerful graphical technique used extensively in statistical analysis and clinical research to assess the agreement between two different methods of measurement. Developed by J. Martin Bland and Douglas G. Altman, this plot provides an intuitive visualization that moves beyond simple correlation analysis, which can often be misleading when assessing true interchangeability.

The primary purpose of the Bland-Altman plot is to determine if two distinct measurement techniques--such as a new, cost-effective device versus a traditional gold standard--are interchangeable. It specifically helps researchers identify the presence of any systematic bias and the extent of random variation across the entire range of measurements.

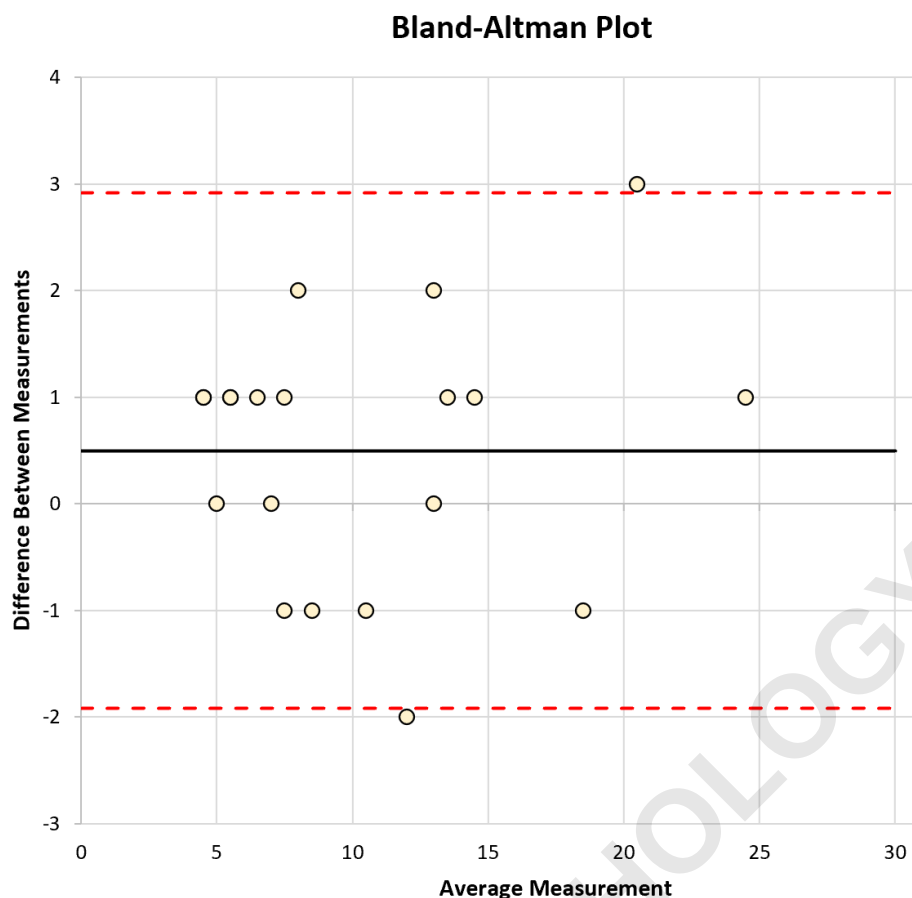
By plotting the difference between the paired measurements against their average, the plot effectively highlights the magnitude and distribution of disagreements, making it an indispensable tool for validating new clinical or analytical methods. This graphical representation is far more informative than simple correlation coefficients when evaluating the true agreement of instruments.

Defining the Bland-Altman Plot Methodology

The **Bland-Altman plot** is fundamentally designed to visualize the discrepancies, or differences, that arise when the same set of subjects or items are measured using two separate instruments or two distinct measurement protocols. This visualization is essential for method comparison studies, particularly when transitioning to newer, potentially less validated, methods.

The core application involves evaluating a novel instrument or procedure against an established, standard method. The central question addressed is: How closely does the new technique replicate the measurements obtained by the standard technique? If the two methods are deemed interchangeable, the new method can be adopted with confidence, often due to advantages such as lower cost, increased speed, or reduced invasiveness.

Understanding the Structure of the Plot Axes



The structure of the Bland-Altman plot is deliberately chosen to maximize interpretability. The horizontal axis (X-axis) represents the average of the two paired measurements, calculated as $(\text{Method A} + \text{Method B}) / 2$. This value serves as a proxy for the true magnitude of the quantity being measured and allows us to detect if the agreement between the methods changes systematically across different measurement levels.

The vertical axis (Y-axis) is the core diagnostic component, displaying the difference between the two measurements (Method A minus Method B). If the two methods perfectly agreed, all data points would align along the zero line on the Y-axis. The scatter of the points around this zero line indicates the random error or lack of agreement between the techniques.

For statistical inference and clinical decision-making, three critical horizontal lines are overlaid onto the scatter plot. These lines provide the basis for assessing systematic and random error:

The central line, which marks the overall average difference in measurements across all pairs.

The upper limit of the 95% Limits of Agreement (LoA).

The lower limit of the 95% Limits of Agreement (LoA).

Interpreting Key Statistical Indicators: Bias and Limits of Agreement

The Bland-Altman methodology provides concise answers to two fundamental questions regarding method comparison, which are visually addressed by the three horizontal lines overlaying the data points. These insights are critical for determining clinical acceptability and method interchangeability.

Question 1: Analyzing the Systematic Bias

The central horizontal line represents the overall average difference in measurements, commonly referred to as the bias. This value quantifies the average systematic disagreement between the two instruments. Ideally, this line should be close to zero, suggesting there is no systematic tendency for one method to produce consistently higher or lower readings than the other.

A significant displacement of this line from zero--whether positive or negative--confirms the presence of systematic bias. For instance, if the mean difference is positive, Method A consistently reads higher than Method B by that average amount. This bias dictates the necessary adjustment or calibration required if the methods are to be considered interchangeable.

Question 2: Establishing the Limits of Agreement

The upper and lower horizontal lines define the 95% Limits of Agreement (LoA). These limits provide a practical range within which 95% of the differences between the two measurement methods are expected to fall. Unlike the central bias line, the LoA accounts for random variation and helps gauge the precision of the agreement.

The width of the LoA is crucial for interpretation. A narrow range implies high agreement and low random error, suggesting the methods are highly interchangeable. Conversely, a wide range indicates substantial random variability, meaning that while the average difference (bias) might be small, individual measurements could differ significantly. Researchers must then compare this observed range against a predefined, clinically acceptable range of error to determine if the methods are interchangeable in practice.

To solidify the theoretical understanding of this plot, we will now proceed through a detailed, step-by-step example demonstrating the calculation and visualization required to construct and interpret a complete Bland-Altman plot from raw data.

Note on Nomenclature: The Bland-Altman plot is also historically known as the Tukey mean-difference plot, and these terms are frequently used interchangeably in statistical literature. However, the methodology popularized by Bland and Altman remains the most recognized standard for assessing agreement.

Step 1: Data Acquisition and Preparation

For this practical illustration, imagine a biologist conducting a method comparison study to evaluate the agreement between two devices, Instrument A and Instrument B, both designed to measure the weight of amphibians in grams. The goal is to determine if Instrument B, perhaps a newer, cheaper model, can reliably replace the existing standard, Instrument A.

The study involves measuring the weight of a sample of 20 different frogs using both instruments. Crucially, each frog serves as its own control, providing a paired measurement for comparison. The resulting raw data for the 20 paired measurements are tabulated below, showing the measured weight (in grams) obtained from Instrument A and Instrument B for each individual frog.

	A	B
1	5	4
2	5	4
3	5	5
4	6	5
5	6	5
6	7	7
7	7	8
8	7	6
9	8	9
10	8	7
11	9	7
12	10	11
13	11	13
14	13	13
15	14	12
16	14	13
17	15	14
18	18	19
19	22	19
20	25	24

Step 2: Calculating Paired Averages and Differences

Next, we must transform the raw data into the necessary coordinates for the Bland-Altman plot. For every pair of measurements (A and B), we calculate two key values: the average measurement and the difference in measurements.

Average Measurement (X-coordinate): Calculated as $(A + B) / 2$. This value represents the best

estimate of the true weight for that frog and will form the horizontal axis of the plot.

Difference in Measurements (Y-coordinate): Calculated as $(A - B)$. This shows how much Instrument A differs from Instrument B for that specific frog and will form the vertical axis of the plot.

Performing these calculations for all 20 data points yields the following augmented dataset, ready for the statistical calculations in the next step:

	A	B	Average	Difference
1	5	4	4.5	1
2	5	4	4.5	1
3	5	5	5	0
4	6	5	5.5	1
5	6	5	5.5	1
6	7	7	7	0
7	7	8	7.5	-1
8	7	6	6.5	1
9	8	9	8.5	-1
10	8	7	7.5	1
11	9	7	8	2
12	10	11	10.5	-1
13	11	13	12	-2
14	13	13	13	0
15	14	12	13	2
16	14	13	13.5	1
17	15	14	14.5	1
18	18	19	18.5	-1
19	22	19	20.5	3
20	25	24	24.5	1

Step 3: Determining Bias and Limits of Agreement (LoA)

To establish the three critical reference lines for the plot, we must perform statistical calculations on the Difference column generated in Step 2. These calculations will quantify the systematic error (bias) and the expected variability (Limits of Agreement).

First, we calculate the **Mean Difference** (the average of all values in the Difference column), which serves as the central bias line, which turns out to be **0.5** grams. We also calculate the standard deviation (s) of the differences, which is **1.235** grams.

The upper and lower limits of the Limits of Agreement are calculated by taking the mean difference and adding or subtracting 1.96 times the standard deviation (s), corresponding to a 95% coverage:

Upper Limit: $X + 1.96*s = 0.5 + 1.96*1.235 = 2.92$

Lower Limit: $X - 1.96*s = 0.5 - 1.96*1.235 = -1.92$

Here's how to interpret the resulting calculated values:

Systematic Error: The mean difference of 0.5 grams signifies that Instrument A consistently provides measurements that are 0.5 grams heavier than Instrument B, on average.

Range of Agreement: The 95% Limits of Agreement range from -1.92 grams to 2.92 grams. This implies that 95% of the observed differences between the two instruments are expected to fall within this range.

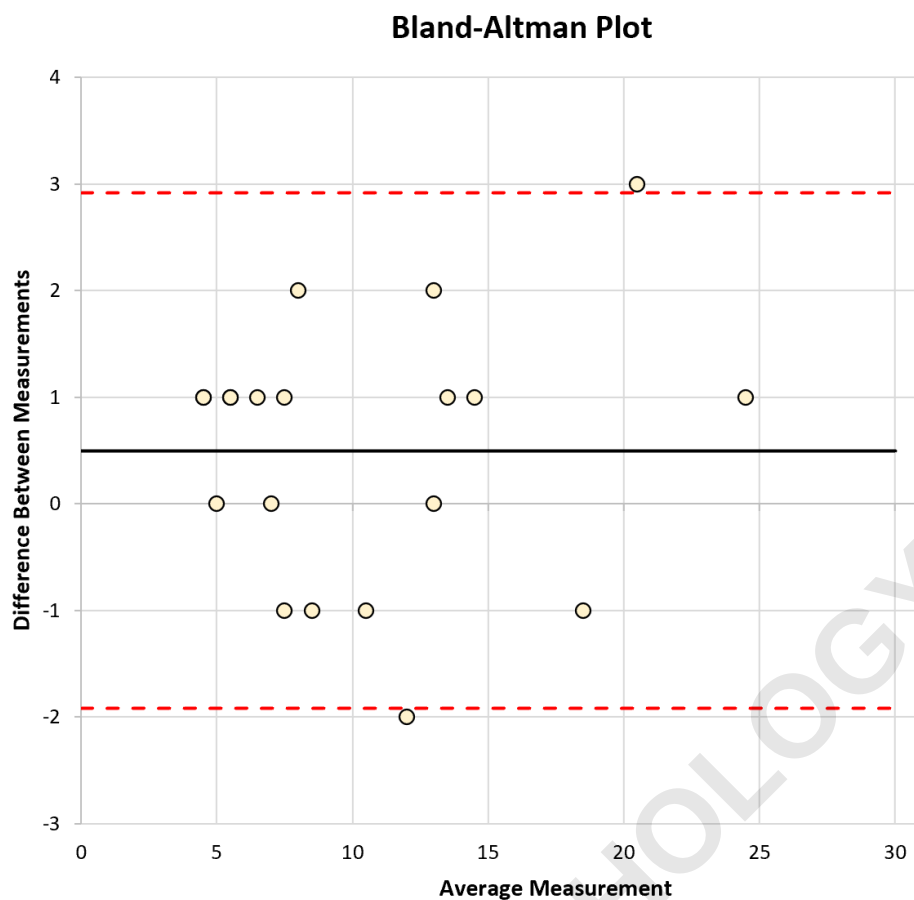
The next step focuses on the visual representation of these findings.

Step 4: Visualizing Agreement with the Bland-Altman Plot

The final step involves generating the Bland-Altman plot by mapping the paired average (X-axis) against the paired difference (Y-axis) for all 20 measurements. This scatter plot immediately reveals the relationship between the magnitude of the measurement and the disagreement between the instruments.

We then overlay the calculated reference lines onto this visualization: the mean difference (0.5), the Upper Limit of Agreement (2.92), and the Lower Limit of Agreement (-1.92). These lines frame the visual assessment of agreement, allowing researchers to quickly identify potential issues such as heteroscedasticity (where the difference widens as the magnitude increases) or outliers falling outside the 95% LoA.

The resulting plot comprehensively visualizes the data analysis, providing both the systematic error and the typical spread of random error:



Conclusion and Practical Application

In this example, the biologist must evaluate if a potential measurement difference of roughly 5 grams (the total range from -1.92 to 2.92) is acceptable for measuring frog weights in their specific context. If the clinical or biological requirement dictates that measurement differences must be less than, say, 1 gram, then Instrument B cannot simply replace Instrument A without recalibration or modification, as the limits of agreement exceed this critical threshold. The Bland-Altman plot thus provides a clear, quantitative basis for making crucial decisions about method interchangeability in scientific research.