

# How to Easily Create a Bland-Altman Plot in Python

Authored by  
**stats writer**

December 6, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Create a Bland-Altman Plot in Python*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106431>

The Bland-Altman Plot is a fundamental graphical tool used extensively in medical statistics and metrology to assess the agreement between two different quantitative measurements. Developed by J. Martin Bland and Douglas G. Altman in 1986, this technique addresses the limitations of simply using correlation coefficients, which measure association but not necessarily agreement.

Unlike standard correlation analysis, which only indicates whether two methods are related, the Bland-Altman method specifically visualizes the systematic and random differences between two measurement techniques or instruments applied to the same subjects. It is essential for determining if a new, potentially cheaper or faster measurement method can reliably replace an established standard method. This plot is sometimes referred to as a difference plot.

Creating a Bland-Altman Plot in Python is efficiently achieved using the powerful statsmodels library. This library provides the specialized `statsmodels.graphics.blandaltman` module, which handles the necessary calculations for the mean difference and the Limits of Agreement (LOA). This comprehensive tutorial will guide you through the process, from data preparation using Pandas to generating and interpreting the final visualization.

## Why Agreement Differs from Correlation

A common mistake when comparing two measurement techniques is relying solely on the correlation coefficient (Pearson's  $r$ ). While a high correlation (e.g.,  $r > 0.9$ ) suggests that the methods respond similarly across the range of values, it does not confirm that the actual values produced by the two methods are interchangeable. Correlation measures the strength of the linear relationship, but not the actual magnitude of the differences.

Consider a scenario where Instrument B consistently measures 5 units higher than Instrument A. The correlation between A and B would be perfect ( $r=1.0$ ), indicating a strong linear association. However, the instruments clearly do not agree because there is a systematic bias of 5 units. The Bland-Altman Plot is designed precisely to expose such systematic bias and quantify the extent of random error, offering a much more rigorous assessment of interchangeability.

The principle behind the Bland-Altman Plot is simple yet powerful: it plots the difference between the two measurements (Y-axis) against their average (X-axis). By plotting the differences against the magnitude of the measurement, we can also identify if the discrepancy between the two methods changes systematically as the value being measured increases (known as heteroscedasticity). The plot automatically calculates the mean difference and the Limits of Agreement (LOA), providing critical context for data interpretation.

## Key Statistical Elements of the Plot

The resulting Bland-Altman Plot displays several critical reference lines that form the basis for

interpretation. Understanding these elements is essential for drawing meaningful conclusions about agreement:

**The X-axis (Mean Value):** This axis represents the average of the two measurements for each subject, typically calculated as  $(\text{Measurement A} + \text{Measurement B}) / 2$ . Plotting against the mean, rather than one of the individual measurements, avoids the assumption that one method is the "true" standard.

**The Y-axis (Difference):** This axis represents the difference between the two measurements, calculated as  $\text{Measurement A} - \text{Measurement B}$ . If the instruments agree perfectly, all points would lie on the zero line.

**Mean Difference Line (Bias):** This solid horizontal line represents the average difference between the two methods (the mean of the Y-axis values). If the mean difference is significantly different from zero, it indicates a fixed systematic bias. For the two methods to be truly interchangeable, this value should ideally be close to zero.

**Limits of Agreement (LOA):** These are the two dashed horizontal lines, typically placed 1.96 standard deviations above and below the mean difference line. The formula for the LOA is:  $\text{Mean Difference} \pm (1.96 \times \text{Standard Deviation of the Differences})$ . The LOA defines the range within which 95% of the differences between the two methods are expected to fall.

The interpretation of the plot hinges on determining whether the discrepancies within the Limits of Agreement are clinically or practically acceptable. If the differences are too large for the context of the study, the methods cannot be considered interchangeable, regardless of how tightly clustered the points appear graphically.

## Prerequisites and Setting Up the Environment

To follow this tutorial and generate the plot, you must have a working [Python](#) environment installed, preferably utilizing Anaconda or a similar package manager, which bundles many scientific computing libraries. The primary libraries required for this task are [Pandas](#) for efficient data handling, `matplotlib` for plotting visualization utilities, and, most importantly, [statsmodels](#), which contains the specialized function for generating the Bland-Altman plot.

If you do not have these libraries installed, you can easily install them using the Python package installer, `pip`. Ensure your environment is up to date before proceeding. Note that `statsmodels` relies on `numpy` and `scipy`, which are usually installed automatically as dependencies when installing `statsmodels`.

The necessary installation commands are straightforward:

## **pip install pandas statsmodels matplotlib**

Once the environment is properly configured, we can import the necessary modules. This setup ensures that all tools needed for data manipulation, statistical calculation, and visual output are ready for use within the [Python](#) script or Jupyter Notebook environment.

## **Step 1: Preparing Sample Data for Analysis**

For demonstration purposes, let us consider a practical scenario in biology or medicine: a researcher is comparing two distinct measurement devices, Instrument A and Instrument B, used to quantify a specific biological characteristic, such as the weight of an animal sample. We assume a biologist uses these two instruments to measure the weight (in grams) of the same set of 20 different frogs.

We need to structure this data into a format suitable for analysis, where each measurement pair corresponds to a single subject (frog). Using the [Pandas](#) library, we create a DataFrame where 'A' represents the measurements from Instrument A and 'B' represents the measurements from Instrument B. This DataFrame structure is optimal for subsequent numerical processing.

The data below represents the paired measurements for the 20 subjects. It is crucial that the data points are correctly aligned, meaning the first value in column 'A' and the first value in column 'B' correspond to the same individual subject. This pairing is fundamental to calculating the differences accurately for the Bland-Altman analysis.

```
import pandas as pd
```

```
df = pd.DataFrame({'A': ,  
'B': })
```

## **Step 2: Generating the Bland-Altman Plot in Python**

With our data successfully prepared in a Pandas DataFrame, the next step is to leverage the functionality provided by the `statsmodels.graphics` module. Specifically, we utilize the powerful `mean_diff_plot()` function. This function abstracts away the tedious steps of calculating the mean, difference, standard deviation, and plotting the confidence intervals, allowing the user to generate a publication-ready plot with minimal code.

We must import `statsmodels` and `matplotlib.pyplot` for visualization control. The `mean_diff_plot()` function accepts the two series of measurements (`df.A` and `df.B`) as required positional arguments. It automatically calculates the average of the measurements (X-axis) and the

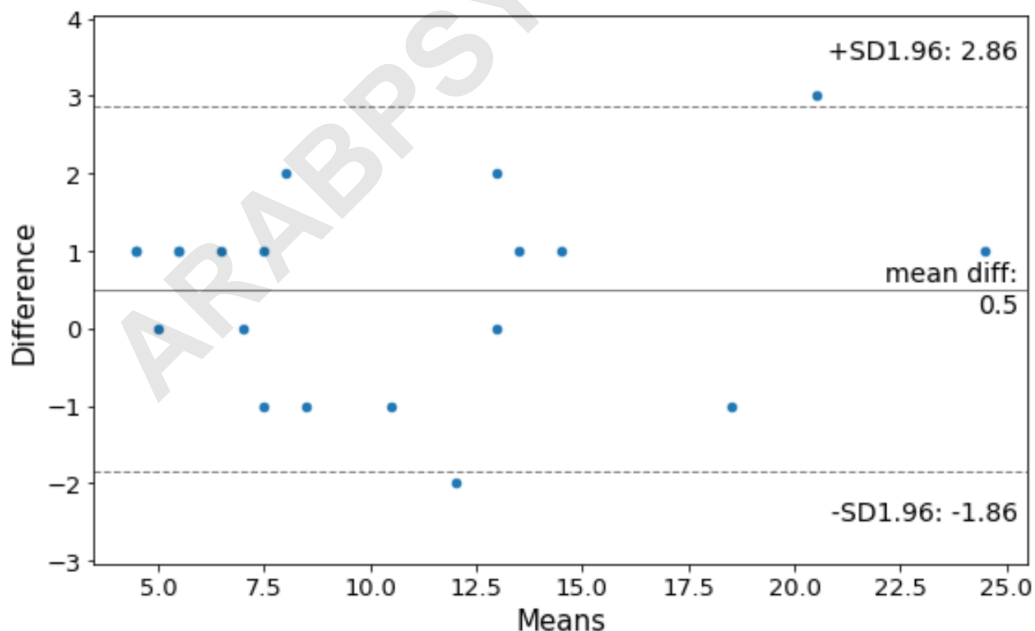
difference between them (Y-axis), plots the scatter points, and renders the crucial bias and LOA lines.

We initialize a figure and an axes object using `plt.subplots()` to ensure we have fine-grained control over the plot dimensions and appearance. The generated plot is then directed to this axes object via the `ax = ax` argument within the `mean_diff_plot()` call. Finally, we display the plot using `plt.show()`.

```
import statsmodels.api as sm  
import matplotlib.pyplot as plt
```

```
# Create figure and axes for the plot  
f, ax = plt.subplots(1, figsize = (8,5))  
  
# Generate Bland-Altman plot using statsmodels  
sm.graphics.mean_diff_plot(df.A, df.B, ax = ax)  
  
# Display the plot visualization  
plt.show()
```

Executing this code block in Python will immediately generate the visual output representing the agreement analysis between Instrument A and Instrument B.



## Interpreting the Visual Output

Once the visualization is generated, the critical phase of interpretation begins. The visual representation conveys a wealth of information regarding the comparability of the two instruments or methods. We must systematically examine the relationship between the plotted points and the reference lines.

The **x-axis** of the plot displays the average measurement obtained from the two instruments. This spread shows the measurement range across which the comparison is valid. The **y-axis** displays the raw difference in measurements (A minus B). If the two methods perfectly agreed across all subjects, all data points would fall along the  $Y=0$  line.

The **black solid line**, positioned at  $Y = 0.5$ , represents the average difference or **bias** between the two instruments. In this sample scenario, the average difference turns out to be **0.5**. Since this value is positive, Instrument A tends to measure slightly higher than Instrument B by an average of 0.5 grams across the sample set. This positive bias indicates a systematic difference that must be considered when using these instruments interchangeably.

The two **dashed lines** delineate the 95% Limits of Agreement (LOA). These limits are calculated based on the variability of the differences. For this dataset, the 95% confidence interval for the average difference is . This means that for a new frog measured by these two instruments, we can expect the difference in measurements to fall between -1.86 grams and 2.86 grams, 95% of the time.

## Analyzing Agreement and Clinical Acceptability

The most crucial step in interpreting the Bland-Altman Plot is determining whether the observed bias (0.5) and the spread defined by the LOA () are clinically or practically acceptable. Statistical significance is often irrelevant here; the focus is on practical significance based on the field of study.

For example, if the tolerance for weight measurement in frogs is  $\pm 1.0$  gram, then the observed LOA range of nearly 5 grams ( $2.86 - (-1.86) = 4.72$ ) suggests that the differences are too large for the instruments to be considered interchangeable. If the acceptable tolerance is much wider, say  $\pm 5.0$  grams, then the instruments might be deemed agreeable for that particular application.

Furthermore, we must visually inspect the scatter of the points. If the points are uniformly scattered across the plot, it indicates that the differences are consistent regardless of the magnitude of the measurement (homoscedasticity). However, if the spread of points widens as the mean measurement increases (the funnel effect), this suggests that the discrepancy between instruments increases for larger weights. If this occurs, standard LOA calculation might be

inappropriate, and methods such as logarithmic transformation or regression-based LOA might be required.

ARABPSYCHOLOGY.COM