

# How to Easily Create a Bland-Altman Plot in R: A Step-by-Step Guide

Authored by  
**stats writer**

December 6, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Create a Bland-Altman Plot in R: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106433>

## Understanding the Bland-Altman Plot

A Bland-Altman plot, often referred to as a difference plot, is an essential graphical tool used in statistical analysis to assess the agreement between two different quantitative measurement methods. It visualizes the differences in measurements between two instruments or two techniques against their mean.

This technique is crucial in fields ranging from clinical research to engineering, where determining the comparability of two measurement systems is vital. It helps analysts determine if the two methods are sufficiently similar for them to be used interchangeably. Unlike correlation, which only measures the strength of the linear relationship, the Bland-Altman method specifically evaluates the existence and magnitude of systematic bias and proportional error.

This comprehensive tutorial provides a rigorous, step-by-step example of how to construct and interpret a Bland-Altman plot using the powerful statistical programming language, R. We will leverage R's data manipulation capabilities and the robust visualization features of the ggplot2 package.

### Step 1: Preparing Your R Environment and Data

To demonstrate this process, consider a common scenario in biological research. Suppose a biologist conducts an experiment where the weight of 20 distinct frogs, measured in grams, is recorded using two separate measuring instruments: Instrument A and Instrument B. The primary objective is to determine if the measurements produced by Instrument A and Instrument B are in acceptable agreement.

We begin by establishing the raw dataset within the R environment. This data frame structure, shown below, organizes the paired observations for subsequent analysis. It is imperative that the data is structured correctly, with one column representing the measurement from the first method (A) and the second column representing the measurement from the second method (B).

The following R code snippet initializes the data frame, ensuring that the necessary numerical vectors are established correctly for all 20 subjects.

```
#create data frame containing measurements from Instrument A and Instrument B
```

```
df <- data.frame(A=c(5, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9,  
10, 11, 13, 14, 14, 15, 18, 22, 25),  
B=c(4, 4, 5, 5, 5, 7, 8, 6, 9, 7, 7, 11,  
13, 13, 12, 13, 14, 19, 19, 24))
```

```
#view the initial structure and first few rows of the data frame
```

```
head(df)
```

```
A B
```

```
1 5 4
```

```
2 5 4
```

```
3 5 5
```

```
4 6 5
```

```
5 6 5
```

```
6 7 7
```

## Step 2: Defining Key Metrics: Average and Difference

The standard Bland-Altman plot is constructed by plotting two critical metrics for each subject: the difference between the two measurements (Y-axis) and the average of the two measurements (X-axis). These calculations must be performed before visualization can occur.

The average measurement (or mean) serves as an estimate of the true magnitude of the quantity being measured. Plotting the difference against this average helps reveal if the disagreement between the two instruments is consistent across the entire range of measurements (i.e., whether the bias is constant or proportional to the measurement magnitude).

Therefore, the subsequent step involves appending two new columns to our data frame: one designated `avg`, containing the mean of A and B for each frog, and another designated `diff`, containing the simple subtraction of B from A. These derived columns form the fundamental coordinates for our plot.

## Step 3: Executing the R Code for Metric Calculation

We utilize the built-in R function `rowMeans()` to efficiently calculate the average for each row, representing the individual frog's estimated weight. The difference calculation is straightforward arithmetic. These operations are performed vectorially, ensuring high computational efficiency.

```
#create new column for average measurement (X-axis value for the plot)
```

```
df$avg <- rowMeans(df)
```

```
#create new column for difference in measurements (Y-axis value for the plot)
```

```
df$diff <- df$A - df$B
```

```
#view the first six rows of the augmented data frame, including the new columns
```

```
head(df)
```

```
A B avg diff
1 5 4 4.5 1
2 5 4 4.5 1
3 5 5 5.0 0
4 6 5 5.5 1
5 6 5 5.5 1
6 7 7 7.0 0
```

## Step 4: Establishing the Limits of Agreement (LOA)

Before plotting the individual data points, we must calculate the statistically significant reference lines that define the acceptable range of agreement. These lines typically include the mean difference (the estimated bias) and the 95% Limits of Agreement (LOA). The LOA defines the range within which 95% of the differences between the two methods are expected to fall, assuming a normal distribution of the differences.

The calculation for the LOA is based on the mean difference ( $\bar{d}$ ) and the standard deviation of the differences ( $SD_d$ ). The formula for the upper and lower limits is  $\bar{d} \pm 1.96$  times  $SD_d$ . The factor 1.96 corresponds to the Z-score for a two-sided 95% confidence interval.

The mean difference, calculated first, quantifies the systematic bias: if this value is far from zero, one instrument consistently reads higher or lower than the other. The standard deviation, using the R function `sd()`, determines the extent of random variation or spread around this mean bias.

## Step 5: Calculating the Statistical Boundaries in R

We now calculate the three horizontal reference lines necessary for interpreting the plot: the mean difference, the lower 95% LOA, and the upper 95% LOA. These values will be used to draw horizontal lines onto the resulting visualization.

**#find average difference (the systematic bias)**

```
mean_diff <- mean(df$diff)
```

```
mean_diff
```

```
0.5
```

**#find lower 95% confidence interval limits, using 1.96 times the standard deviation**

```
lower <- mean_diff - 1.96*sd(df$diff)
```

```
lower
```

```
-1.921465
```

```
#find upper 95% confidence interval limits
```

```
upper <- mean_diff + 1.96*sd(df$diff)
```

```
upper
```

```
2.921465
```

Upon executing these calculations, we determine that the average difference (systematic bias) between the two instruments is **0.5** grams. This indicates that Instrument A generally measures 0.5 grams higher than Instrument B. Furthermore, the 95% confidence interval for the average difference--our Limits of Agreement--is calculated to be . This range provides the boundaries for judging the acceptability of the agreement.

## Step 6: Generating the Visualization using ggplot2

The preferred method for creating high-quality, reproducible statistical graphics in R is by utilizing the `ggplot2` package. This package, based on the grammar of graphics, allows for precise control over aesthetic elements and graphical layers. If `ggplot2` is not already installed on your system, it must be installed and then loaded using the `library()` function.

The construction of the plot starts by mapping the calculated average (`avg`) to the aesthetic X-axis and the difference (`diff`) to the Y-axis. We use `geom_point()` to plot the 20 individual observations, ensuring that each frog's disagreement is visualized relative to its overall magnitude of measurement.

## Step 7: Adding Reference Lines for Interpretation

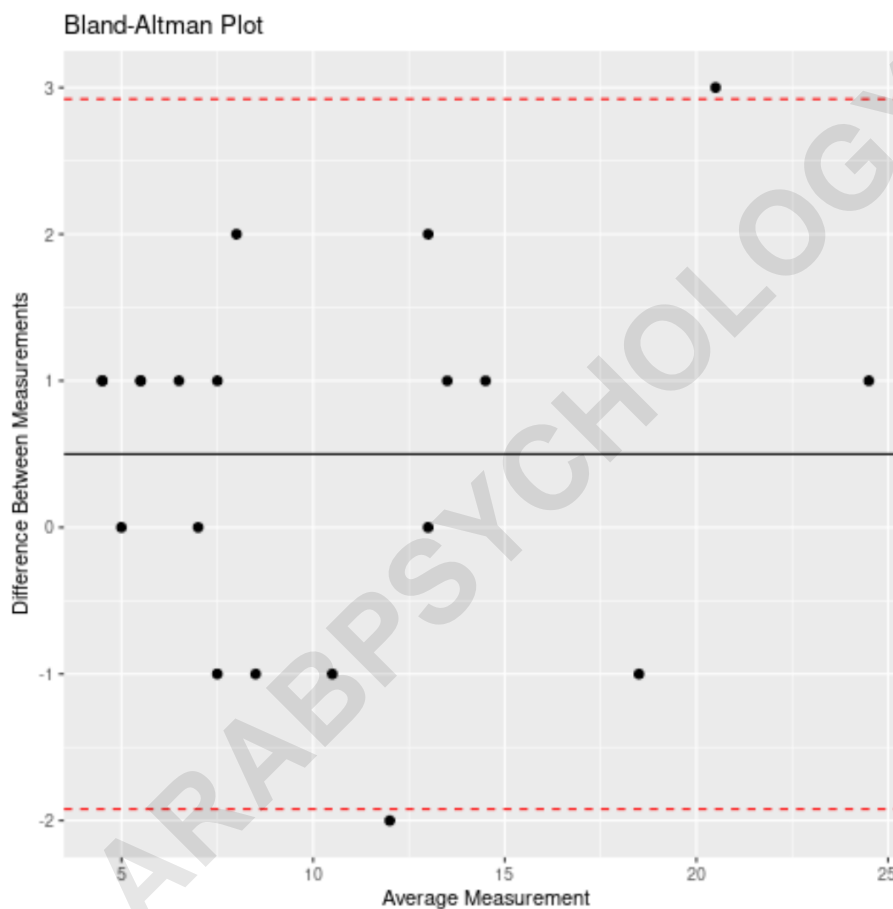
A crucial aspect of the Bland-Altman plot is the inclusion of the reference lines derived in the previous steps. We use `geom_hline()` to draw these lines, which facilitate immediate visual assessment of agreement and bias. The mean difference is represented by a solid line, indicating the systematic offset, while the LOA boundaries are typically drawn as dashed red lines.

The final code block compiles the loading of the `ggplot2` library, the aesthetic mapping, the geometric elements (points and horizontal lines), and the labels (title and axis descriptions) to produce the final, professional visualization.

```
#load ggplot2 visualization package
```

```
library(ggplot2)
```

```
#create Bland-Altman plot
ggplot(df, aes(x = avg, y = diff)) +
  geom_point(size=2) +
  geom_hline(yintercept = mean_diff) +
  geom_hline(yintercept = lower, color = "red", linetype="dashed") +
  geom_hline(yintercept = upper, color = "red", linetype="dashed") +
  ggtitle("Bland-Altman Plot: Agreement Between Instrument A and B") +
  ylab("Difference Between Measurements (A - B)") +
  xlab("Average Measurement ((A + B) / 2)")
```



## Step 8: Interpreting the Final Bland-Altman Diagram

The resulting plot provides a clear visual summary of the agreement. The **x-axis** represents the average measurement of the two instruments, allowing us to examine whether the agreement changes as the magnitude of the measured weight increases. The **y-axis** displays the difference in measurements, showing the actual disparity between the two instruments for each observation.

The central solid **black line** represents the mean difference (0.5). If this line were exactly at zero, it would indicate no systematic bias. Since it is slightly positive, it reinforces the finding that Instrument A tends to measure slightly higher than Instrument B across the sample.

The two **red dashed lines** represent the 95% Limits of Agreement (LOA). These limits define the variability of the difference. If nearly all points fall within these two limits, it suggests good agreement. Furthermore, the plot must be examined for patterns. If the scatter of points widens at higher average measurements, this indicates proportional bias, meaning the instruments disagree more widely when measuring larger weights--a critical finding that simply calculating the mean difference alone would miss. In this specific example, the points appear randomly scattered around the mean difference line, suggesting that the systematic bias is relatively consistent across the range of measurements. A high or low standard deviation of the differences (which defines the distance between the LOA lines) determines the clinical or practical acceptability of the disagreement.