

What are some examples of comparing box plots and how can it be done?

Authored by
stats writer

April 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What are some examples of comparing box plots and how can it be done?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138721>

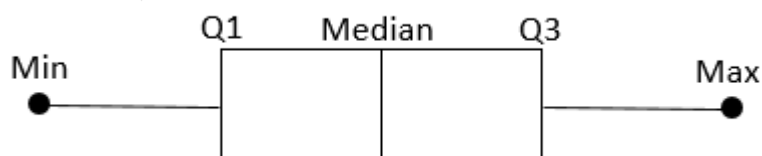
Comparing box plots involves visually examining the distribution and central tendency of data sets. It can be done by looking at the shape, position, and spread of the boxes, whiskers, and outliers. For example, if two box plots have similar box sizes and positions but one has longer whiskers, it may suggest a greater variability in that data set. Similarly, comparing the medians and quartiles of two box plots can indicate differences in central tendency. Additionally, box plots can be overlaid or placed side by side for easier comparison. Overall, comparing box plots allows for a quick and effective way to identify similarities and differences between data sets.

Compare Box Plots (With Examples)

A box plot is a type of plot that displays the five number summary of a dataset, which includes:

The minimum value
The first quartile (the 25th percentile)
The median value
The third quartile (the 75th percentile)
The maximum value

To make a box plot, we draw a box from the first to the third quartile. Then we draw a vertical line at the median. Lastly, we draw "whiskers" from the quartiles to the minimum and maximum value.



Box plots are useful because they allow us to gain a

quick understanding of the distribution of values in a dataset. They're also useful for comparing two different datasets.

When comparing two or more box plots, we can answer four different questions:

1. How do the median values compare? We can compare the vertical line in each box to determine which dataset has a higher median value.

2. How does the dispersion compare? We can compare the length of each box (which represents the distance between Q1 and Q3 - the interquartile range) to determine which dataset is more spread out.

3. How does the skewness compare? The closer the vertical line is to Q1, the more positively skewed the dataset. The closer the vertical line is to Q3, the more negatively skewed the dataset.

4. Are outliers present? In box plots, outliers are typically represented by tiny circles that extend beyond either whisker. An observation is defined to be an outlier if it meets one of the following criteria:

An observation is less than $Q1 - 1.5 \cdot IQR$ An observation is greater than $Q3 + 1.5 \cdot IQR$

The following example shows how to compare two different box plots and answer these four questions.

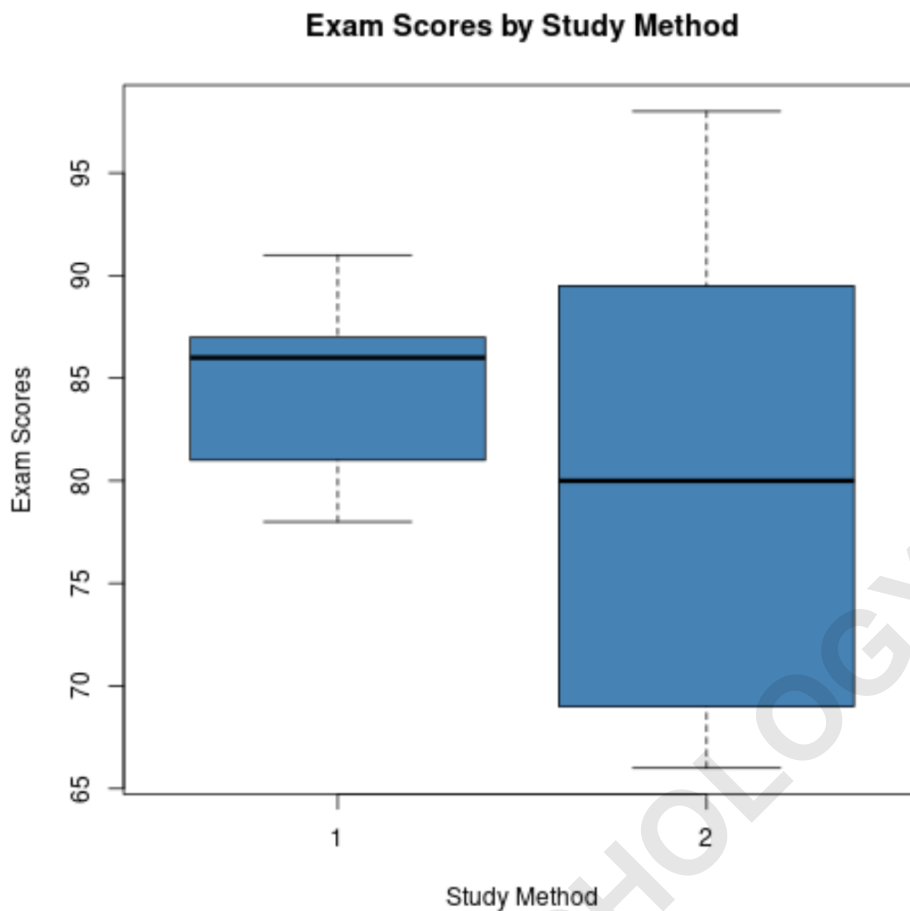
Example: Comparing Box Plots

The following datasets display the exam scores for students who used one of two studying techniques to prepare for the exam:

Method 1: 78, 78, 79, 80, 80, 82, 82, 83, 83, 86, 86, 86, 86, 87, 87, 87, 88, 88, 88, 91

Method 2: 66, 66, 66, 67, 68, 70, 72, 75, 75, 78, 82, 83, 86, 88, 89, 90, 93, 94, 95, 98

If we create box plots for each dataset, here's what they would look like:



We can compare these two box plots and answer the following four questions:

1. How do the median values compare? The line in the middle of the box plot for Study Method 1 is higher than the line for Study Method 2, which indicates that the students who used Study Method 1 had a higher median exam score.

2. How does the dispersion compare? The box plot for Study Method 2 is much longer than Study Method 1,

which indicates that the exam scores are much more spread out among students who used Study Method 2.

3. How does the skewness compare? The line in the middle of the box plot for Study Method 1 is close to Q3, which indicates that the distribution of exam scores for students who used Study Method 1 is negatively skewed. Conversely, the line in the middle of the box plot for Study Method 2 is near the center of the box, which means the distribution of scores has little skew at all.

4. Are outliers present? Neither box plot has tiny circles that extend beyond the top or bottom whiskers, which means neither dataset had any clear outliers.