

Box Plot

What: A Box Plot is a graphical analysis tool used to investigate the effect of discrete inputs on continuous outputs. They can represent the skewness of a distribution of data in an intuitive way.

When: Box Plots are used to summarize large quantities of data. In the Analyze phase of the DMAIC process to explore potential issues identified by the team. They can be used in the Improve phase to graphically represent the results of changes made by the team.

How: 1) **Gather the Data:** Gather the data to be used for calculating the Box Plot.

2) **Calculate 1st Quartile:** aka 25th percentile or Q1 is calculated using the following formula:
 $k = (\text{quart}/4) * (n-1) + 1$ (where quart is replaced with "1") Q1 is the upper end of the lower 25% of the data range.

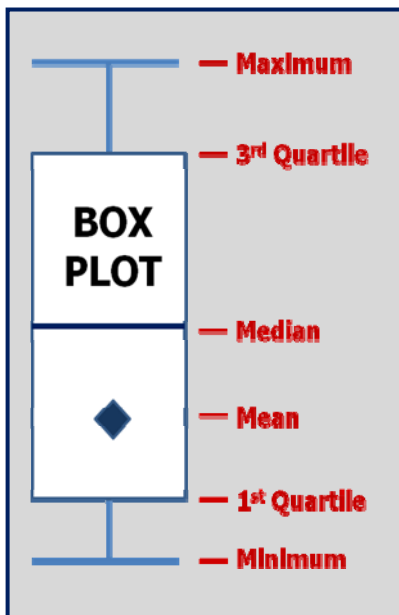
3) **Calculate 3rd Quartile:** aka 75th percentile or Q3 is calculated using the following formula:
 $k = (\text{quart}/4) * (n-1) + 1$ (where quart is replaced with "3") Q3 is the lower end of the upper 25% of the data range.

4) **Draw Box:** The box represents 50% of data that falls between Q1 and Q3.

5) **Calculate Median:** Draw a horizontal line through the box at the median.

6) **Determine Minimum & Maximum Points:** Use one of two methods (A) actual minimum/maximum or (B) calculated based on quartile

Calculation for Minimum $Q3 - 1.5 \times (Q3 - Q1)$, for Maximum $Q3 + 1.5 \times (Q3 - Q1)$



7) **Connect Maximum/Minimum with Box:** Draw a line between the Maximum and the 3rd Quartile and another between the 1st Quartile and the Minimum.

8) **Calculate Outliers:** Outliers are data points that fall outside the Minimum/Maximum points. To calculate the outliers use the following formula:

$$L1 = Q1 - 1.5 \times (Q3 - Q1)$$

$$U1 = Q3 + 1.5 \times (Q3 - Q1)$$

Outlier points are plotted with an asterisk.

Box Plot

Definitions/Calculations for Box Plot

Maximum: The minimum of

- The maximum value in the data set and
- $Q3 + 1.5 * IQR$

Q3 – 3rd Quartile: The 75th percentile of the data set.

IQR – Inter-quartile Range: This is the value of the 75th percentile (Q3) minus the value of the 25th percentile (Q1). This measure of scale attempts to measure the variability of points near the center. $Q3 - Q1$

Median – the median is the value of the point which has half the data smaller than that point and half the data larger than that point. That is, if X_1, X_2, \dots, X_N is a random sample sorted from smallest value to largest value, then the median is defined as:

$$\hat{Y} = Y_{(N+1)/2} \quad \text{if } N \text{ is odd}$$

$$\hat{Y} = (Y_{N/2} + Y_{(N/2+1)})/2 \quad \text{if } N \text{ is even}$$

Mean – the mean is the sum of the data points divided by the number of data points. That is,

$$\hat{Y} = \sum_{i=1}^N Y_i / N$$

Q1 – 1st Quartile: The 25th percentile of the data set.

$$k = (\text{quart}/4) * (n-1) + 1$$

Minimum: The maximum of

- the minimum value in the data set and
- $Q1 - 1.5 * IQR$

Box: The box portion of a Box Plot that represents the data between Q1 and Q3.

Outliers: Data points that lie beyond the IQR-based **Minimum** or **Maximum**

Whiskers: The lines extending from Q3 to the Maximum and from Q1 to the Minimum