

Statistics

Lecture 20: | Boxplot

Prof. Dr. Mostafa Elhosseini

<https://youtube.com/drmelhosseini>

| What is a Boxplot

- For some distributions/datasets, you will find that you **need more information than** the measures of **central tendency** (median, mean, and mode)
- Box plots show the **five-number summary of a set of data**: including the minimum score, first (lower) quartile, median, third (upper) quartile, and maximum score

| Why Boxplot?

- You need to have information on the variability or dispersion of the data.
- A **boxplot** is a graph that gives you a **good indication of how the values in the data are spread out**.
- It can be used to **determine the regions in which the sample values are more densely crowded** and the regions in which they are **more sparse**

| How to Plot?

- A **boxplot** is a standardized way of **displaying the distribution** of data based on a **five number summary**
 - Minimum: $Q1 - 1.5 * IQR$
 - First quartile (Q1),
 - Median (Q2),
 - Third quartile (Q3), and
 - Maximum: $Q3 + 1.5 * IQR$

| Outliers

- An **outlier** is an observation that is numerically **distant from the rest of the data**
- When reviewing a box plot, an **outlier** is defined as a data point that is **located outside the whiskers of the box plot**
- Outside 1.5 times the IQR above the Q3 and below Q1 ($Q1 - 1.5 * IQR$ or $Q3 + 1.5 * IQR$)

| References

- Understanding Boxplots <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>
- What does a box plot tell you?
<https://www.simplypsychology.org/boxplots.html>
- Statistics for Engineers and Scientists, Fourth Edition, William Navidi
Colorado School of Mines
- Khan Academy
<https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/cc-6th-box-whisker-plots/v/constructing-a-box-and-whisker-plot>