



Statistics for DATA SCIENCE

This Photo by Unknown author is licensed under [CC BY-SA-NC](#).

Bindu K R

Contents

Statistical Methods

Mean

Median

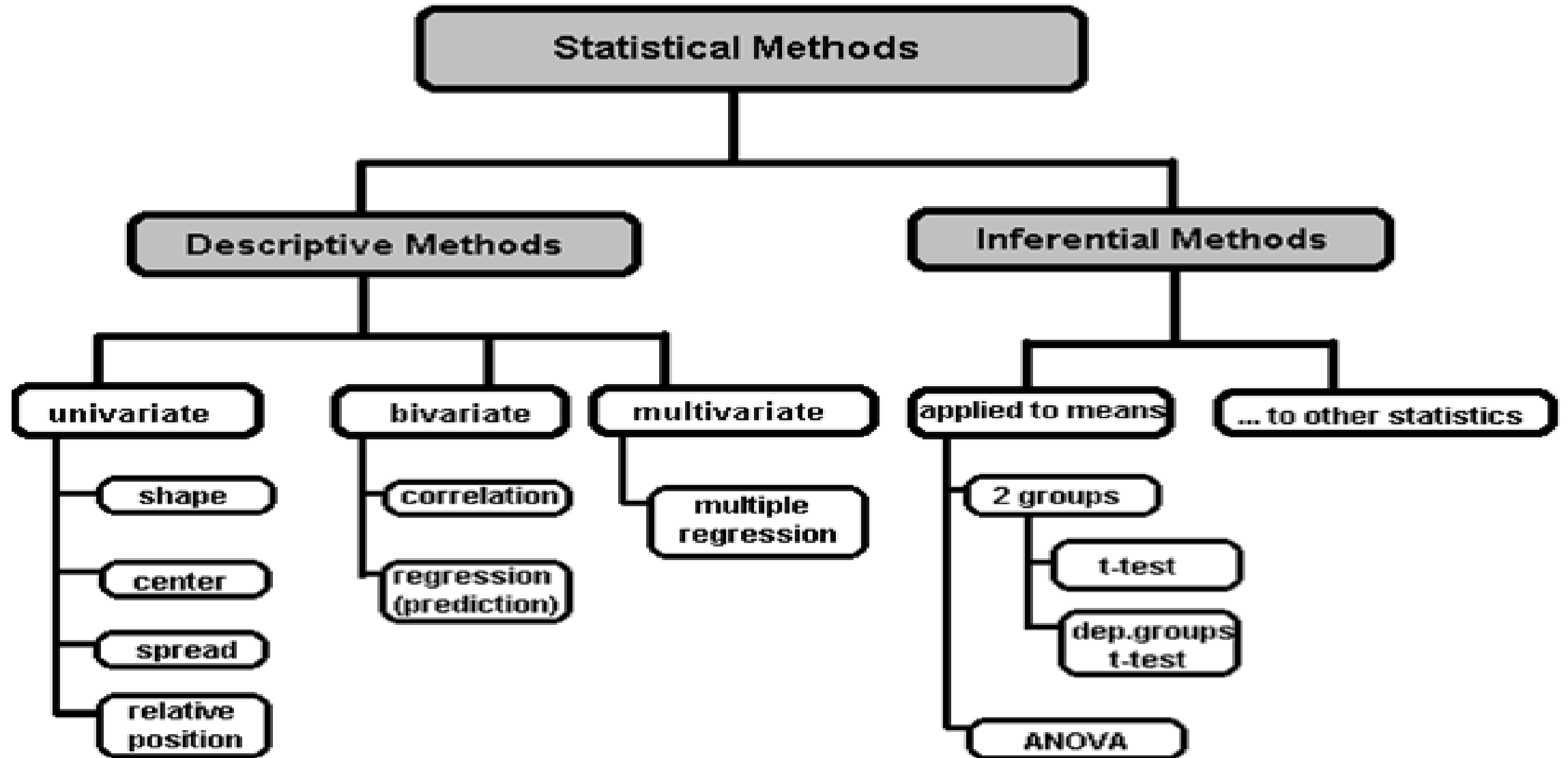
Mode

STDEV

Variance

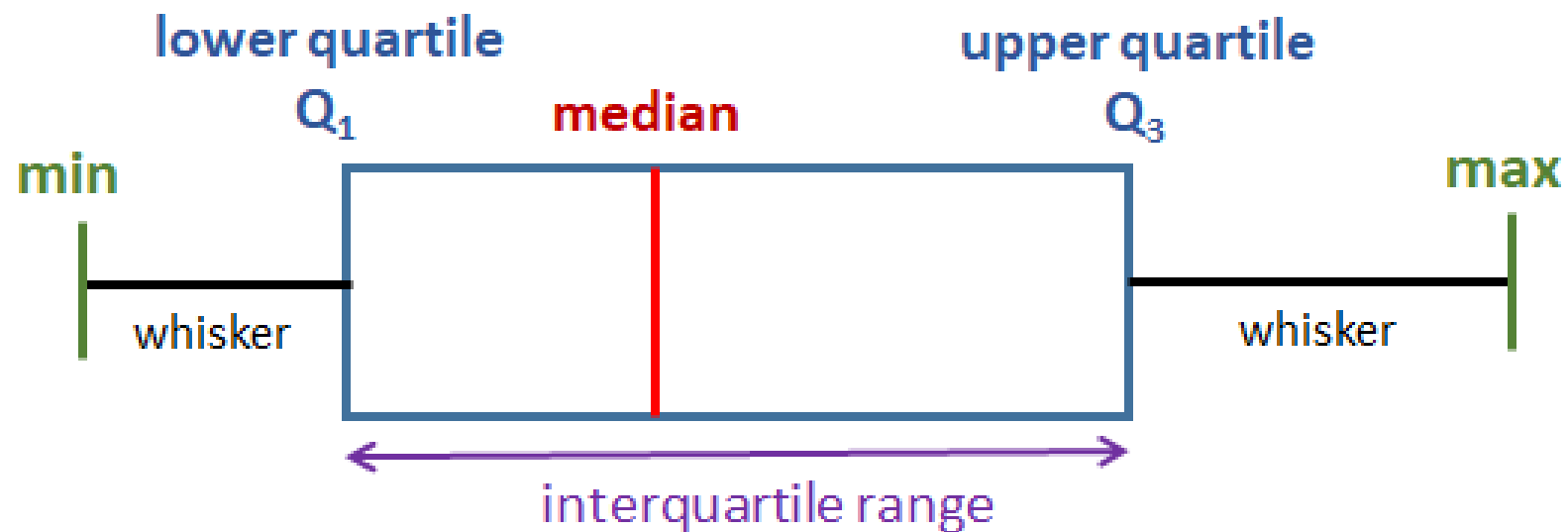
Univariate Analysis - Boxplot

Statistical methods

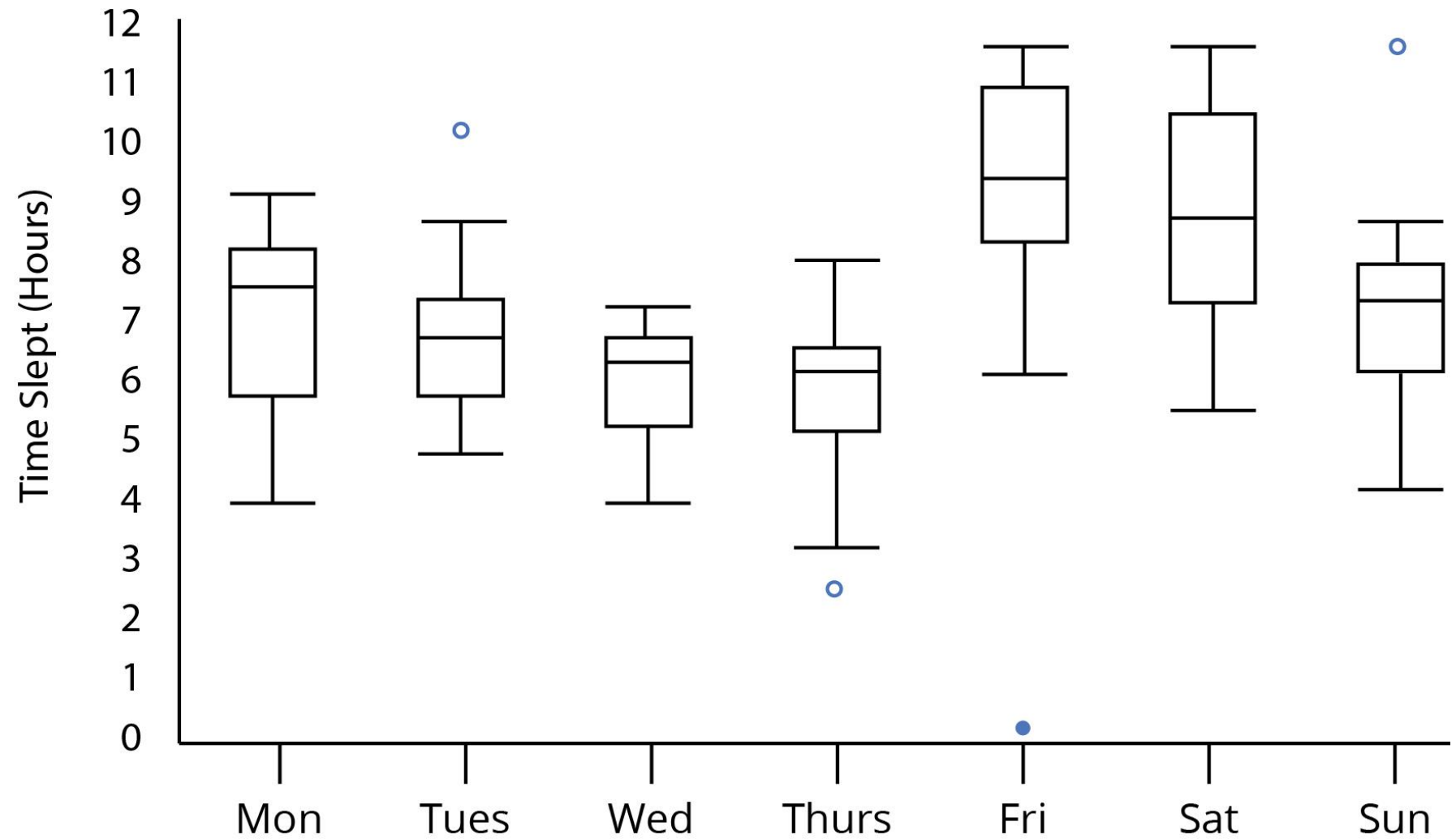


Box plot

A box and whisker plot (also called a box plot) shows the five-number summary of a set of data: **minimum**, **lower quartile**, **median**, **upper quartile**, and **maximum**.



Box plot



Box plot

- **Drawing A Box And Whisker Plot**

- **Example:**

Construct a box plot for the following data:

12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25

- **Solution:**

Step 1: Arrange the data in ascending order.

- Step 2: Find the median, lower quartile and upper quartile.

- **Median (middle value) = 22**

Lower quartile (middle value of the lower half) = 12

Upper quartile (middle value of the upper half) = 36

- **IQR Inter Quartile Range** interquartile range, $IQR = Q_3 - Q_1$

- **lower 1.5*IQR whisker = $Q_1 - 1.5 * IQR$** (If there is no data point at 4, then the lowest point greater than 4.)

- **upper 1.5*IQR whisker = $Q_3 + 1.5 * IQR$**

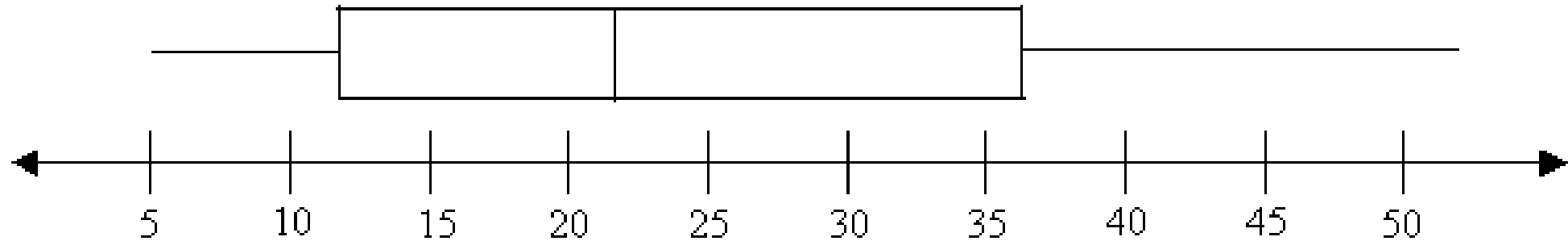
Box plot

5, 7, 12, 14, 15, 22, 25, 30, 36, 42, 53

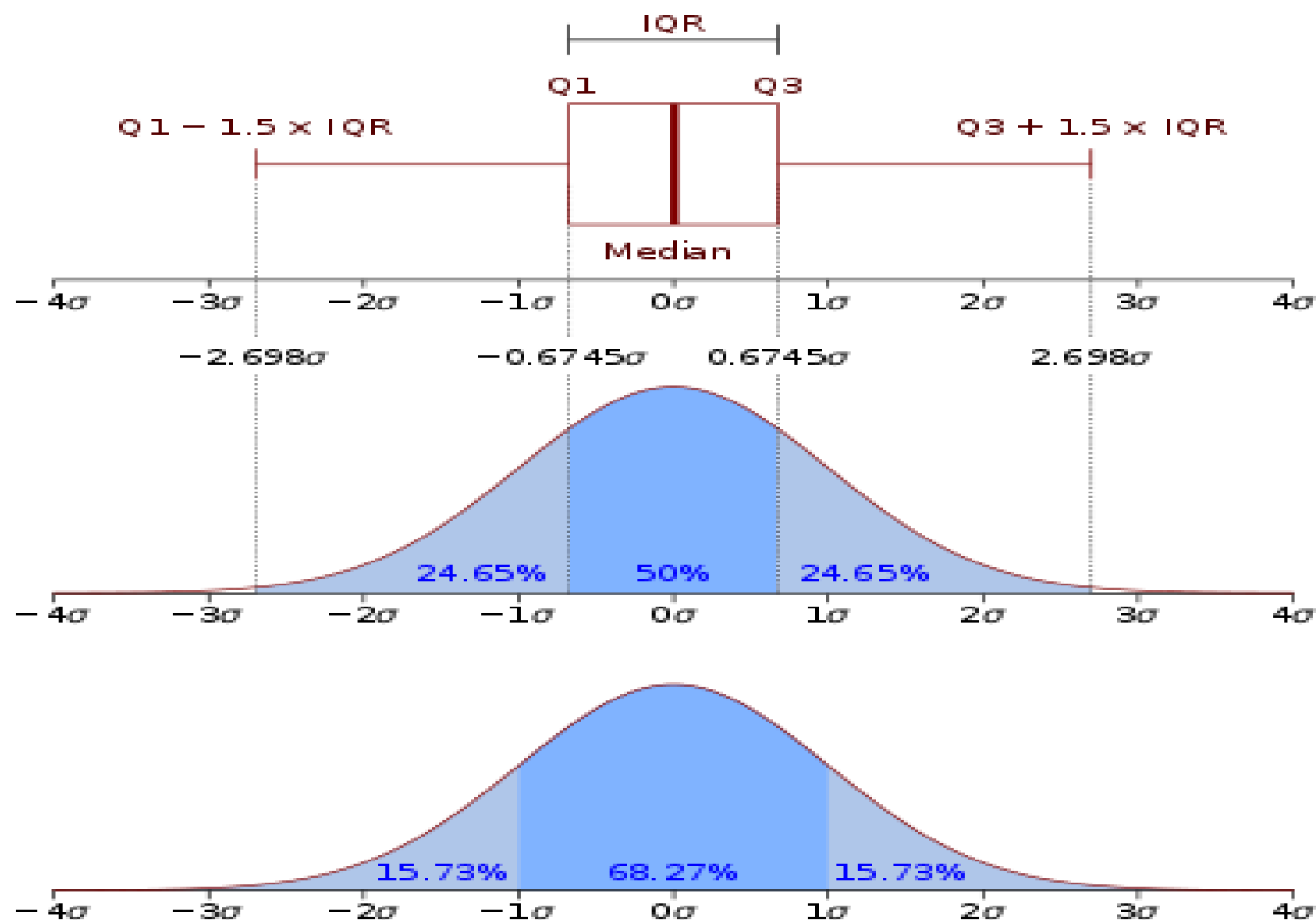
lower quartile

median

upper quartile



Box plot

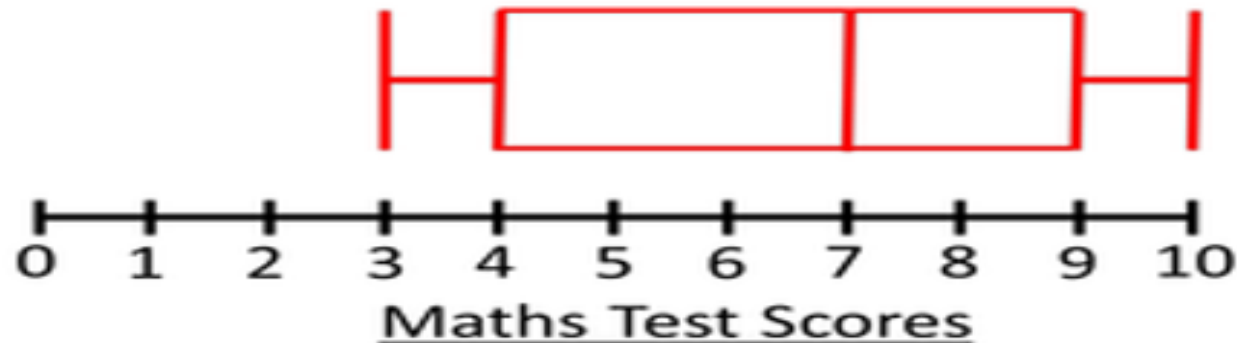


Box plot

- Here are the maths test results of 23 male students:

3, 3, 3, 3, 4, 4, 4, 5, 6, 6, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10

- Lower Quartile: $\frac{n+1}{4} = \frac{23+1}{4} = \frac{24}{4} = 6^{\text{th}} = 4$
- Median: $\frac{n+1}{2} = \frac{23+1}{2} = \frac{24}{2} = 12^{\text{th}} = 7$
- Upper Quartile: $3 \times \text{LQ} = 18^{\text{th}} = 9$



Summary

Statistical Methods

Mean

Median

Mode

STDEV

Variance

Univariate Analysis - Boxplot