0.1 Basic Statistics

We denote the number of data points by n.

Mean / Average: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Median / Second Quartile: The middle value when arranged from lowest to highest or vice versa. The *position* of the median is given by $Q_2 = \frac{1}{2}(n+1)$

Mode: The most commonly occurring term is called the *mode*. Note: It is possible to have two modes, three modes and so on. Such data is termed as *bi-modal*, *tri-modal* and so on, respectively

Range: The *range* is given by the highest value minus the lowest value

First Quartile: The *position* of the first quartile is given by $Q_1 = \frac{1}{4}(n+1)$. Note: The data must be arranged from lowest to highest or vice versa

Third Quartile: The *position* of the second quartile is given by $Q_3 = \frac{3}{4}(n+1)$. Note: The data must be arranged from lowest to highest or vice versa

Interquartile Range: The interquartile range is given by the *value* of the thrid quartile minus the *value* of the first quartile **Minimum Value:** The smallest value in the data is called the *minimum*

Maximum Value: The largest value in the data is called the *maximum*

Box-and-Whisker Diagram: A graphical rendition of the minimum value, first quartile, median (second quartile), thrid quartile and the maximum value. The first, second and third quartiles form the so called "box" of the plot

0.2 Variance (Sample and Population) and Standard Deviation (Sample and Population)

0.3 Linear Regression

Given a set of n data points (x_i, y_i), where x is the independent variable and y is the dependent variable. We note that the means are $\bar{x}$ and $\bar{y}$ respectively. The least squares regression line that fits the data is given by

$$\hat{y} = \hat{a} + \hat{b}x$$

$$\hat{b} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$