Statistics:
Descriptive Statistics

When we are given a large data set, it is necessary to describe the data in some way. The raw data is just too large and meaningless on its own to make sense of. We will sue the following exam scores data throughout:

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$$
$$45, 67, 87, 21, 43, 98, 28, 23, 28, 75$$

Summary Statistics

We can use the raw data to calculate summary statistics so that we have some idea what the data looks like and how sprad out it is.

Max, Min and Range

The maximum value of the dataset and the minimum value of the dataset are very simple measures. The range of the data is difference between the maximum and minimum value.

$$\begin{aligned} \text{Range } &= \text{Max Value } - \text{Min Value} \\ &= 98 - 21 = 77 \end{aligned}$$

Mean, Median and Mode

The mean, median and mode are measures of central tendency of the data (i.e. where is the center of the data).

Mean (μ)

The mean is sum of all values divided by how many values there are

$$\frac{1}{N}\sum_{i=1}^{N} x_i = \frac{45+67+87+21+43+98+28+23+28+75}{10} = 51.5$$

Median

The median is the middle data point when the dataset is arranged in order from smallest to largest. If there are two middle values then we take the average of the two values.
Using the data above check that the median is: 44

Mode

The mode is the value in the dataset that appears most.
Using the data above check that the mode is: 28

Standard Deviation

The standard deviation (σ) measures how spread out the data is. A simple formula for calculating the standard deviation is:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2}$$

The following example shows hoe the standard deviation can be calculated.

x_i	μ	$x_i - \mu$	$(x_i-\mu)^2$
45	51.5	-6.5	42.25
67	51.5	15.5	240.25
87	51.5	35.5	1260.25
21	51.5	-30.5	930.25
43	51.5	-8.5	72.25
98	51.5	46.5	2162.25
28	51.5	-23.5	552.25
23	51.5	-28.5	812.25
28	51.5	-23.5	552.25
75	51.5	23.5	552.25
			7176.5

$$\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N-1} = \frac{7176.5}{9} = 797.3889$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N-1}} = \sqrt{797.3889} = 28.238$$