Statistics

Types of Data

Data can be one of three types:

- categorical - the data fall into two or more categories with no natural order, such as hair colour
- ordinal - the data fall into two or more categories, but have a natural order, such as education level
- cardinal - the data count quantity. Cardinal data can be discrete (the number of items in a box) or continuous (the temperature measured by a thermometer)

Averages

There are two commonly-used types of average - the *mean* and the *median*.

The mean (calculated using equation (1) to the right) is the sum of the data points divided by the number of data points. It is sensitive to individual measurements, and so is useful for detecting outliers.

The median (found using equation (2) to the right) is the midpoint (or midway between the two midpoints) of the data points. It is less sensitive to individual measurements, and therefore useful for tracking trends.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}(x_1 + x_2 + x_3 + ... + x_n) \tag{1}$$

$$\tilde{x} = \begin{pmatrix} \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) \text{ if n is even} \\ x_{\frac{n+1}{2}} \text{ if n is odd} \end{pmatrix} \tag{2}$$

Averages and Constants

Averages respond intuitively to their data being modified by constants.

For some constant a, given the data $(x_1 + a, x_2 + a, x_3 + a, ..., x_n + a)$, the mean and median of the data will both increase by a.

Likewise, for some constant b, given the data $(bx_1, bx_2, bx_3, ..., bx_n)$, the mean and median of the data will both be multiplied by b.

Standard Deviation

Standard deviation measures the average distance between data points and the mean. It can be thought of as a measure of how wide a variance the data contains.

The symbol σ represents standard deviation.

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}[(x_i - \overline{x})^2]$$

Frequency Tables

Frequency tables compile data to show frequency of cardinal data readings. For instance, the table here records tyre tread depth as measured by car mechanics during MOTs.

The table can be used with adjusted versions of the mean and standard deviation equations.

Depth (mm)	Frequency (f_i)
1.6	11
1.8	22
2.0	26
2.2	40

Figure 1: Tyre Tread Depths

Mean and Standard Deviation

The equation (1) is used to find the mean of data in a frequency table. Using the data in Tyre Tread Depths, we can apply this to find a mean of $1.988mm$.

$$\overline{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} \tag{1}$$

$$\begin{aligned}\overline{x} =& \frac{\Sigma f_i x_i}{\Sigma f_i} \\ =& \frac{1.6 \cdot 12 + 1.8 \cdot 22 + 2.0 \cdot 26 + 2.2 \cdot 40}{11 + 22 + 26 + 40} \\ =& \frac{198.8}{100} \\ =& 1.988mm\end{aligned}$$

The equation (2) can be used to find the standard deviation of data in a frequency table. Again using the data in Tyre Tread Depths:

$$\sigma = \frac{\Sigma f_i (x_i - \overline{x})^2}{(\Sigma f_i) - 1} \tag{2}$$

$$\begin{aligned}\sigma =& \frac{\Sigma f_i (x_i - \overline{x})^2}{(\Sigma f_i) - 1} \\ =& \frac{11 \cdot (1.6 - 1.988)^2 + 22 \cdot (1.8 - 1.988)^2 + 26 \cdot (2.0 - 1.988)^2 + 40 \cdot (2.2 - 1.988)^2}{11 + 22 + 26 + 40 - 1} \\ =& 0.21\end{aligned}$$