Standard deviation

There are three commonly used measures of spread (or dispersion) – the range, the inter-quartile range and the **standard deviation**.

The standard deviation is widely used in statistics to measure spread. It is based on **all** the values in the data, so it is sensitive to the presence of outliers in the data.

The **variance** is related to the standard deviation:

$$\boxed{\text{variance} = (\text{standard deviation})^2}$$

The following formulae can be used to find the variance and s.d.

$$\boxed{\text{variance} = \frac{\sum (x_i - \bar{x})^2}{n}}$$

$$\boxed{\text{s.d.} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}}$$

Standard deviation

Example: The mid-day temperatures (in °C) recorded for one week in June were: 21, 23, 24, 19, 19, 20, 21

First we find the mean: $\bar{x} = \dfrac{21+23+\ldots+21}{7} = \dfrac{147}{7} = 21°\text{C}$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
21	0	0
23	2	4
24	3	9
19	-2	4
19	-2	4
20	-1	1
21	0	0
	Total:	**22**

$$\text{variance} = \frac{\sum (x_i - \bar{x})^2}{n}$$

So variance = 22 ÷ 7 = **3.143**

So, s.d. = **1.77°C** (3 s.f.)

Standard deviation

There is an **alternative formula** which is usually a more convenient way to find the variance:

$$\text{variance} = \frac{\sum (x_i - \bar{x})^2}{n}$$

But, $\sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$

$$= \sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2$$

$$= \sum x_i^2 - 2\bar{x} \times n\bar{x} + n\bar{x}^2$$

$$= \sum x_i^2 - n\bar{x}^2$$

Therefore, $\boxed{\text{variance} = \frac{\sum x_i{}^2}{n} - \bar{x}^2}$ and $\boxed{\text{s.d.} = \sqrt{\frac{\sum x_i{}^2}{n} - \bar{x}^2}}$

Standard deviation

Example (continued): Looking again at the temperature data for June: 21, 23, 24, 19, 19, 20, 21

We know that $\bar{x} = \frac{147}{7} = 21°\text{C}$

Also, $\sum x_i^{\,2} = 21^2 + 23^2 + \ldots + 21^2$ $= 3109$

So, $\text{variance} = \frac{\sum x_i^{\,2}}{n} - \bar{x}^2 = \frac{3109}{7} - 21^2 = 3.143$

$\text{s.d.} = 1.77\,°\text{C}$

Note: Essentially the standard deviation is a measure of how close the values are to the mean value.

Calculating standard deviation from a table

When the data is presented in a frequency table, the formula for finding the standard deviation needs to be adjusted slightly:

$$\text{s.d.} = \sqrt{\frac{\sum f_i \times x_i^{\ 2}}{\sum f_i} - \overline{x}^2}$$

Example: A class of 20 students were asked how many times they exercise in a normal week.

Find the mean and the standard deviation.

Number of times exercise taken	Frequency
0	5
1	3
2	5
3	4
4	2
5	1

Calculating standard deviation from a table

No. of times exercise taken, x	Frequency, f	$x \times f$	$x^2 \times f$
0	5	0	0
1	3	3	3
2	5	10	20
3	4	12	36
4	2	8	32
5	1	5	25
TOTAL:	**20**	**38**	**116**

The table can be extended to help find the mean and the s.d.

$$\bar{x} = \frac{38}{20} = 1.9 \qquad \text{s.d.} = \sqrt{\frac{\sum f_i \times x_i{}^2}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{116}{20} - 1.9^2} = 1.48$$

Worked example

Find an estimate of the variance and standard deviation of the following data for the marks obtained in a test by 88 students.

Marks (x)	$0 \le x < 10$	$10 \le x < 20$	$20 \le x < 30$	$30 \le x < 40$	$40 \le x < 50$
Frequency (f)	6	16	24	25	17

Marks	**Mid Interval Value (x)**	f	fx	x^2	fx^2
$0 \le x < 10$	5	6	30	25	150
$10 \le x < 20$	15	16	240	225	3 600
$20 \le x < 30$	25	24	600	625	15 000
$30 \le x < 40$	35	25	875	1 225	30 625
$40 \le x < 50$	45	17	765	2 025	34 425
Total		88	2 510		83 800

$$\text{Mean } \bar{x} = \frac{\sum fx}{n} = \frac{2510}{88}$$

$$\text{Variance } \sigma^2 = \frac{\sum fx^2}{n} - \bar{x}^2$$

$$= \frac{83800}{88} - \left(\frac{2510}{88}\right)^2$$

$$= 952.273 - 813.546 = 138.727$$

$$= 138.73 \text{ (2 dp)}$$

$$\text{Standard deviation} = \sqrt{138.727}$$

$$= 11.78$$