

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:

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

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

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

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



# 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:  $\bar{x} = \frac{21+23+\dots+21}{7} = \frac{147}{7} = 21^{\circ}\text{C}$

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
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 =  $22 \div 7 = 3.143$

So, s.d. =  $1.77^{\circ}\text{C}$  (3 s.f.)



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

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

$$\begin{aligned} \text{But, } \sum (x_i - \bar{x})^2 &= \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \times n\bar{x} + n\bar{x}^2 \\ &= \sum x_i^2 - n\bar{x}^2 \end{aligned}$$

Therefore,

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

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



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

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

Also,  $\sum x_i^2 = 21^2 + 23^2 + 24^2 + 19^2 + 19^2 + 20^2 + 21^2 = 3109$

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

s.d. =  $1.77^{\circ}\text{C}$

**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:

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

**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.

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



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

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



$$\begin{aligned}\text{Mean } \bar{x} &= \frac{\sum fx}{n} \\ &= \frac{2510}{88}\end{aligned}$$

$$\begin{aligned}\text{Variance } \sigma^2 &= \frac{\sum fx^2}{n} - \bar{x}^2 \\ &= \frac{83800}{88} - \left(\frac{2510}{88}\right)^2 \\ &= 952.273 - 813.546 = 138.727 \\ &= 138.73 \text{ (2 dp)}\end{aligned}$$

$$\begin{aligned}\text{Standard deviation} &= \sqrt{138.727} \\ &= 11.78\end{aligned}$$

