Variance and standard deviation (grouped data)

Introduction

In this leaflet we extend the definitions of variance and standard deviation to data which has been grouped.

Variance

The **variance** of a set of values, which we denote by σ^2, is defined as

$$\sigma^2 = \frac{\sum f(x - \bar{x})^2}{n}$$

where $\bar{x}$ is the mean, x stands for each data value in turn, and f is the frequency with which data value, x, occurs. Note that $\sum f = n$.

An alternative, yet equivalent formula, which is often easier to use is

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

$$\text{where} \quad \bar{x} = \sum \frac{fx}{n} \quad \text{(The Mean)}$$

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

We can show the calculations in a table as follows:

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

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

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

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

$$= 952.273 - 813.546 = 138.727$$

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

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

$$= 11.78$$

Exercises

Find an estimate of the standard deviation of the following:

1.

a) **Lifetime (hours) of components**	**Frequency**	b) **Income (1000's £)**	**Frequency**	c) **House prices (1000's £)**	**Frequency**
$300 \leq l < 400$	13	$10 \leq i < 15$	9	$40 \leq p < 60$	5
$400 \leq l < 500$	25	$15 \leq i < 20$	16	$60 \leq p < 80$	9
$500 \leq l < 600$	66	$20 \leq i < 25$	22	$80 \leq p < 100$	15
$600 \leq l < 700$	58	$25 \leq i < 30$	8	$100 \leq p < 120$	8
$700 \leq l < 800$	38	$30 \leq i < 35$	5	$120 \leq p < 140$	3

2. The grouped frequency table shows the length of service in years of employees who have been working for a company for at least ten years.

Calculate an estimate of the standard deviation of the length of service of these employees.

Length of Service (x)	$10 \leq x < 15$	$15 \leq x < 20$	$20 \leq x < 25$	$25 \leq x < 30$	$30 \leq x < 40$	$40 \leq x < 50$
Frequency (f)	30	42	23	13	8	4

Answers

1. a) 112.37 hours   b) £5 617   c) £21 994

2. 7.70 years (2 dp)

Note that the units of standard deviation are the same as the original data units. This is not true of the variance where the units would be the square of the original data units.