Paper 6 – Topic 1 - REPRESENTATION OF DATA

Average/Measure of Central Tendency

An average value is useful when describing a set of data. This is a typical or representative value and is known as a measure of central tendency.

One such average is the mean.

The mean

The mean is the most commonly used average. It is calculated by dividing the sum of all the observations by the number of observation

Example 1

The members of an orchestra were asked how many instruments each could play. These are their replies.

2 5 2 4 1 1 1 2 1 3 3 2 1 2 1
1 2 4 3 2 1 2 3 1 4 2 3 1 1 2

Calculate the mean number of instruments played.

or

Number of instruments, x	1	2	3	4	5
Frequency, f	11	10	5	3	1

Calculate the mean number of instruments played.

Example 2

In a spot check, the speeds of 120 vehicles on a particular stretch of road through a village were noted. The results are shown in the table.

Speed x km/h	21–25	26–30	31–35	36–45	46–60
Frequency f	22	48	25	16	9

Estimate the mean speed of these vehicles.

Example 3

The following table shows the results of a survey to find the average daily time, in hours, that people spent watching television.

Time spent per day (t hours)	$0 \leqslant t < 1$	$1 \leqslant t < 2$	$2 \leqslant t < 4$	$4 \leqslant t < 8$
Number of people	10	18	f	4

An estimate of the mean time was calculated to be 2 hours.

Form an equation involving f and hence find the number of people in the survey.

Measure of Spread

Standard Deviation

Is a very useful measure of spread. It gives a measure of the spread of the data in relation to the mean, $\bar{x}$, of the distribution. It is calculated using all the values in the distribution.

Standard deviation is very useful for comparing distributions: the **lower** the standard deviation, the **less variation** there is and the **more consistent** the data are.

Example 4

The mean of the numbers 2, 3, 5, 6, 8 is 4.8. Calculate the standard deviation.

Method 1: Definition version		
x	$x - 4.8$	$(x - 4.8)^2$
2	−2.8	7.84
3	−1.8	3.24
5	0.2	0.04
6	1.2	1.44
8	3.2	10.24
		$\sum(x - \bar{x})^2 = 22.8$

$$\text{s.d.} = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$
$$= \sqrt{\frac{22.8}{5}}$$
$$= 2.14 \text{ (3 s.f.)}$$

Method 2: Calculation version	
x	x^2
2	4
3	9
5	25
6	36
8	64
	$\sum x^2 = 138$

$$\text{s.d.} = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$
$$= \sqrt{\frac{138}{5} - 4.8^2}$$
$$= 2.14 \text{ (3 s.f.)}$$

Example 5

An online test was taken by 115 students. The time spent on each question was recorded by the computer. The following table shows the time taken, in minutes, on the final question.

Time (mins)	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 5$	$5 \leqslant x < 10$
Frequency	16	32	42	25

Calculate estimates of the mean and standard deviation of the time spent on the final question.

The formula:

For raw data:

$$\bar{x} = \frac{\sum x}{n} \qquad \text{standard deviation} = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Use the calculation versions.

For data in a frequency table:

$$\bar{x} = \frac{\sum xf}{\sum f} \qquad \text{standard deviation} = \sqrt{\frac{\sum(x - \bar{x})^2 f}{\sum f}} = \sqrt{\frac{\sum x^2 f}{\sum f} - \bar{x}^2}$$

Variance

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

Combining Sets of Data

When you are not given the actual data but instead are given totals such as $\sum x$ and $\sum x^2$, you will need to substitute them into the formulae for the mean and the standard deviation. This is particularly useful when combining sets of data.

Example 6

The ages, x years, of 18 people attending an evening class are summarised by the following totals:

$$\sum x = 745, \sum x^2 = 33\,951.$$

(i) Calculate the mean and standard deviation of the ages of this group of people.

(ii) One person leaves the group and the mean age of the remaining 17 people is exactly 41 years. Find the age of the person who left and the standard deviation of the ages of the remaining 17 people.

Cambridge Paper 6 Q4 N04

(i) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{745}{18} = 41.388\ldots = 41.4$ years (3 s.f.)

s.d. $= \sqrt{\dfrac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\dfrac{33\,951}{18} - (41.38\ldots)^2} = 13.157\ldots = 13.2$ years (3 s.f.)

(ii) For 17 people:

Mean = 41, so $\sum x = 41 \times 17 = 697$

Previous total = 745, so age of leaver = $745 - 697 = 48$

New $\sum x^2 = 33\,951 - 48^2 = 31\,647$

New standard deviation

$$= \sqrt{\frac{31\,647}{17} - 41^2} = 13.43\ldots = 13.4 \text{ years (3 s.f.)}$$

Example 7

The following table shows the mean and standard deviation of the heights of 20 boys and 30 girls.

	Mean	Standard deviation
Boys	160 cm	4 cm
Girls	155 cm	3.5 cm

Find the mean and standard deviation of the heights of the 50 children.

[157, 4.44]

Scaling Data

Example 8

Sweets are packed into bags with a nominal weight of 75 grams. Ten bags are picked at random from the production line and weighed. Their weights, in grams, are

76.0, 74.2, 75.1, 73.7, 72.0, 74.3, 75.4, 74.0, 73.1, 72.8

(i) Use your calculator to find the mean and the standard deviation.

(ii) It is later discovered that the scales were reading 3.2 grams below the correct weight
   (a) What was the correct mean weight of the 10 bags?
   (b) What was the correct standard deviation of the 10 bags?

(i) Using a calculator, you should find that
Mean = 74.06 grams, standard deviation = 1.166... = 1.17 grams (3 s.f.)

(ii) The correct readings are
79.2, 77.4, 78.3, 76.9, 75.2, 77.5, 78.6, 77.2, 76.3, 76.0
   (a) Correct mean = 77.26 grams
   (b) Correct standard deviation = 1.166... = 1.17 grams (3 s.f.)

Note:

In general, if each data value is increased by a constant a
- the mean is increased by a
- the standard deviation is unaltered.

This is particularly useful when finding the mean and standard deviation using $\sum(x - a)$ and $\sum(x - a)^2$, where a is a constant.

Example 9

A summary of 24 observations of x gave the following information:

$$\sum(x - a) = -73.2 \quad \text{and} \quad \sum(x - a)^2 = 2115$$

The mean of these values of x is 8.95.

(i) Find the value of the constant a.

(ii) Find the standard deviation of these values of x. Cambridge Paper 6 Q1 N07