

## AS Mathematics Exam Questions by Topic Chapter 15a: Averages, spread, boxplots, outliers

---

These questions are taken from the Specimen Exam materials and the real 2018 papers for the new syllabus AS and A-level mathematics courses and arranged by chapter of the textbooks by Goldie et al (available here: <https://amzn.to/39umfr5> and <https://amzn.to/3hE8kBL> ). There are a mixture of questions from OCR A, OCR B (MEI), Edexcel and AQA. Although the style of questions varies a little across the exam boards the content of the syllabus is almost identical so these are suitable for students preparing for any exam board.

Free problem sets for all other chapters, as well as video solutions, full past papers and other content for GCSE and A-level maths can be found at:

<https://mathsaurus.com/>

---

### OCR B MEI 2018 Paper 2 Question 4:

- 4 A survey of the number of cars per household in a certain village generated the data in Fig. 4.

Number of cars	0	1	2	3	4
Number of households	8	22	31	27	7

**Fig. 4**

- (i) Calculate the mean number of cars per household. [1]
- (ii) Calculate the standard deviation of the number of cars per household. [1]
-

---

OCR B MEI 2018 Paper 2 Question 9:

- 9 At the end of each school term at North End College all the science classes in year 10 are given a test. The marks out of 100 achieved by members of set 1 are shown in Fig. 9.

3	5
4	0 9
5	2 3 6
6	0 1 3 5 6
7	0 1 2 5 6 8 9 9
8	3 4 6 6 8 8 9
9	5 5 5 6 7

Key 5 | 2 represents a mark of 52

**Fig. 9**

- (i) Describe the shape of the distribution. [1]
- (ii) The teacher for set 1 claimed that a typical student in his class achieved a mark of 95. How did he justify this statement? [1]
- (iii) Another teacher said that the average mark in set 1 is 76. How did she justify this statement? [1]
- Benson's mark in the test is 35. If the mark achieved by any student is an outlier in the lower tail of the distribution, the student is moved down to set 2.
- (iv) Determine whether Benson is moved down to set 2. [2]

---

OCR B MEI AS 2018 Paper 2 Question 7:

- 7 Rose and Emma each wear a device that records the number of steps they take in a day. All the results for a 7-day period are given in Fig. 7.

Day	1	2	3	4	5	6	7
Rose	10014	11262	10149	9361	9708	9921	10369
Emma	9204	9913	8741	10015	10261	7391	10856

**Fig. 7**

The 7-day mean is the mean number of steps taken in the last 7 days. The 7-day mean for Rose is 10 112.

- (i) Calculate the 7-day mean for Emma. [1]

At the end of day 8 a new 7-day mean is calculated by including the number of steps taken on day 8 and omitting the number of steps taken on day 1. On day 8 Rose takes 10 259 steps.

- (ii) Determine the number of steps Emma must take on day 8 so that her 7-day mean at the end of day 8 is the same as for Rose. [4]

In fact, over a long period of time, the mean of the number of steps per day that Emma takes is 10 341 and the standard deviation is 948.

- (iii) Determine whether the number of steps Emma needs to take on day 8 so that her 7-day mean is the same as that for Rose in part (ii) is unusually high. [3]

---

OCR B MEI AS Sample Paper 2 Question 7:

- 7 An apple farmer has 200 apple trees. She is investigating the masses of the crops of apples from individual trees. She decides to select a sample of these trees and find the mass of the crop for each tree.

(i) Explain how she can select a random sample of 10 different trees from the 200 trees. [2]

The masses of the crops from the 10 trees, measured in kg, are recorded as follows.

23.5 27.4 26.2 29.0 25.1 27.4 26.2 28.3 38.1 24.9

- (ii) For these data find
- the mean,
  - the sample standard deviation. [2]

(iii) Show that there is one outlier at the upper end of the data. How should the farmer decide whether to use this outlier in any further analysis of the data? [3]

---

OCR B MEI Sample Paper 2 Question 8:

- 8 Alison selects 10 of her male friends. For each one she measures the distance between his eyes. The distances, measured in mm, are as follows:

51 57 58 59 61 64 64 65 67 68

The mean of these data is 61.4. The sample standard deviation is 5.232, correct to 3 decimal places.

One of the friends decides he does not want his measurement to be used. Alison replaces his measurement with the measurement from another male friend. This increases the mean to 62.0 and reduces the standard deviation. Give a possible value for the measurement which has been removed and find the measurement which has replaced it. [3]

---

AQA AS 2018 Paper 2 Question 14:

- 14 Given that  $\sum x = 364$ ,  $\sum x^2 = 19412$ ,  $n = 10$ , find  $\sigma$ , the standard deviation of  $X$ .

Circle your answer.

[1 mark]

24.8                      44.1                      616.2                      1941.2

---

AQA AS 2018 Paper 2 Question 17:

- 17 The table below is an extract from the Large Data Set, showing the purchased quantities of fats and oils for the South East of England in 2014.

Description	Purchased quantity
Butter	42
Soft margarine	16
Olive oil	17
Other vegetable and salad oils	28

Kim claims that more olive oil was purchased in the South East than soft margarine.

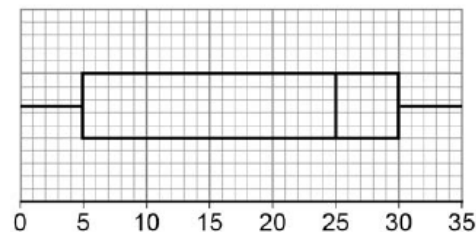
Explain why Kim may be incorrect.

[2 marks]

---

AQA AS Sample Paper 2 Question 16:

- 16 The boxplot below represents the time spent in hours by students revising for a history exam.



- 16 (a) Use the information in the boxplot to state the value of a measure of central tendency of the revision times, stating clearly which measure you are using.

[1 mark]

---

---

---

- 16 (b) Use the information in the boxplot to explain why the distribution of revision times is negatively skewed.

[1 mark]

17 The table below is an extract from the Large Data Set.

Make	Region	Engine size	Mass	CO <sub>2</sub>	CO
VAUXHALL	South West	1398	1163	118	0.463
VOLKSWAGEN	London	999	1055	106	0.407
VAUXHALL	South West	1248	1225	85	0.141
BMW	South West	2979	1635	194	0.139
TOYOTA	South West	1995	1650	123	0.274
BMW	South West	2979	0	244	0.447
FORD	South West	1596	0	165	0.518
TOYOTA	South West	1299	1050	144	
VAUXHALL	London	1398	1361	140	0.695
FORD	North West	4951	1799	299	0.621

17 (a) (i) Calculate the standard deviation of the engine sizes in the table.

[1 mark]

---

---

---

17 (a) (ii) The mean of the engine sizes is 2084

Any value more than 2 standard deviations from the mean can be identified as an outlier.

Using this definition of an outlier, show that the sample of engine sizes has **exactly one** outlier.

Fully justify your answer.

[3 marks]

17 (b) Rajan calculates the mean of the masses of the cars in this extract and states that it is 1094 kg.

Use your knowledge of the Large Data Set to suggest what error Rajan is likely to have made in his calculation.

[1 mark]

17 (c) Rajan claims there is an error in the data recorded in the table for one of the Toyotas from the South West, because there is no value for its carbon monoxide emissions.

Use your knowledge of the Large Data Set to comment on Rajan's claim.

[1 mark]

Edexcel 2018 Paper 3 Question 4:

4. Charlie is studying the time it takes members of his company to travel to the office. He stands by the door to the office from 08 40 to 08 50 one morning and asks workers, as they arrive, how long their journey was.

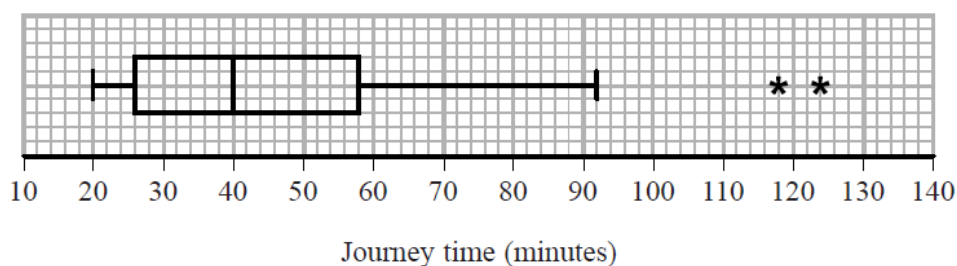
(a) State the sampling method Charlie used. (1)

(b) State and briefly describe an alternative method of non-random sampling Charlie could have used to obtain a sample of 40 workers. (2)

Taruni decided to ask every member of the company the time,  $x$  minutes, it takes them to travel to the office.

(c) State the data selection process Taruni used. (1)

Taruni's results are summarised by the box plot and summary statistics below.



$$n = 95 \quad \sum x = 4133 \quad \sum x^2 = 202294$$

(d) Write down the interquartile range for these data. (1)

(e) Calculate the mean and the standard deviation for these data. (3)

(f) State, giving a reason, whether you would recommend using the mean and standard deviation or the median and interquartile range to describe these data. (2)

Rana and David both work for the company and have both moved house since Taruni collected her data.

Rana's journey to work has changed from 75 minutes to 35 minutes and David's journey to work has changed from 60 minutes to 33 minutes.

Taruni drew her box plot again and only had to change two values.

(g) Explain which two values Taruni must have changed and whether each of these values has increased or decreased. (3)

---

Edexcel AS 2018 Paper 2 Question 4:

4. Helen is studying the daily mean wind speed for Camborne using the large data set from 1987. The data for one month are summarised in Table 1 below.

<b>Windspeed</b>	n/a	6	7	8	9	11	12	13	14	16
<b>Frequency</b>	13	2	3	2	2	3	1	2	1	2

**Table 1**

- (a) Calculate the mean for these data. (1)
- (b) Calculate the standard deviation for these data and state the units. (2)

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

<b>Month</b>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<b>Mean</b>	7.58	8.26	8.57	8.57	11.57
<b>Standard Deviation</b>	2.93	3.89	3.46	3.87	4.64

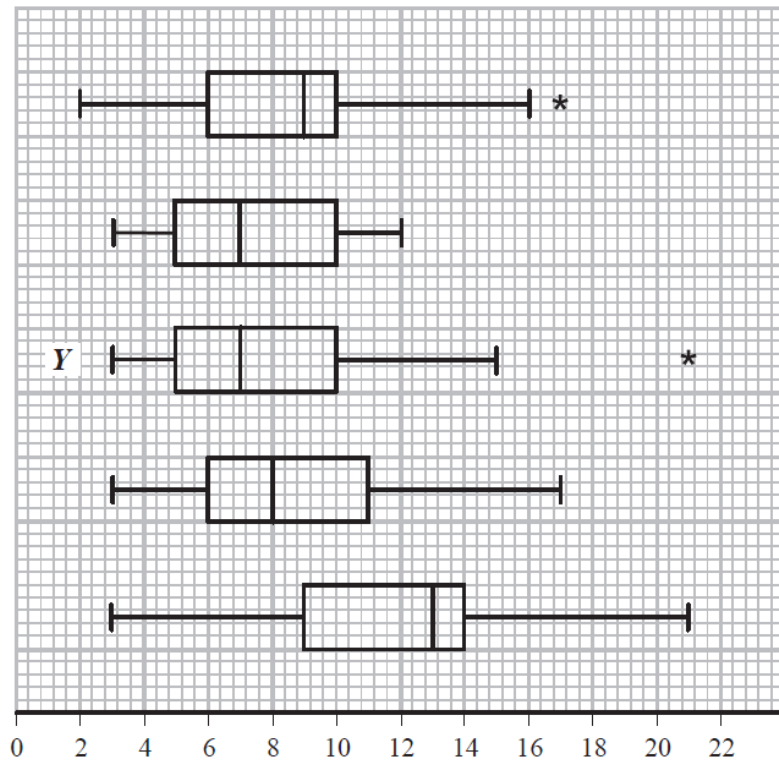
**Table 2**

- (c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57 (2)

The data for these months are summarised in the box plots on the opposite page. They are not in month order or the same order as in Table 2.

- (d) (i) State the meaning of the \* symbol on some of the box plots.
- (ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*. (3)

Question 4 continued




---

Edexcel AS Sample Paper 2 Question 1:

1. Sara is investigating the variation in daily maximum gust,  $t$  kn, for Camborne in June and July 1987.

She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

- (a) State the sampling technique Sara used. (1)

- (b) From your knowledge of the large data set explain why this process may not generate a sample of size 20. (1)

The data Sara collected are summarised as follows

$$n = 20 \quad \sum t = 374 \quad \sum t^2 = 7600$$

- (c) Calculate the standard deviation. (2)
-