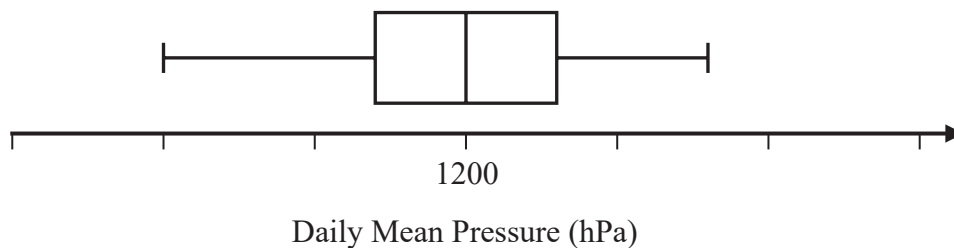


1.

Jiang is studying the variable Daily Mean Pressure from the large data set.

He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

- he failed to label all the values on the scale
- he gave an incorrect value for the median



Using your knowledge of the large data set, suggest a suitable value for

(a) the median,

(1)

(b) the range.

(1)

(You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers.)

2. 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.

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

Table 1

- Calculate the mean for these data. (1)
- 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.

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

Table 2

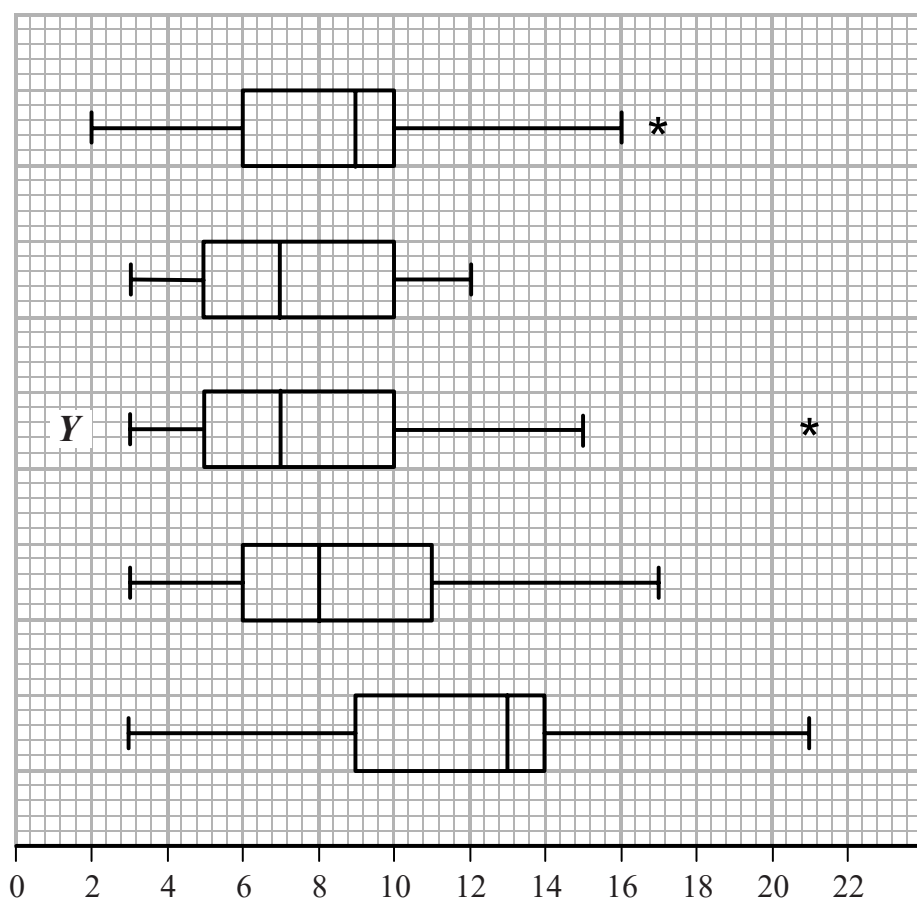
- (c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

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)

[illegible]

Question 2 continued



3. Sara was studying the relationship between rainfall, r mm, and humidity, h %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1 \quad Q_2 = 0.9 \quad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.

- (a) Show that $r = 20.6$ is an outlier.

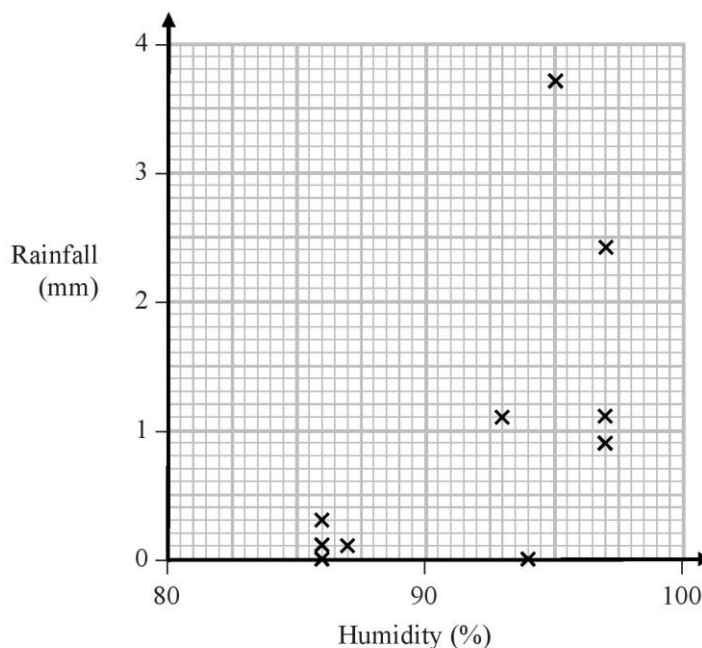
(1)

- (b) Give a reason why Sara might (i) include
(ii) exclude

this day's reading.

(2)

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of r and h .



- (c) Give an interpretation of the correlation between rainfall and humidity.

(1)

Question 3 continued

The equation of the regression line of r on h for these 10 days is $r = -12.8 + 0.15h$

- (d) Give an interpretation of the gradient of this regression line. (1)
- (e) (i) Comment on the suitability of Sara's sampling method for this study.
- (ii) Suggest how Sara could make better use of the large data set for her study. (2)

This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.