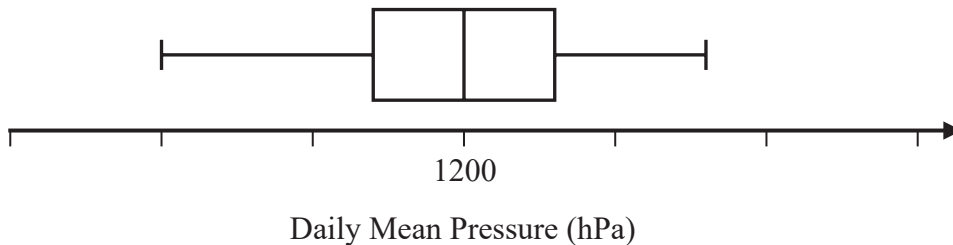


**1.**

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. Fred and Nadine are investigating whether there is a linear relationship between Daily Mean Pressure,  $p$  hPa, and Daily Mean Air Temperature,  $t$  °C, in Beijing using the 2015 data from the large data set.

Fred randomly selects one month from the data set and draws the scatter diagram in Figure 1 using the data from that month.

The scale has been left off the horizontal axis.

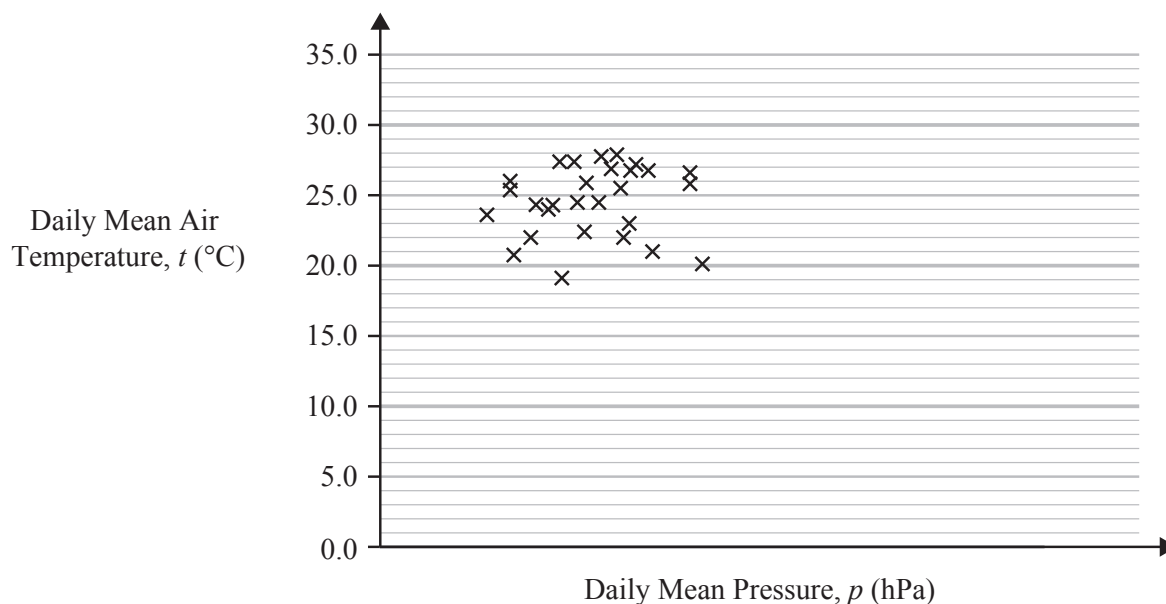


Figure 1

- (a) Describe the correlation shown in Figure 1.

(1)

Nadine chooses to use all of the data for Beijing from 2015 and draws the scatter diagram in Figure 2.

She uses the same scales as Fred.

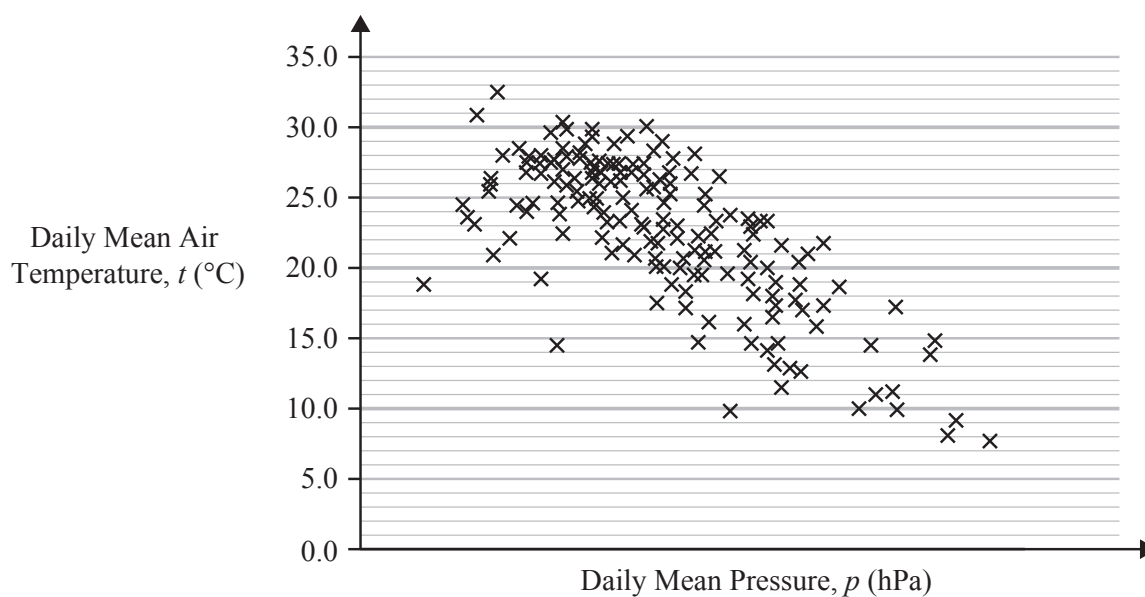


Figure 2

### Question 2 continued

- (b) Explain, in context, what Nadine can infer about the relationship between  $p$  and  $t$  using the information shown in Figure 2. (1)
- (c) Using your knowledge of the large data set, state a value of  $p$  for which interpolation can be used with Figure 2 to predict a value of  $t$ . (1)
- (d) Using your knowledge of the large data set, explain why it is not meaningful to look for a linear relationship between Daily Mean Wind Speed (Beaufort Conversion) and Daily Mean Air Temperature in Beijing in 2015. (1)

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

3. Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

<b>Outcome</b>	<i>A</i>	<i>B</i>	<i>C</i>
<b>Frequency</b>	16	2	1

- (a) State the sampling technique Helen used.

(1)

- (b) From your knowledge of the large data set

(i) suggest which variable was being studied,

(ii) state the name of outcome *A*.

(2)

George is also studying the same variable from the large data set for Heathrow from 2015. He started with the data from 5th May and then took every 10th reading and obtained the following

<b>Outcome</b>	<i>A</i>	<i>B</i>	<i>C</i>
<b>Frequency</b>	16	1	1

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following

<b>Outcome</b>	<i>A</i>	<i>B</i>	<i>C</i>
<b>Frequency</b>	155	26	3

- (c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015.

(1)

---

---

---

---

---

---

---

4. Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

Date	Daily Maximum Relative Humidity	Daily Mean Visibility
Units	%	
10/06/1987	90	5300
28/06/1987	100	0

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

- (a) write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility. (1)

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

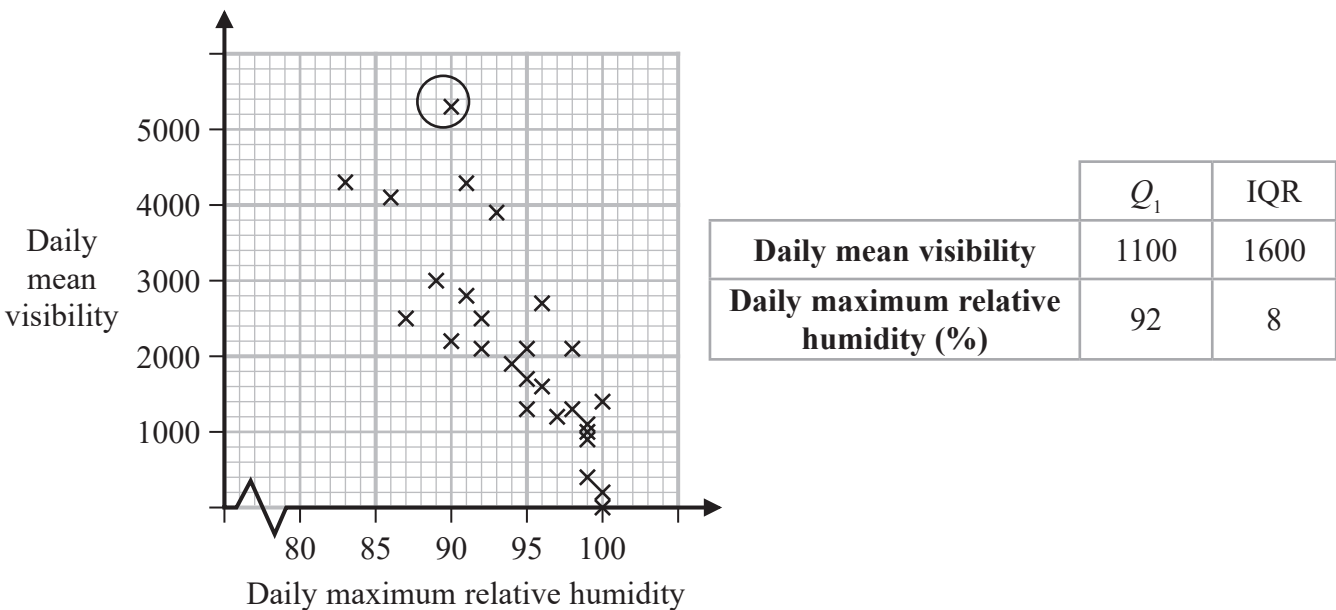
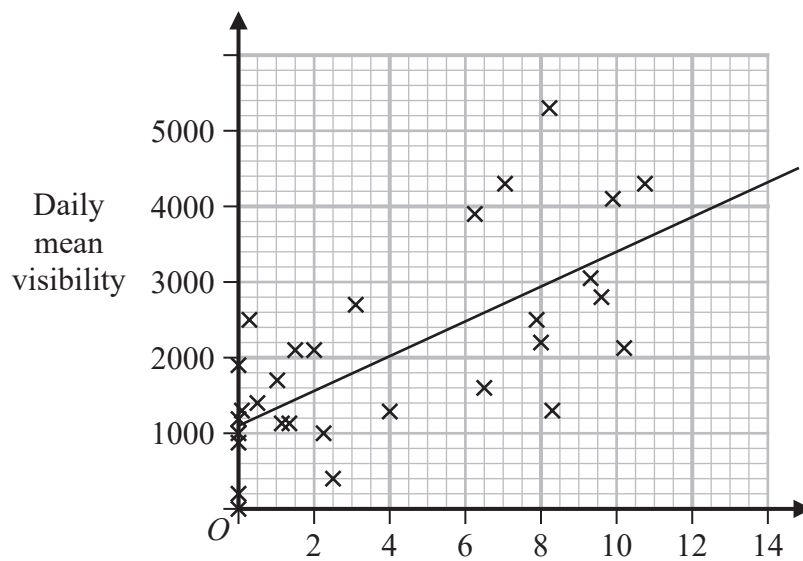


Figure 2

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above  $Q_3$  or more than 1.5 times the interquartile range below  $Q_1$ .

- (b) Show that the point circled on the scatter diagram is an outlier for visibility. (2)
- (c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity. (1)

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the  $x$ -axis.



### Figure 3

- (d) Using your knowledge of the large data set, suggest which variable the  $x$ -axis on this scatter diagram represents.

(1)

5. 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

- 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

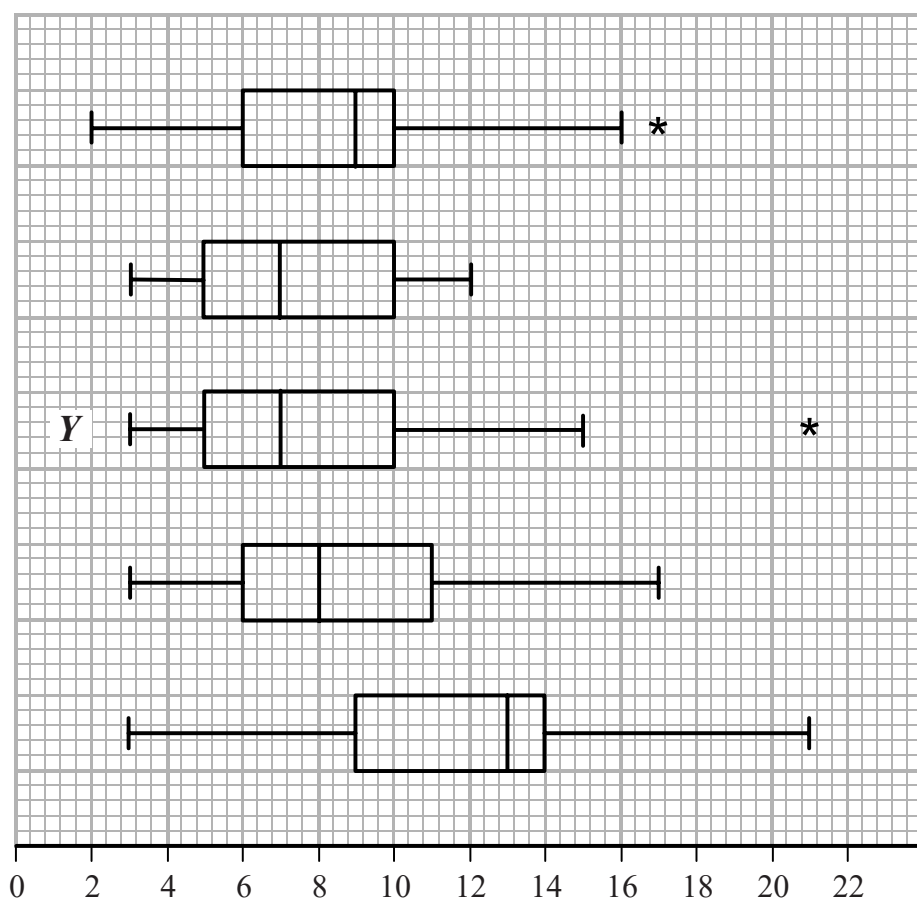
- (c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

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 5 continued





6.

Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

- (a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall ( $r$ mm)	Frequency	Midpoint ( $x$ mm)
$0 \leq r < 0.5$	121	0.25
$0.5 \leq r < 1.0$	10	0.75
$1.0 \leq r < 5.0$	24	3.0
$5.0 \leq r < 10.0$	12	7.5
$10.0 \leq r < 30.0$	17	20.0

You may use  $\sum fx = 539.75$  and  $\sum fx^2 = 7704.1875$

- (b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

- (c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

- (d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

- (ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

- (iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

---

---

---

---

---

---

---

---

