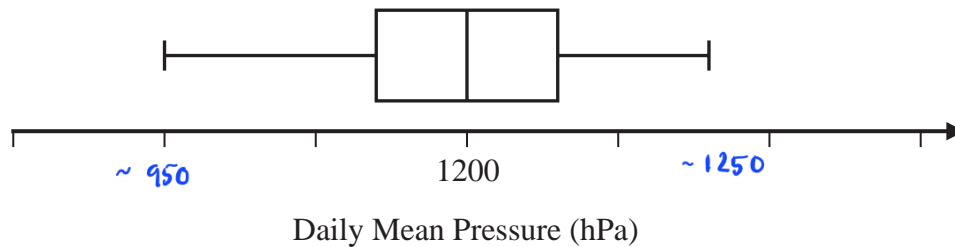


1. Jiang is studying the variable Daily Mean Pressure from the large data set.

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, For this sort of questions, the only way is to familiarise yourself with range of values for various data sets- (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.)

a) 1000 hPa (1)

b) 30 hPa (1)

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.

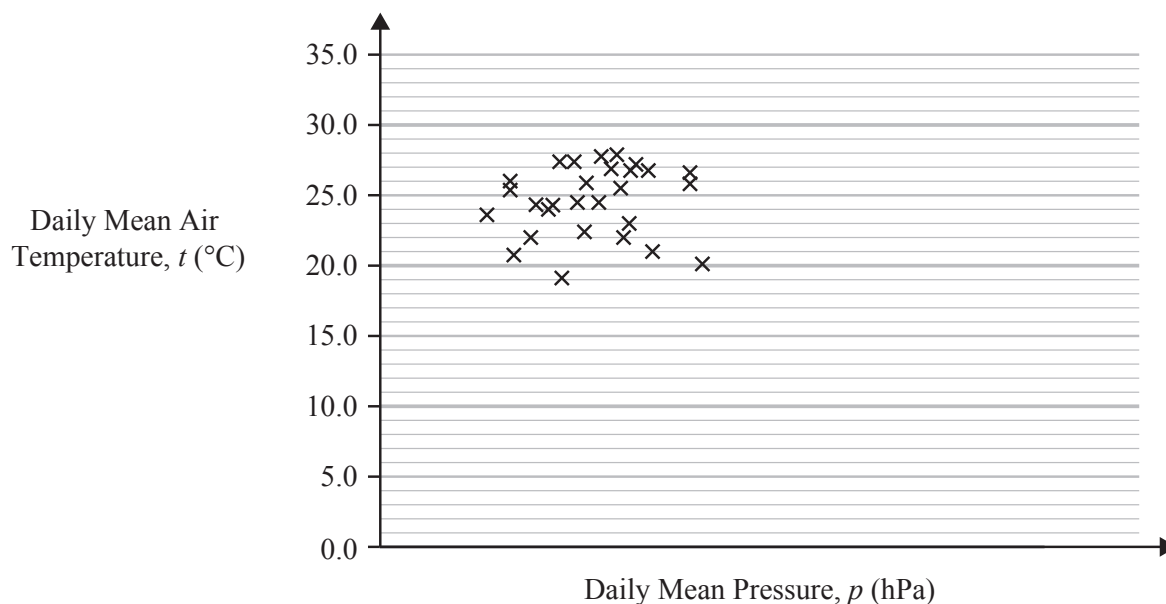


Figure 1

- (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.

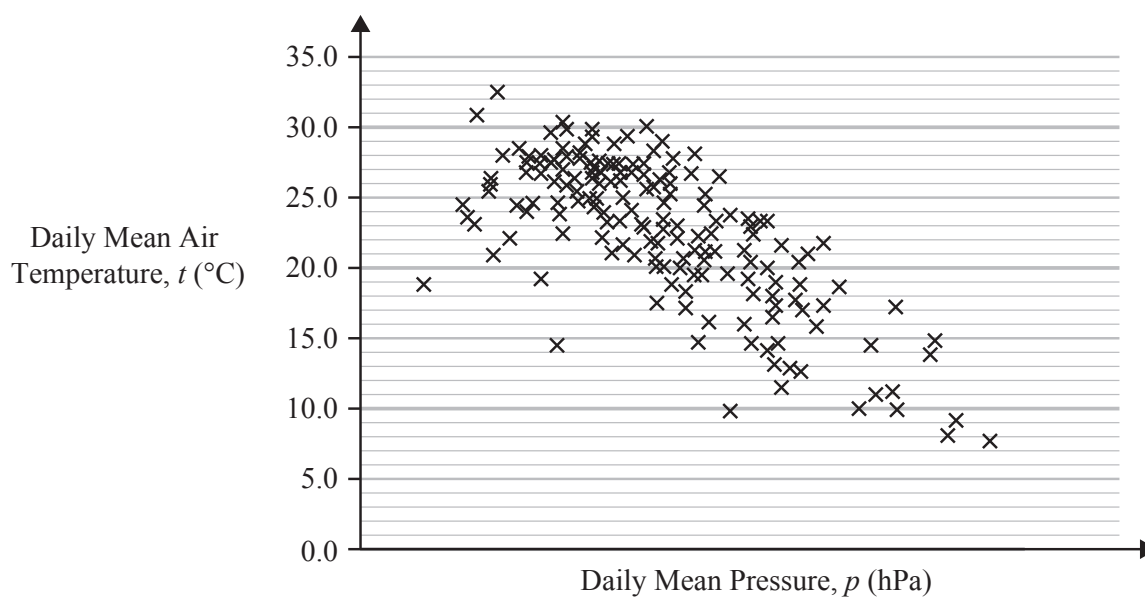


Figure 2

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)

a) No correlation (1)

b) As air pressure increases, temperature decreases
(Negative correlation) (1)

c) 1010 (1)

for Beijing 2015, air pressure ranges from 990hPa to 1040hPa, so any value between these is suitable.

d) Daily mean wind speed (Beaufort Conversion) is a qualitative variable. (1)

(Total for Question 2 is 4 marks)

3. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x , using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \quad \sum y^2 = 5912$$

(a) State the **units** of the variable x (1)

(b) Find the **mean** Daily Mean Pressure for these 30 days. (2)

(c) Find the **standard deviation** of Daily Mean Pressure for these 30 days. (3)

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction	NE	E	W (1)

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

entire pressure system
→

You may assume that these 3 locations were under a single region of pressure.

(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer. (2)

$$1 \text{ hPa} = 100 \text{ Pa} \quad , \quad 1 \text{ Pa} = 1 \text{ Nm}^{-2}$$

a) hPa (1)

b) $x = y + 1010$

$$\therefore \bar{x} = \bar{y} + 1010 \quad (1) \quad \leftarrow \bar{x} = \text{mean value of } x$$

$$= \frac{214}{30} + 1010$$

$$= 1017.1333 = 1017 \text{ hPa} \quad (1)$$

Question 3 continued.

c) $\sigma_x = \sigma_y$ standard deviation is not affected by this type of coding. (1)

$$= \sqrt{\frac{\sum y^2}{n} - \bar{y}^2}$$

$$= \sqrt{\frac{5912}{30} - \left(\frac{214}{30}\right)^2}$$

$$= \sqrt{146.1822} \quad (1)$$

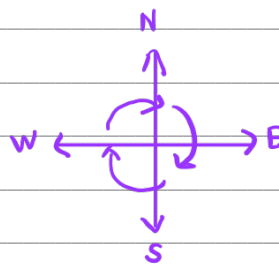
$$= 12.0905\dots = 12.1 \text{ (3 s.f.)} \quad (1)$$

d) mean + standard deviation = 1017 + 12.1

$$= 1029.1 \text{ hPa}$$

\therefore entire pressure system is high pressure

so, wind moves in clockwise direction (1)



Locations from North to South: Leuchars, Heathrow, Hurn

• Leuchars (W)

• Heathrow (NE)

• Hurn (E)

• Leuchars (W)

• Heathrow (NE)

• Hurn (E)

(Total for Question 3 is 8 marks)

4. Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leq 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \quad \sum r = 174.9 \quad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.

The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

(d) Explain why the distribution $B(14, 0.27)$ might **not** be a reasonable model for the number of days without rain for a 14-day summer event.

(1)

a) 0 (1)

b) (i) $\text{mean} = \frac{\sum r}{n} = \frac{174.9}{31} = 5.64 \text{ (3sf)}$ (1)

(ii) $\text{sd} = \sqrt{\frac{\sum r^2}{n} - \left(\frac{\sum r}{n}\right)^2} = \sqrt{\frac{3523.283}{31} - \left(\frac{174.9}{31}\right)^2}$ (1)

$= 9.05 \text{ (3sf)}$ (1)

c) Leuchars is further North than Camborne, but the mean Daily Total Rainfall is smaller for Leuchars than for Camborne. Therefore there is no evidence that Dian's belief is true. (1)

d) $p=0.27$ is unlikely to be constant ①

It's unlikely that there is a 27% chance of no rain every day for 14 days. A condition on the binomial distribution is that p must be constant.

5. Ben is studying the Daily Total Rainfall, x mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the following table.

x	0	0.1–0.5	0.6–1.0	1.1–1.9	2.0–4.0	4.1–6.9	7.0–12.0	12.1–20.9	21.0–32.0	tr
Frequency	55	18	18	21	17	9	9	6	2	29

- (a) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.

(2)

Using all 184 of these values, Ben estimates $\sum x = 390$ and $\sum x^2 = 4336$

- (b) Calculate estimates for

- (i) the mean Daily Total Rainfall,
 (ii) the standard deviation of the Daily Total Rainfall.

(3)

Ben suggests using the statistic calculated in part (b)(i) to estimate the annual mean Daily Total Rainfall in Leeming for 1987

- (c) Using your knowledge of the large data set,

- (i) give a reason why these data would not be suitable,
 (ii) state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

(2)

a) Replace 'tr' with a numerical value between 0 and 0.05.

For example, 0.025. $\textcircled{1}$ \leftarrow if the total amount of rainfall recorded is less than 0.05 mm, then it is recorded as 'tr'.

$$\text{b) (i) mean, } \bar{x} = \frac{\sum x}{n}$$

$$= \frac{390}{184}$$

$$= 2.119 \dots = 2.12 \text{ (3 s.f.)}$$

 $\textcircled{1}$

$$\begin{aligned} \text{(ii) Standard deviation} &= \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \\ &= \sqrt{\frac{4336}{184} - 2.119^2} \\ &= 4.367\dots = 4.37 \text{ (3 s.f.)} \end{aligned}$$

c) (i) The data only covers May to October. So, it is not a representative of the whole year. (1)

(ii) Winter months are missing when we'd expect more rain during this season. So, estimation in b(i) is expected to be an underestimate. (1)