The Large Data Set – Textbook exercise

(Answers are given at the end of the document)

1 From the eight weather stations featured in the large data set, write down:

a the station which is furthest north

b the station which is furthest south

c an inland station

d a coastal station

e an overseas station.

2 Explain, with reasons, whether daily maximum relative humidity is a discrete or continuous variable.

Questions 3 and 4 in this exercise use the following extracts from the large data set.

LEEMING

Date	Daily mean temperature (°C)	Daily total rainfall (mm)	Daily total sunshine (hrs)	Daily mean windspeed (kn)
01/06/2015	8.9	10	5.1	15
02/06/2015	10.7	tr	8.9	17
03/06/2015	12.0	0	10.0	8
04/06/2015	11.7	0	12.8	7
05/06/2015	15.0	0	8.9	9
06/06/2015	11.6	tr	5.4	17
07/06/2015	12.6	0	13.9	10
08/06/2015	9.4	0	9.7	7
09/06/2015	9.7	0	12.1	5
10/06/2015	11.0	0	14.6	4

HEATHROW

Date	Daily mean temperature (°C)	Daily total rainfall (mm)	Daily total sunshine (hrs)	Daily mean windspeed (kn)
01/06/2015	12.1	0.6	4.1	15
02/06/2015	15.4	tr	1.6	18
03/06/2015	15.8	0	9.1	9
04/06/2015	16.1	0.8	14.4	6
05/06/2015	19.6	tr	5.3	9
06/06/2015	14.5	0	12.3	12
07/06/2015	14.0	0	13.1	5
08/06/2015	14.0	tr	6.4	7
09/06/2015	11.4	0	2.5	10
10/06/2015	14.3	0	7.2	10

(P) **3** **a** Work out the mean of the daily total sunshine for the first 10 days of June 2015 in:

i Leeming

ii Heathrow.

b Work out the range of the daily total sunshine for the first 10 days of June 2015 in:

i Leeming

ii Heathrow.

c Supraj says that the further north you are, the fewer the number of hours of sunshine. State, with reasons, whether your answers to parts **a** and **b** support this conclusion.

Hint State in your answer whether Leeming is north or south of Heathrow.

(P) **4** Calculate the mean daily total rainfall in Heathrow for the first 10 days of June 2015. Explain clearly how you dealt with the data for 2/6/2015, 5/6/2015 and 8/6/2015.

(P) **5** Dominic is interested in seeing how the average monthly temperature changed over the summer months of 2015 in Jacksonville. He decides to take a sample of two days every month and average the temperatures before comparing them.

a Give one reason why taking two days a month might be:
i a good sample size
ii a poor sample size.

b He chooses the first day of each month and the last day of each month. Give a reason why this method of choosing days might not be representative.

c Suggest a better way that he can choose his sample of days.

(P) **6** The table shows the mean daily temperatures at each of the eight weather stations for August 2015:

	Camborne	Heathrow	Hurn	Leeming	Leuchars	Beijing	Jacksonville	Perth
Mean daily mean temp (°C)	15.4	18.1	16.2	15.6	14.7	26.6	26.4	13.6

a Give a geographical reason why the temperature in August might be lower in Perth than in Jacksonville.

b Comment on whether this data supports the conclusion that coastal locations experience lower average temperatures than inland locations.

(P) **7** Brian calculates the mean cloud coverage in Leeming in September 1987. He obtains the answer 9.3 oktas. Explain how you know that Brian's answer is incorrect.

(E/P) **8** The large data set provides data for 184 consecutive days in 1987. Marie is investigating daily mean windspeeds in Camborne in 1987.

a Describe how Marie could take a systematic sample of 30 days from the data for Camborne in 1987. **(3 marks)**

b Explain why Marie's sample would not necessarily give her 30 data points for her investigation. **(1 mark)**

Large data set

You will need access to the large data set and spreadsheet software to answer these questions.

1 **a** Find the mean daily mean pressure in Beijing in October 1987.

b Find the median daily rainfall in Jacksonville in July 2015.

c **i** Draw a grouped frequency table for the daily mean temperature in Heathrow in July and August 2015. Use intervals $10 \leqslant t < 15$, etc.

ii Draw a histogram to display this data.

iii Draw a frequency polygon for this data.

2 **a** **i** Take a simple random sample of size 10 from the data for daily mean windspeed in Leeming in 1987.

ii Work out the mean of the daily windspeeds using your sample.

b **i** Take a sample of the last 10 values from the data for daily mean windspeed in Leuchars in 1987.

ii Work out the mean of the daily mean windspeeds using your sample.

c State, with reasons, which of your samples is likely to be more representative.

d Suggest two improvements to the sampling methods suggested in part **a**.

e Use an appropriate sampling method and sample size to estimate the mean windspeeds in Leeming and Leuchars in 1987. State with a reason whether your calculations support the statement 'Coastal locations are likely to have higher average windspeeds than inland locations'.

Hint You can use the **CountIf** command in a spreadsheet to work out the frequency for each class.

Answers

1 **a** Leuchars
b Perth
c ANY ONE FROM: Leeming, Heathrow, Beijing
d ANY ONE FROM: Leuchars, Hurn, Camborne, Jacksonville, Perth
e ANY ONE FROM: Beijing, Jacksonville, Perth

2 Continuous – it can take any value in the range 0 to 100

3 **a** **i** 10.14 hours **ii** 7.6 hours
b **i** 9.5 hours **ii** 12.8 hours
c The mean of the daily total sunshine in Leeming is higher than that in Heathrow. Leeming is north of Heathrow, so these data do not support Supraj's conclusion.

4 0.14 mm, treat tr. as 0 in numerical calculations.

5 **a** **i** Covers several months **ii** Small sample size
b Two consecutive days chosen all the time – not random, possibly have similar weather.
c Number the days and choose a simple random sample.

6 **a** Perth is in the southern hemisphere so August is a winter month
b The lowest temperatures in the UK are at coastal locations (Camborne and Leuchars). The highest temperature is at an inland location (Beijing). There is some evidence to support this conclusion.

7 Oktas measure the cloud coverage in eighths. The highest value is 8 which represents full cloud coverage.

8 **a** She needs to select days at regular intervals in an ordered list. Put the days into date order. Select every sixth day ($184 \div 30 = 6.13$).
b Some of the data values might not be available (n/a).

Large data set

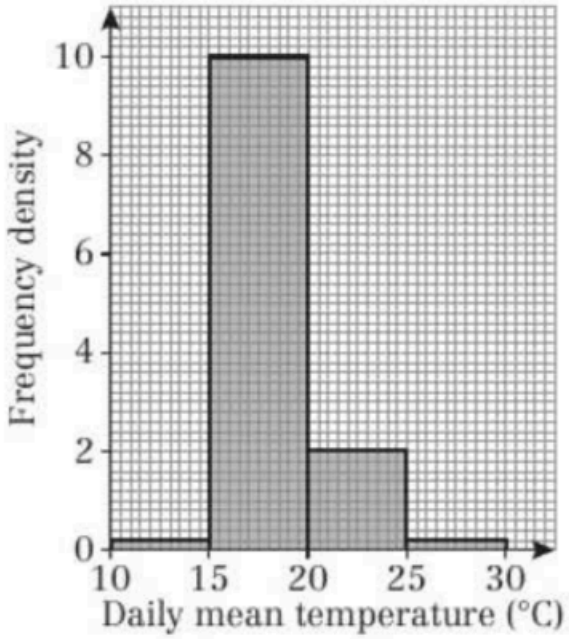
1 a 1020 hPa

b 0.0 mm

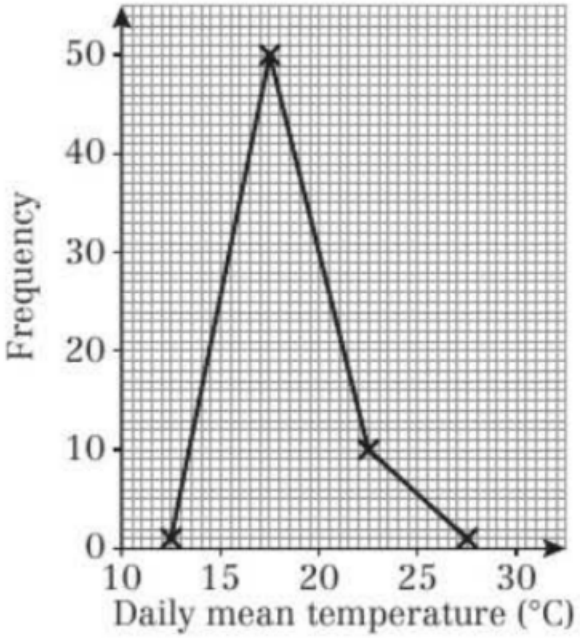
c i

Temperature, t (°C)	Frequency
$10 \leqslant t < 15$	1
$15 \leqslant t < 20$	50
$20 \leqslant t < 25$	10
$25 \leqslant t < 30$	1

ii



iii



2 Students' own answer.