# Stats1 Chapter 1: Data Collection

1.5 The Large Data Set

# Name That Sampling Method!

Simple Random	Systematic	Stratified	Quota	Opportunity
Sampling	Sampling	Sampling	Sampling	Sampling

### Suggest a suitable sampling method.

"You wish to test lightbulbs produced by a factory in a daily batch."	?
"You wish to survey consumer opinion on your new drink <i>FizzGuzz</i> released in the UK."	?
"You wish to determine students' favourite TV programmes in your school, that is fairly representative of each year group."	?

## Name That Sampling Method!

Simple Random Systematic Stratified Quota Opportunity Sampling Sampling Sampling Sampling

#### Suggest a suitable sampling method.

"You wish to test lightbulbs produced by a factory in a daily batch." Probably **systematic sampling**, as the method of choosing items is simpler than simple random sampling (where it would be time-consuming to find specifically chosen random light bulbs). Sampling frame is known.

"You wish to survey consumer opinion on your new drink *FizzGuzz* released in the UK."

**Quota sampling** or **opportunity sampling**. We'd realistically not have access to the sampling frame (i.e. a list of all UK residents).

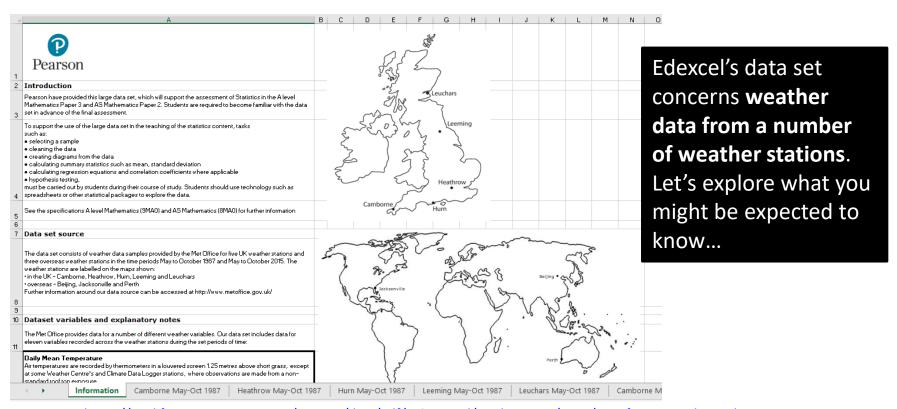
"You wish to determine students' favourite TV programmes in your school, that is fairly representative of each year group."

**Stratified sampling**. We (probably) have access to the sampling frame (i.e. a list of all students). Stratified sampling ensures that each stratum (year group) is proportionately represented.

### Large Data Set

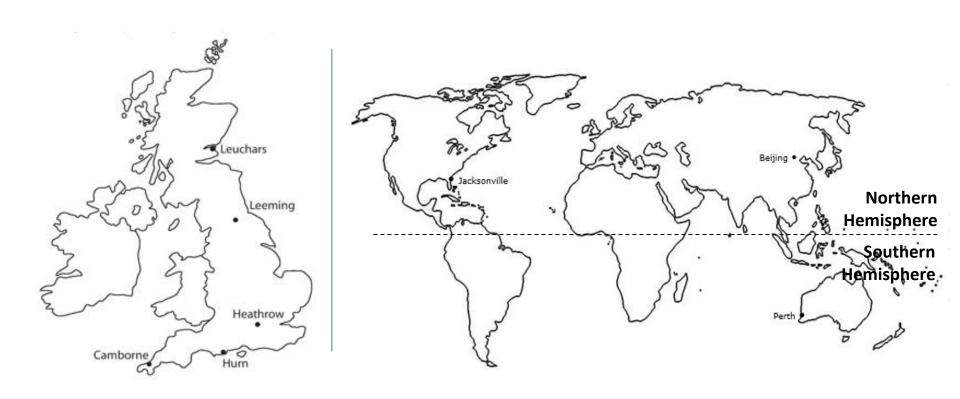
All A Level exam boards are obligated to provide a 'large data set'. Data in exam questions will often be from this set, and you are encouraged to explore this data (which is publicly available) in Microsoft Excel.

It is important to note that you are expected to be familiar with this data set before you go into your exam, including some basic geographic knowledge!



https://qualifications.pearson.com/content/dam/pdf/A%20Level/Mathematics/2017/specification-and-sample-assesment/Pearson%20Edexcel%20GCE%20AS%20and%20AL%20Mathematics%20data%20set%20-%20Issue%201%20(1).xls

### What You Need To Be Familiar With...





You should know the names and rough locations of the 5 UK weather stations, as well as the 3 international weather stations.

#### The data was recorded for:

- May-Oct 1987
- May-Oct 2015

All the following are daily...

but it is mean temp

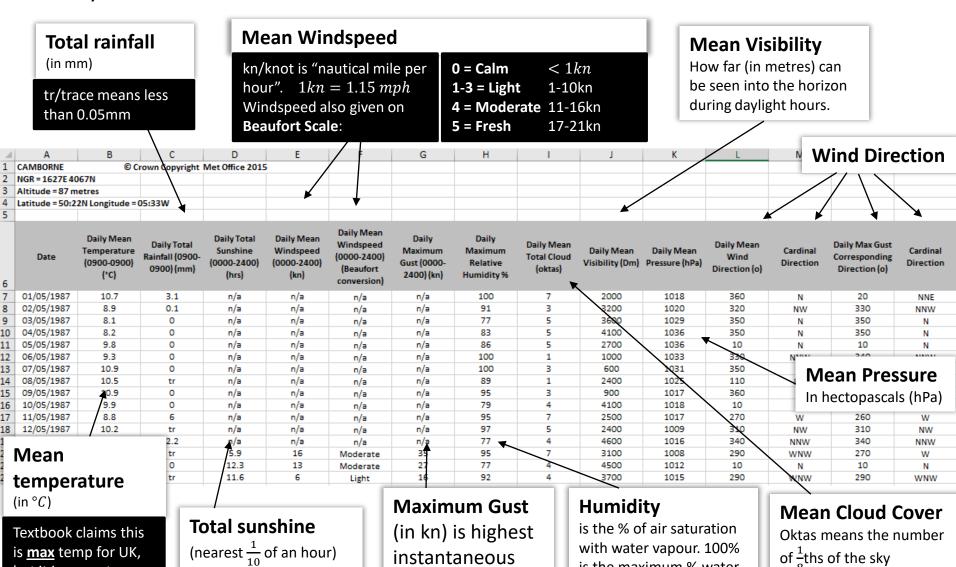
for all locations.

You should be familiar with the variables involved and their respective units.

is the maximum % water

content air can contain.

covered.



wind speed.



# You should have a vague idea of the range of values for each location.

UK Location (2015)	Temp Range	Wind Speed Range
Camborne	10-20	3-18
Heathrow	8- <b>29</b>	3-19
Hurn	6-24	2-19
Leeming	4-23	3-17
Leuchars	4-19	3-23

World Location (2015)	Temp Range	Wind Speed Range
Beijing	8-33	2-9
Jacksonville	15-31	1-12
Perth	8-25	4-14
	<b>A</b>	

Beijing temp range relatively large. Min Jacksonville temp high. Perth similar to UK.

Mean wind speed in UK across full period was roughly 9 nm. But 4 nm in Beijing (i.e. lower), 5 in Jacksonville (again lower), 8 in Perth (similar to UK).

#### From new A Level sample assessment materials:

"A meteorologist believes that there is a relationship between the daily mean windspeed, w kn, and the daily mean temperature, t °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below. ...

Using the same 9 days, a location from the large data set gave  $\bar{t}=27.2$  and  $\bar{w}=3.5$ .

(d) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics."

# You should have a vague idea of the range of values for each variable for the data set as a whole.

Variable	Typical value(s)
Gust (UK only)	8 – 52 nm
Rainfall	0 – 60 mm in UK, but more extreme maximums elsewhere (e.g. 102mm in Perth)
Pressure	988 – 1038 hPa
Wind Speed on Beaufort scale	Max is 'fresh' (5). Most Light or Moderate.
Sunshine (UK only)	0 – 16 hrs
Cloud Cover	0 – 8 ocktas (i.e. full spread)

#### Hurn

© Crown Copyright Met Office 1987

Date	Daily Max Temp (09- 00-0900 C)	Daily Total Rainfall (0900- 0900) (mm)	Daily Total Sunshine (0000- 2400) (hrs)	Daily Mean Windspee d (0000- 2400) (kn)	Daily Mean Windspeed (0000-2400) (Beaufort conversion)	Daily Maximum Gust (0000- 2400) (kn)
01/06/1987	15.1	0.6	4.5	7	Light	19
02/06/1987	12.5	4.7	0	7	Light	22
03/06/1987	13.8	tr	5.6	11	Moderate	25
04/06/1987	15.5	5.3	7.8	7	Light	17
05/06/1987	13.1	19	0.5	10	Light	33
06/06/1987	13.8	0	8.9	19	Fresh	46
07/06/1987	13.2	tr	3.8	11	Moderate	27
08/06/1987	12.9	1	1.7	9	Light	19
09/06/1987	11.2	tr	5.4	6	Light	19
10/06/1987	9.2	1.3	9.7	4	Light	n/a
11/06/1987	12.6	0	12.5	6	Light	18
12/06/1987	10.4	0	11.9	5	Light	n/a
13/06/1987	9.6	0	8.6	5	Light	15
14/06/1987	10.2	0	13.1	5	Light	18
15/06/1987	9.2	3.7	7.1	4	Light	25
16/06/1987	10.4	5.6	8.3	6	Light	25
17/06/1987	12.8	0.1	5.3	10	Light	27
18/06/1987	13.0	7.4	3.2	9	Light	24
19/06/1987	14.0	tr	0.4	12	Moderate	33
20/06/1987	12.6	0	7.7	6	Light	17

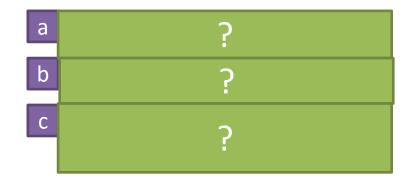
As previously noted, the **actual** data set has mean temperature for <u>all</u> locations. I changed to maximum temperature for this example for consistency with the textbook.

#### [Textbook]

(a) Describe the type of data represented by daily total rainfall.

Alison is investigating daily maximum gust. She wants to select a sample of size 5 from the first 20 days in Hurn in June 1987. She uses the first two digits of the date as a sampling frame and generates five random numbers between 1 and 20. b) State the type of sample selected by Alison.

c) Explain why Alison's process might not generate a sample of size 5.



#### Hurn

© Crown Copyright Met Office 1987

		676				
Date	Daily Max Temp (09- 00-0900 C)	Daily Total Rainfall (0900- 0900) (mm)	Daily Total Sunshine (0000- 2400) (hrs)	Daily Mean Windspee d (0000- 2400) (kn)	Daily Mean Windspeed (0000-2400) (Beaufort conversion)	Daily Maximum Gust (0000- 2400) (kn)
01/06/1987	15.1	0.6	4.5	7	Light	19
02/06/1987	12.5	4.7	0	7	Light	22
03/06/1987	13.8	tr	5.6	11	Moderate	25
04/06/1987	15.5	5.3	7.8	7	Light	17
05/06/1987	13.1	19	0.5	10	Light	33
06/06/1987	13.8	0	8.9	19	Fresh	46
07/06/1987	13.2	tr	3.8	11	Moderate	27
08/06/1987	12.9	1	1.7	9	Light	19
09/06/1987	11.2	tr	5.4	6	Light	19
10/06/1987	9.2	1.3	9.7	4	Light	n/a
11/06/1987	12.6	0	12.5	6	Light	18
12/06/1987	10.4	0	11.9	5	Light	n/a
13/06/1987	9.6	0	8.6	5	Light	15
14/06/1987	10.2	0	13.1	5	Light	18
15/06/1987	9.2	3.7	7.1	4	Light	25
16/06/1987	10.4	5.6	8.3	6	Light	25
17/06/1987	12.8	0.1	5.3	10	Light	27
18/06/1987	13.0	7.4	3.2	9	Light	24
19/06/1987	14.0	tr	0.4	12	Moderate	33
20/06/1987	12.6	0	7.7	6	Light	17

As previously noted, the **actual** data set has mean temperature for <u>all</u> locations. I changed to maximum temperature for this example for consistency with the textbook.

#### [Textbook]

(a) Describe the type of data represented by daily total rainfall.

Alison is investigating daily maximum gust. She wants to select a sample of size 5 from the first 20 days in Hurn in June 1987. She uses the first two digits of the date as a sampling frame and generates five random numbers between 1 and 20. b) State the type of sample selected by Alison.

- c) Explain why Alison's process might not generate a sample of size 5.
  - a Continuous quantitative data.
  - **b** Simple random sample.
  - Some of the data values are not available (n/a).

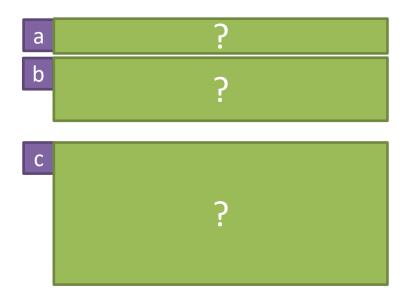
#### Hurn

#### © Crown Copyright Met Office 1987

Date	Daily Max Temp (09- 00-0900 C)	Daily Total Rainfall (0900- 0900) (mm)	Daily Total Sunshine (0000- 2400) (hrs)	Daily Mean Windspee d (0000- 2400) (kn)	Daily Mean Windspeed (0000-2400) (Beaufort conversion)	Daily Maximum Gust (0000- 2400) (kn)
01/06/1987	15.1	0.6	4.5	7	Light	19
02/06/1987	12.5	4.7	0	7	Light	22
03/06/1987	13.8	tr	5.6	11	Moderate	25
04/06/1987	15.5	5.3	7.8	7	Light	17
05/06/1987	13.1	19	0.5	10	Light	33
06/06/1987	13.8	0	8.9 19		Fresh	46
07/06/1987	13.2	tr	3.8	11	Moderate	27
08/06/1987	12.9	1	1.7	9	Light	19
09/06/1987	11.2	tr	5.4	6	Light	19
10/06/1987	9.2	1.3	9.7	4	Light	n/a
11/06/1987	12.6	0	12.5	6	Light	18
12/06/1987	10.4	0	11.9	5	Light	n/a
13/06/1987	9.6	0	8.6	5	Light	15
14/06/1987	10.2	0	13.1	5	Light	18
15/06/1987	9.2	3.7	7.1	4	Light	25
16/06/1987	10.4	5.6	8.3	6	Light	25
17/06/1987	12.8	0.1	5.3	10	Light	27
18/06/1987	13.0	7.4	3.2	9	Light	24
19/06/1987	14.0	tr	0.4	12	Moderate	33
20/06/1987	12.6	0	7.7	6	Light	17

#### [Textbook] Calculate:

- a) The mean daily maximum temperature for the first five days of June in Hurn in 1987.
- b) The median daily total rainfall for the week of 14<sup>th</sup> June to 20<sup>th</sup> June inclusive.
- c) The median daily total rainfall for the same week in Perth was 19.00mm. Karl states that more southerly countries experience higher rainfall during June. State with a reason whether your answer to part (b) supports this statement.



#### Hurn

#### © Crown Copyright Met Office 1987

Date	Daily Max Temp (09- 00-0900 C)	Daily Total Rainfall (0900- 0900) (mm)	Daily Total Sunshine (0000- 2400) (hrs)	Daily Mean Windspee d (0000- 2400) (kn)	Daily Mean Windspeed (0000-2400) (Beaufort conversion)	Daily Maximum Gust (0000- 2400) (kn)
01/06/1987	15.1	0.6	4.5	7	Light	19
02/06/1987	12.5	4.7	0	7	Light	22
03/06/1987	13.8	tr	5.6	11	Moderate	25
04/06/1987	15.5	5.3	7.8	7	Light	17
05/06/1987	13.1	19	0.5	10	Light	33
06/06/1987	13.8	0	8.9 19		Fresh	46
07/06/1987	13.2	tr	3.8	11	Moderate	27
08/06/1987	12.9	1	1.7	9	Light	19
09/06/1987	11.2	tr	5.4	6	Light	19
10/06/1987	9.2	1.3	9.7	4	Light	n/a
11/06/1987	12.6	0	12.5	6	Light	18
12/06/1987	10.4	0	11.9	5	Light	n/a
13/06/1987	9.6	0	8.6	5	Light	15
14/06/1987	10.2	0	13.1	5	Light	18
15/06/1987	9.2	3.7	7.1	4	Light	25
16/06/1987	10.4	5.6	8.3	6	Light	25
17/06/1987	12.8	0.1	5.3	10	Light	27
18/06/1987	13.0	7.4	3.2	9	Light	24
19/06/1987	14.0	tr	0.4	12	Moderate	33
20/06/1987	12.6	0	7.7	6	Light	17

#### [Textbook] Calculate:

- a) The mean daily maximum temperature for the first five days of June in Hurn in 1987.
- b) The median daily total rainfall for the week of 14<sup>th</sup> June to 20<sup>th</sup> June inclusive.
- c) The median daily total rainfall for the same week in Perth was 19.00mm. Karl states that more southerly countries experience higher rainfall during June. State with a reason whether your answer to part (b) supports this statement.

a 
$$70.0 \div 5 = 14.0 \,{}^{\circ}C \, (1dp)$$

- b Values in ascending order: 0, 0, tr, 0.1, 3.7, 5.6, 7.4. ∴ Median is 0.1mm.
- Perth is in Australia, which is south of the UK, and the median rainfall was higher. However, this is a very small sample from a single location in each country so does not provide enough evidence to support Karl's statement.

## Exercise 1.5

Pearson Statistics & Mechanics Year 1/AS Page 5

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

LEEMING	LEEMING © Crown Copyright Met Office 2015								
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 © Crown Copyright Met Office 2015							
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			

- 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.
- 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.

- 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.
- 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.

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

Crown Copyright Met Office

- 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.
- 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.
- 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)
  - Explain why Marie's sample would not necessarily give her 30 data points for her investigation.

    (1 mark)

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  $10 \le t < 15$ , 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.
  - 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'.

Countif command in a spreadsheet to work out the frequency for each class.

### **Homework 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 ÷ 30 = 6.13).
  - b Some of the data values might not be available (n/a).

# Homework Answers

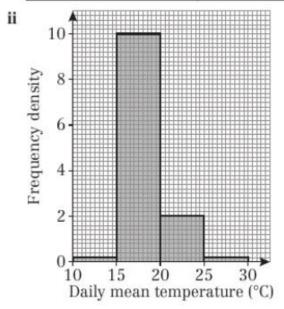
#### Large data set

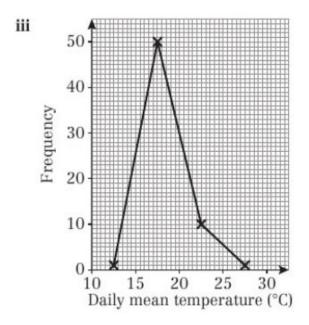
1 a 1020 hPa

**b** 0.0 mm

c

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





2 Students' own answer.