1 A council wants to sample households in a town in order to find out what proportion of their waste is recycled. There are 20 000 houses in the town and the council want to take a sample of 10% of these houses.

  a Describe how a systematic random sample could be used. **[2 marks]**

The council intends to contact the selected households by telephone and request permission to track how much material they recycle over a 3-month period.

  b Suggest a practical difficulty that may be encountered when contacting households. **[1]**

  c Give a reason why the results from the sampled households may not be representative of the population. **[1]**

2 A battery manufacturer needs to test the average life of its batteries.

  a What is the population in this case? **[1]**

  b Aside from cost, give a reason why only a sample should be tested. **[1]**

  c What would be a sensible statistic to calculate? **[1]**

3 The daily total sunshine, in hours, recorded over a 12-day period during the summer at Hurn weather station are given.

0.3	6.2	10.5	12.4	12.3	14.3
15	14.9	10.3	22	14.5	10.1

Large data set

The mean hours of daily total sunshine over this period was 10.95

  a One of the measurements has been written down incorrectly.

Identify the error and calculate the correct value. **[3]**

  b Calculate the variance of the correct data. **[2]**

  c Calculate the median and explain whether the mean or the median would be the best average to use in this case. **[3]**

4 The box-and-whisker plots show the maximum daily temperature in August 2015 for two different weather stations.

Large data set

Leuchars

Hurn

15 16 17 18 19 20 21 22 23 24 25 26
Maximum daily temperature (°C)
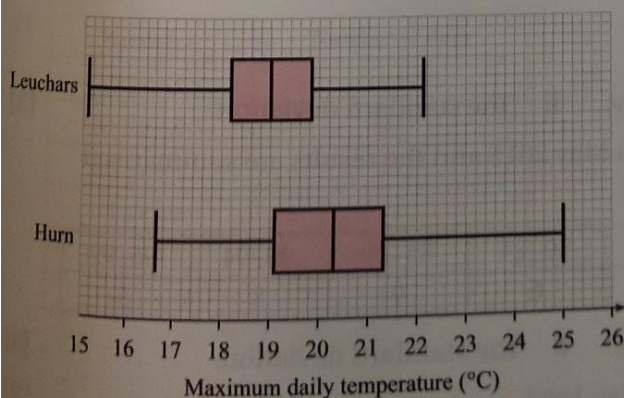
  a Estimate the interquartile range at each station. **[3]**

  b Write down the median temperature at each station. **[2]**

  c Compare the maximum daily temperatures during August at each weather station. **[3]**

An outlier is defined as a value outside of the range $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$
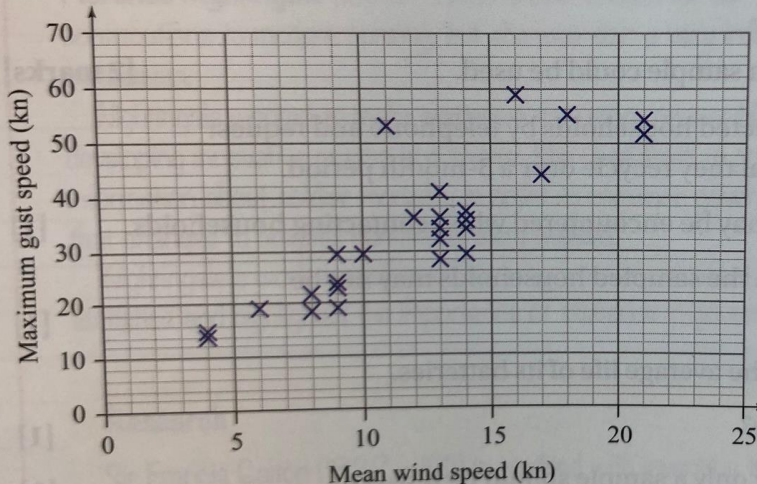
  d Work out whether or not these values are outliers, showing your working clearly.

   **i** 15.1°C at Leuchars   **ii** 22.1°C at Leuchars   **iii** 25°C at Hurn **[6]**

5 The scatter diagram shows the mean windspeed and maximum gust speed in Camborne in October 1987

Describe the relationship between the mean windspeed and maximum gust speed. [2]

6 A teacher wants to find out the average number of homework assignments that students have been given each week. He takes a sample of 30 students. The results are shown in this graph.
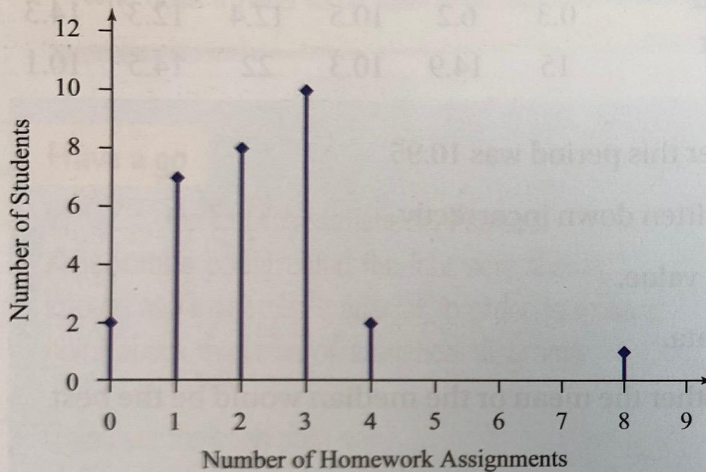
a What is the modal number of assignments? [1]

b Calculate the mean and standard deviation of the sample. [4]

c State the median and first and third quartiles. [2]

d Advise which measure of average would be best to use, stating your reasons clearly. [1]

7 The maximum daily relative humidity, h, was measured at Heathrow weather station every day in May 1987

You are given $\sum h = 2824$ and $\sum h^2 = 25\,8304$

a Calculate **i** The mean relative humidity, **ii** The standard deviation. [3]

The same measurement was taken every day in May 2015 and the results are summarised in the table.

Relative Humidity (%)	70–74	75–79	80–84	85–89	90–94	95–99
Number of days	1	2	6	9	8	5

b Estimate **i** The mean relative humidity, **ii** The standard deviation. [5]

c Compare the humidity in May 2015 with May 1987 [2]

**d** Draw a frequency polygon for the 2015 results. [3]

8 A sample of students in a sixth form is to be taken and surveyed regarding their use of the library. The numbers of boys and girls in each year is given in the table.

	Year 12	Year 13
Girls	60	51
Boys	39	30

a It is suggested that a sample of size 60 should be taken by randomly selecting 15 boys and 15 girls from each year. State a disadvantage of taking a sample in this way. **[1]**

b It is instead decided to use a stratified sample of 60 students. Calculate the number of boys and the number of girls that should be sampled in each year. **[4]**

c In order to select individuals for the survey, an interviewer will randomly choose students as they leave their common room. Explain why the results of the survey could be biased. **[1]**

9 An outlier is an observation greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$

A box-and-whisker plot is drawn for a large volume of data. Four extra observations are then recorded. Which of the extra observations A, B, C and D are outliers? Show your working. **[3]**
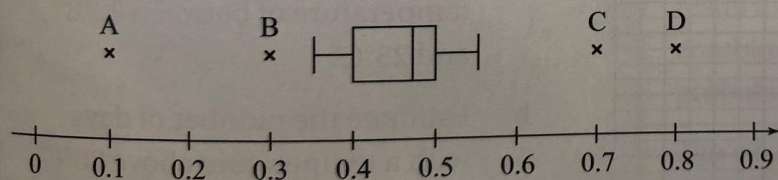
10 An experiment was carried out using tomato plant seeds. Trays of seeds were planted and each tray was placed in a controlled environment with a different temperature for each tray. All other variables, such as light and water, were the same for each of the trays. After 10 days, the number of seeds that had germinated in each was counted. The results are shown in the scatter diagram.
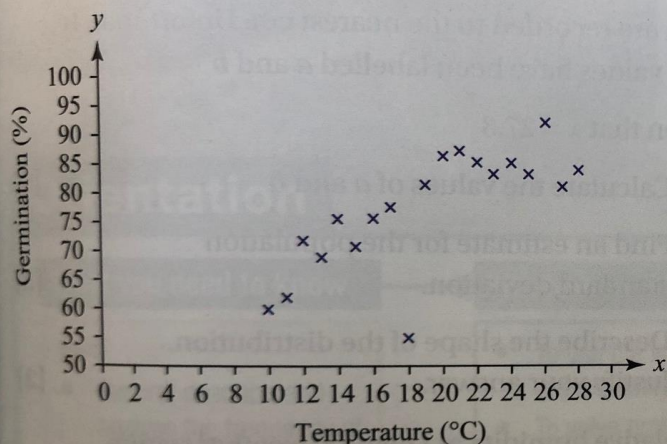
a Which is the explanatory variable? **[1]**

b Describe the relationship observed. **[2]**

c It is suggested that a temperature of 35°C would result in almost all seeds germinating. Comment on whether this is a sensible suggestion. **[2]**

11

Rainfall, r	Number of days
$0 \le r < 0.1$	84
$0.1 \le r < 0.5$	27
$0.5 \le r < 1$	19
$1 \le r < 2$	14
$2 \le r < 5$	17
$5 \le r < 30$	23

The rainfall is measured (in mm) at Leeming weather station every day from May to October 2015. The results are summarised in this table.

A histogram is to be drawn to show this data. The bar for $0.5 \le r < 1$ is 5 mm wide and 76 mm tall.

a Calculate the height and width of the $1 \le r < 2$ bar. **[4]**

b Estimate to 2 dp

  i The median,

  ii The interquartile range. **[6]**

12 The police in a town wish to survey members of a number of "Neighbourhood Watch" schemes. They wish to survey people from each scheme and the number selected is to be in proportion to the size of the scheme. Two possible methods of selecting the sample are suggested.

Method A: 10 people are randomly selected from each scheme from a list of all members.

Method B: The person who manages each scheme is asked to choose a proportionally sized sample from their population at random.

a State the name of each of these methods of sampling. **[2]**

b Which method is preferable? Clearly explain why this method is better. **[2]**

13 This histogram shows the maximum daily temperature at Heathrow weather station in July 2015
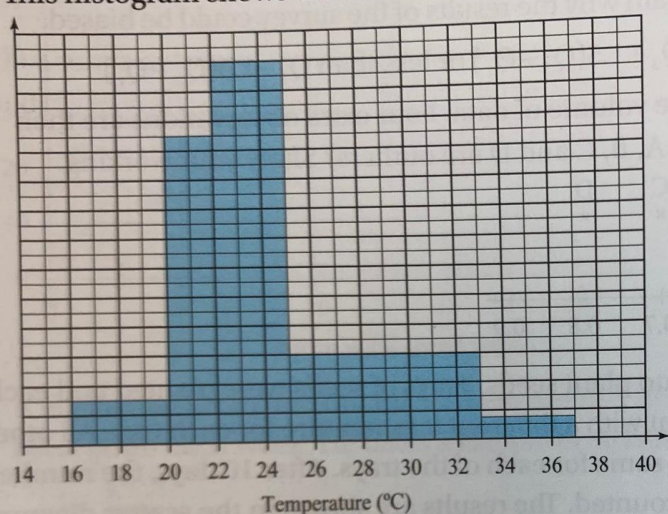
Temperature (°C)

The maximum temperature was below 20°C on just two days.

a How many days had a maximum temperature of between 22°C and 25°C? **[3]**

b Estimate the number of days with a temperature above 26°C. **[4]**

c Use the histogram to estimate the mean maximum daily temperature. **[4]**

14 The lengths of 40 fish caught in a competition are recorded to the nearest cm. Unfortunately, some of the numbers are now illegible. These values have been labelled a and b
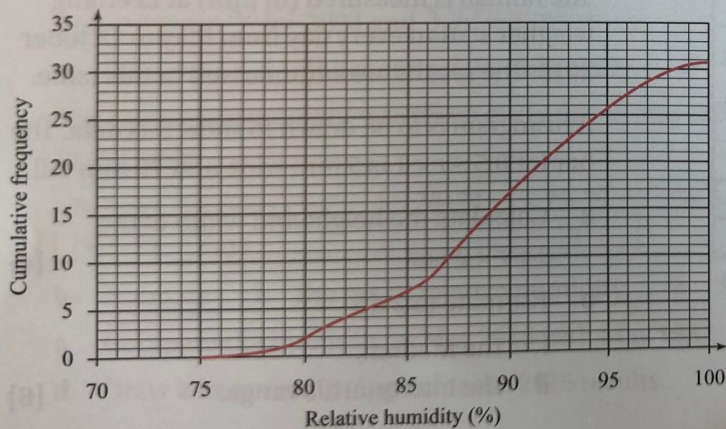
Length (cm)	Frequency
18–22	3
23–25	10
26–28	14
29–31	a
32–40	b

Given that $\bar{x} = 27.3$

a Calculate the values of a and b [5]

b Find an estimate for the population standard deviation. [3]

c Describe the shape of the distribution. Justify your answer. [2]

15 The cumulative frequency curve shows the relative humidity at Leeming weather station in May 2015

a What is the median relative humidity? [1]

b Calculate the interquartile range. [2]

A day with relative humidity over 95% is likely to have been foggy or misty.

c Estimate the percentage of days that were foggy or misty. [3]