

# Data collection

# 1

## Objectives

After completing this chapter you should be able to:

- Understand 'population', 'sample' and 'census', and comment on the advantages and disadvantages of each → pages 2–3
- Understand the advantages and disadvantages of simple random sampling, systematic sampling, stratified sampling, quota sampling and opportunity sampling → pages 4–9
- Define qualitative, quantitative, discrete and continuous data, and understand grouped data → pages 9–10
- Understand the large data set and how to collect data from it, identify types of data and calculate simple statistics → pages 11–16

## Prior knowledge check

- 1** Find the mean, median, mode and range of these data sets:

**a** 1, 3, 4, 4, 6, 7, 8, 9, 11      **b** 20, 18, 17, 20, 14, 23, 19, 16

← GCSE Mathematics

- 2** Here is a question from a questionnaire surveying TV viewing habits.

How much TV do you watch?

☐ 0–1 hours    ☐ 1–2 hours    ☐ 3–4 hours

Give two criticisms of the question and write an improved question.

← GCSE Mathematics

- 3** Rebecca records the shoe size,  $x$ , of the female students in her year. The results are given in the table.

Find:

- a** the number of female students who take shoe size 37
- b** the shoe size taken by the smallest number of female students
- c** the shoe size taken by the greatest number of female students
- d** the total number of female students in the year.

$x$	Number of students, $f$
35	3
36	17
37	29
38	34
39	12

← GCSE Mathematics

Meteorologists collect and analyse weather data to help them predict weather patterns. Selecting weather data from specific dates and places is an example of sampling.

→ Section 1.5

## 1.1 Populations and samples

■ **In statistics, a population is the whole set of items that are of interest.**

For example, the population could be the items manufactured by a factory or all the people in a town. Information can be obtained from a population. This is known as raw data.

■ **A census observes or measures every member of a population.**

■ **A sample is a selection of observations taken from a subset of the population which is used to find out information about the population as a whole.**

There are a number of advantages and disadvantages of both a census and a sample.

	Advantages	Disadvantages
<b>Census</b>	<ul style="list-style-type: none"> <li>It should give a completely accurate result</li> </ul>	<ul style="list-style-type: none"> <li>Time consuming and expensive</li> <li>Cannot be used when the testing process destroys the item</li> <li>Hard to process large quantity of data</li> </ul>
<b>Sample</b>	<ul style="list-style-type: none"> <li>Less time consuming and expensive than a census</li> <li>Fewer people have to respond</li> <li>Less data to process than in a census</li> </ul>	<ul style="list-style-type: none"> <li>The data may not be as accurate</li> <li>The sample may not be large enough to give information about small sub-groups of the population</li> </ul>

The size of the sample can affect the validity of any conclusions drawn.

- The size of the sample depends on the required accuracy and available resources.
- Generally, the larger the sample, the more accurate it is, but you will need greater resources.
- If the population is very varied, you need a larger sample than if the population were uniform.
- Different samples can lead to different conclusions due to the natural variation in a population.

■ **Individual units of a population are known as sampling units.**

■ **Often sampling units of a population are individually named or numbered to form a list called a sampling frame.**

### Example 1

A supermarket wants to test a delivery of avocados for ripeness by cutting them in half.

**a** Suggest a reason why the supermarket should not test all the avocados in the delivery.

The supermarket tests a sample of 5 avocados and finds that 4 of them are ripe.

They estimate that 80% of the avocados in the delivery are ripe.

**b** Suggest one way that the supermarket could improve their estimate.

**a** Testing all the avocados would mean that there were none left to sell.

When testing a product destroys it, a 'census' is not appropriate.

**b** They could take a larger sample, for example 10 avocados. This would give a better estimate of the overall proportion of ripe avocados.

In general, larger samples produce more accurate predictions about a population.



**Exercise 1A**

- 1 A school uses a census to investigate the dietary requirements of its students.
  - a Explain what is meant by a census.
  - b Give one advantage and one disadvantage to the school of using a census.
  
- 2 A factory makes safety harnesses for climbers and has an order to supply 3000 harnesses. The buyer wishes to know that the load at which the harness breaks exceeds a certain figure.
  - a Suggest a reason why a census would not be used for this purpose.

The factory tests four harnesses and the load for breaking is recorded:

320 kg	260 kg	240 kg	180 kg
--------	--------	--------	--------

  - b The factory claims that the harnesses are safe for loads up to 250 kg. Use the sample data to comment on this claim.
  - c Suggest one way in which the company can improve their prediction.
  
- 3 A city council wants to know what people think about its recycling centre. The council decides to carry out a sample survey to learn the opinion of residents.
  - a Write down one reason why the council should not take a census.
  - b Suggest a suitable sampling frame.
  - c Identify the sampling units.
  
- 4 A manufacturer of microswitches is testing the reliability of its switches. It uses a special machine to switch them on and off until they break.
  - a Give one reason why the manufacturer should use a sample rather than a census.

The company tests a sample of 10 switches, and obtains the following results:

23 150	25 071	19 480	22 921	7455
--------	--------	--------	--------	------

  - b The company claims that its switches can be operated an average of 20 000 times without breaking. Use the sample data above to comment on this claim.
  - c Suggest one way the company could improve its prediction.
  
- 5 A manager of a garage wants to know what their mechanics think about a new pension scheme designed for them. The manager decides to ask all the mechanics in the garage.
  - a Describe the population the manager will use.
  - b Write down the main advantage in asking all of their mechanics.

## 1.2 Sampling

In random sampling, every member of the population has an equal chance of being selected. The sample should therefore be **representative** of the population. Random sampling also helps to remove **bias** from a sample.

There are three methods of random sampling:

- Simple random sampling
- Systematic sampling
- Stratified sampling

■ **A simple random sample of size  $n$  is one where every sample of size  $n$  has an equal chance of being selected.**

To carry out a simple random sample, you need a sampling frame, usually a list of people or things. Each person or thing is allocated a unique number and a selection of these numbers is chosen at random.

There are two methods of choosing the numbers: generating random numbers (using a calculator, computer or random number table) and **lottery** sampling.

In lottery sampling, the members of the sampling frame could be written on tickets and placed into a 'hat'. The required number of tickets would then be drawn out.

### Example 2

The 100 members of a yacht club are listed alphabetically in the club's membership book.

The committee wants to select a sample of 12 members to fill in a questionnaire.

- Explain how the committee could use a calculator or random number generator to take a simple random sample of the members.
- Explain how the committee could use a lottery sample to take a simple random sample of the members.

- Allocate a number from 1 to 100 to each member of the yacht club. Use your calculator or a random number generator to generate 12 random numbers between 1 and 100. Go back to the original population and select the people corresponding to these numbers.
- Write all the names of the members on (identical) cards and place them into a hat. Draw out 12 names to make up the sample of members.

If your calculator generates a number that has already been selected, ignore that number and generate an extra random number.

■ **In systematic sampling, the required elements are chosen at regular intervals from an ordered list.**

For example, if a sample of size 20 was required from a population of 100, you would take every fifth person since  $100 \div 20 = 5$ .

The first person to be chosen should be chosen at random. So, for example, if the first person chosen is number 2 in the list, the remaining sample would be persons 7, 12, 17 etc.

- **In stratified sampling, the population is divided into mutually exclusive strata (males and females, for example) and a random sample is taken from each.**

The proportion of each strata sampled should be the same. A simple formula can be used to calculate the number of people we should sample from each stratum:

$$\text{The number sampled in a stratum} = \frac{\text{number in stratum}}{\text{number in population}} \times \text{overall sample size}$$

### Example 3

A factory manager wants to find out what his workers think about the factory canteen facilities. The manager decides to give a questionnaire to a sample of 80 workers. It is thought that different age groups will have different opinions.

There are 75 workers between ages 18 and 32.

There are 140 workers between 33 and 47.

There are 85 workers between 48 and 62.

- Write down the name of the method of sampling the manager should use.
- Explain how he could use this method to select a sample of workers' opinions.

a Stratified sampling.

b There are:  $75 + 140 + 85 = 300$  workers altogether.

$$18-32: \frac{75}{300} \times 80 = 20 \text{ workers.}$$

$$33-47: \frac{140}{300} \times 80 = 37\frac{1}{3} \approx 37 \text{ workers.}$$

$$48-62: \frac{85}{300} \times 80 = 22\frac{2}{3} \approx 23 \text{ workers.}$$

Number the workers in each age group. Use a random number table (or generator) to produce the required quantity of random numbers. Give the questionnaire to the workers corresponding to these numbers.

Find the total number of workers.

For each age group find the number of workers needed for the sample.

Where the required number of workers is not a whole number, round to the nearest whole number.

Each method of random sampling has advantages and disadvantages.

Simple random sampling	
Advantages	Disadvantages
<ul style="list-style-type: none"> <li>• Free of bias</li> <li>• Easy and cheap to implement for small populations and small samples</li> <li>• Each sampling unit has a known and equal chance of selection</li> </ul>	<ul style="list-style-type: none"> <li>• Not suitable when the population size or the sample size is large</li> <li>• A sampling frame is needed</li> </ul>



**Systematic sampling****Advantages**

- Simple and quick to use
- Suitable for large samples and large populations

**Disadvantages**

- A sampling frame is needed
- It can introduce bias if the sampling frame is not random

**Stratified sampling****Advantages**

- Sample accurately reflects the population structure
- Guarantees proportional representation of groups within a population

**Disadvantages**

- Population must be clearly classified into distinct strata
- Selection within each stratum suffers from the same disadvantages as simple random sampling

**Exercise 1B**

- 1 a The head teacher of an infant school wishes to take a stratified sample of 20% of the pupils at the school. The school has the following numbers of pupils.

Year 1	Year 2	Year 3
40	60	80

Work out how many pupils in each age group will be in the sample.

**Problem-solving**

When describing advantages or disadvantages of a particular sampling method, always refer to the context of the question.

- b Describe one benefit to the head teacher of using a stratified sample.
- 2 A survey is carried out on 100 members of the adult population of a city suburb. The population of the suburb is 2000. An alphabetical list of the inhabitants of the suburb is available.
- a Explain one limitation of using a systematic sample in this situation.
- b Describe a sampling method that would be free of bias for this survey.
- 3 A gym wants to take a sample of its members. Each member has a 5-digit membership number, and the gym selects every member with a membership number ending 000.
- a Is this a systematic sample? Give a reason for your answer.
- b Suggest one way of improving the reliability of this sample.
- 4 A head of sixth form wants to get the opinion of year 12 and year 13 students about the facilities available in the common room. The table shows the numbers of students in each year.

	Year 12	Year 13
Male	70	50
Female	85	75

- a Suggest a suitable sampling method that might be used to take a sample of 40 students.
- b How many students from each gender in each of the two years should the head of sixth form ask?

- 5 A factory manager wants to get information about the ways their workers travel to work. There are 480 workers in the factory, and each has a clocking-in number. The numbers go from 1 to 480. Explain how the manager could take a systematic sample of size 30 from these workers.
- 6 The director of a sports club wants to take a sample of members. The members each have a unique membership number. There are 121 members who play cricket, 145 members who play hockey and 104 members who play squash. No members play more than one sport.
- a Explain how the director could take a simple random sample of 30 members and state one disadvantage of this sampling method.
- The director decides to take a stratified sample of 30 members.
- b State one advantage of this method of sampling.
- c Work out the number of members who play each sport that the director should select for the sample.

### 1.3 Non-random sampling

There are two types of non-random sampling that you need to know:

- Quota sampling
- Opportunity sampling

- **In quota sampling, an interviewer or researcher selects a sample that reflects the characteristics of the whole population.**

The population is divided into groups according to a given characteristic. The size of each group determines the proportion of the sample that should have that characteristic.

As an interviewer, you would meet people, assess their group and then, after interview, allocate them into the appropriate quota.

This continues until all quotas have been filled. If a person refuses to be interviewed or the quota into which they fit is full, then you simply ignore them and move on to the next person.

- **Opportunity sampling consists of taking the sample from people who are available at the time the study is carried out and who fit the criteria you are looking for.**

**Notation** Opportunity sampling is sometimes called **convenience sampling**.

This could be the first 20 people you meet outside a supermarket on a Monday morning who are carrying shopping bags, for example.

There are advantages and disadvantages of each type of sampling.

Quota sampling	
Advantages	Disadvantages
<ul style="list-style-type: none"> <li>• Allows a small sample to still be representative of the population</li> <li>• No sampling frame required</li> <li>• Quick, easy and inexpensive</li> <li>• Allows for easy comparison between different groups within a population</li> </ul>	<ul style="list-style-type: none"> <li>• Non-random sampling can introduce bias</li> <li>• Population must be divided into groups, which can be costly or inaccurate</li> <li>• Increasing scope of study increases number of groups, which adds time and expense</li> <li>• Non-responses are not recorded as such</li> </ul>

**Opportunity sampling****Advantages**

- Easy to carry out
- Inexpensive

**Disadvantages**

- Unlikely to provide a representative sample
- Highly dependent on individual researcher

**Exercise 1C**

- Interviewers in a shopping centre collect information on the spending habits from a total of 40 shoppers.
  - Explain how they could collect the information using:
    - quota sampling
    - opportunity sampling
  - Which method is likely to lead to a more representative sample?
- Describe the similarities and differences between quota sampling and stratified random sampling.
- An interviewer asks the first 50 people he sees outside a fish and chip shop on a Friday evening about their eating habits.
  - What type of sampling method did he use?
  - Explain why the sampling method may not be representative.
  - Suggest two improvements he could make to his data collection technique.
- A researcher is collecting data on the radio-listening habits of people in a local town. She asks the first 5 people she sees on Monday morning entering a supermarket. The number of hours per week each person listens is given below:
 

4     7     6     8     2

  - Use the sample data to work out a prediction for the average number of hours listened per week for the town as a whole.
  - Describe the sampling method used and comment on the reliability of the data.
  - Suggest two improvements to the method used.
- In a research study on the masses of wild deer in a particular habitat, scientists catch the first 5 male deer they find and the first 5 female deer they find.
  - What type of sampling method are they using?
  - Give one advantage of this method.

The masses of the sampled deer are listed below.

<b>Male (kg)</b>	75	80	90	85	82
<b>Female (kg)</b>	67	72	75	68	65

- Use the sample data to compare the masses of male and female wild deer.
- Suggest two improvements the scientists could make to the sampling method.



6 The heights, in metres, of 20 ostriches are listed below:

1.8, 1.9, 2.3, 1.7, 2.1, 2.0, 2.5, 2.7, 2.5, 2.6, 2.3, 2.2, 2.4, 2.3, 2.2, 2.5, 1.9, 2.0, 2.2, 2.5

- Take an opportunity sample of size five from the data.
- Starting from the second data value, take a systematic sample of size five from the data.
- Calculate the mean height for each sample.
- State, with reasons, which sampling method is likely to be more reliable.

**Hint** An example of an opportunity sample from this data would be to select the first five heights from the list.

## 1.4 Types of data

- Variables or data associated with numerical observations are called quantitative variables or quantitative data.**

For example, you can give a number to shoe size so shoe size is a quantitative variable.

- Variables or data associated with non-numerical observations are called qualitative variables or qualitative data.**

For example, you can't give a number to hair colour (blonde, red, brunette). Hair colour is a qualitative variable.

- A variable that can take any value in a given range is a continuous variable.**

For example, time can take any value, e.g. 2 seconds, 2.1 seconds, 2.01 seconds etc.

- A variable that can take only specific values in a given range is a discrete variable.**

For example, the number of girls in a family is a discrete variable as you can't have 2.65 girls in a family.

Large amounts of data can be displayed in a frequency table or as grouped data.

- When data is presented in a grouped frequency table, the specific data values are not shown. The groups are more commonly known as classes.**
  - Class boundaries tell you the maximum and minimum values that belong in each class.**
  - The midpoint is the average of the class boundaries.**
  - The class width is the difference between the upper and lower class boundaries.**

### Example 4

The lengths,  $x$  mm, to the nearest mm, of the forewings of a random sample of male adult butterflies are measured and shown in the table.

Length of forewing (mm)	Number of butterflies, $f$
30–31	2
32–33	25
34–36	30
37–39	13

- a State whether length is
- quantitative or qualitative
  - discrete or continuous.
- b Write down the class boundaries, midpoint and class width for the class 34–36.

- a i Quantitative  
ii Continuous
- b Class boundaries 33.5 mm, 36.5 mm  
Midpoint =  $\frac{1}{2}(33.5 + 36.5) = 35$  mm  
Class width =  $36.5 - 33.5 = 3$  mm

**Watch out** Be careful when finding class boundaries for continuous data. The data values have been rounded to the nearest mm, so the upper class boundary for the 30–31 mm class is 31.5 mm.



### Exercise 1D

- 1 State whether each of the following variables is qualitative or quantitative.
- Height of a tree
  - Colour of car
  - Time waiting in a queue
  - Shoe size
  - Names of pupils in a class
- 2 State whether each of the following quantitative variables is continuous or discrete.
- Shoe size
  - Length of leaf
  - Number of people on a bus
  - Weight of sugar
  - Time required to run 100 m
  - Lifetime in hours of torch batteries
- 3 Explain why:
- 'Type of tree' is a qualitative variable
  - 'The number of pupils in a class' is a discrete quantitative variable
  - 'The weight of a collie dog' is a continuous quantitative variable.
- 4 The distribution of the masses of two-month-old lambs is shown in the grouped frequency table.

Mass, $m$ (kg)	Frequency
$1.2 \leq m < 1.3$	8
$1.3 \leq m < 1.4$	28
$1.4 \leq m < 1.5$	32
$1.5 \leq m < 1.6$	22

**Hint** The class boundaries are given using inequalities, so the values given in the table are the actual class boundaries.

- Write down the class boundaries for the third group.
- Work out the midpoint of the second group.
- Work out the class width of the first group.



## Answers

### Prior knowledge 1

- Mean 5.89 (2 d.p.); Median 6; Mode 4; Range 10
  - Mean 18.38 (2 d.p.); Median 18.5; Mode 20; Range 9
- ANY TWO FROM: Overlapping categories; No option for > 4 hours; Question doesn't specify a period of time.  
How much TV do you watch each day?  
0–1 hours      2–3 hours      4 hours or more
- 29
  - 35
  - 38
  - 95

### Exercise 1A

- A census observes or measures every member of a population.
  - Advantage: will give a completely accurate result. Disadvantage: ANY ONE FROM: time consuming, expensive.
- The testing process will destroy the harness, so a census would destroy *all* the harnesses.
  - 250 kg is the median load at which the harnesses in the sample break. This means that half of the harnesses will break at a load less than 250 kg.
  - Test a larger number of harnesses.
- ANY ONE FROM:  
It would be expensive.  
It would be time consuming.  
It would be difficult.
  - A list of residents.
  - A resident.
- The testing process will destroy the microswitches, so a census would destroy *all* the switches.
  - The mean is less than the stated average but one of the switches lasted a significantly lower number of operations which suggests the median might be a better average to take – not affected by outliers. The data supports the company claim.
  - Test a larger number of microswitches.
- All the mechanics in the garage.
  - Everyone's views will be known.

### Exercise 1B

- Year 1: 8, Year 2: 12, Year 3: 16
  - ANY ONE FROM: sample accurately reflects the population structure of the school; guarantees proportional representation of different year groups in the sample.
- Patterns in the sample data might occur when taking every 20th person.
  - A simple random sample using the alphabetical list as the sampling frame.
- No: A systematic sample requires the first selected person to be chosen at random.
  - Take a simple random sample using the list of members as the sampling frame.
- Stratified sampling.
  - Male Y12: 10, Male Y13: 7, Female Y12: 12, Female Y13: 11
- $k = \frac{480}{30} = 16$   
Randomly select a number between 1 and 16. Starting with the worker with this clocking-in number, select the workers that have every 16th clocking-in number after this.
- Any method in which every member of the population has an equal chance of being selected, e.g. lottery. Disadvantage: the sample may not accurately reflect the proportions of members at the club who play each sport.

- The sample will have proportional representation of the members who play the different sports.
- Cricket: 10, Hockey: 12, Squash: 8

### Exercise 1C

- Divide the population into groups according to given characteristics. The size of each group determines the proportion of the sample that should have that characteristic. The interviewer assesses which group people fall into as part of the interview. Once a quota has been filled, no more people in that group are interviewed.
    - Opportunity sampling consists of taking the sample from the people who are available at the time the study is carried out, e.g. the first 40 shoppers who are available to be interviewed.
  - Quota sampling.
- Similarities: The population is divided according to the characteristics of the whole population (into strata for stratified sampling, and groups for quota sampling)  
Differences: Stratified sampling uses random sampling whereas quota sampling does not.
- Opportunity sampling
  - Sample is likely to be biased towards people who eat fish and chips on a Friday.
  - Survey people at different times of day. Survey people in other parts of the town, not outside the fish and chip shop.
- 5.4 hours
  - Opportunity sampling; unlikely to provide a representative sample of the town as a whole
  - Increase the number of people asked. Ask people at different times/in different locations.
- Quota sampling.
  - ANY ONE FROM: no sampling frame required, quick, easy, inexpensive, allows for comparison between male and female deer.
  - Males are on average heavier and have a greater spread.
  - Increase the sample size. Catch deer at random times during the day.
- Student's opportunity sample: For example, first five values
  - 1.9, 2.0, 2.6, 2.3, 2.0
  - 1.96 m, 2.16 m
  - Systematic sample – is random and likely to be more representative. Opportunity sample might get all the small values, for example.

### Exercise 1D

- Quantitative
  - Qualitative
  - Quantitative
  - Quantitative
  - Qualitative
- Discrete
  - Continuous
  - Discrete
  - Continuous
  - Continuous
  - Continuous
- It is descriptive rather than numerical.
  - It is quantitative because it is numerical. It is discrete because its value must be an integer; you cannot have fractions of a pupil.
  - It is quantitative because it is numerical. It is continuous because weight can take any value in a given range.

- 4 a 1.4 kg and 1.5 kg    b 1.35 kg  
c 0.1 kg

### Exercise 1E

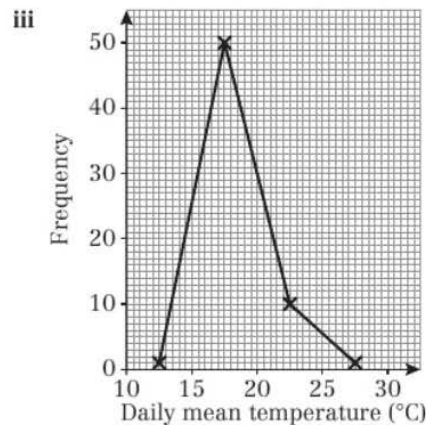
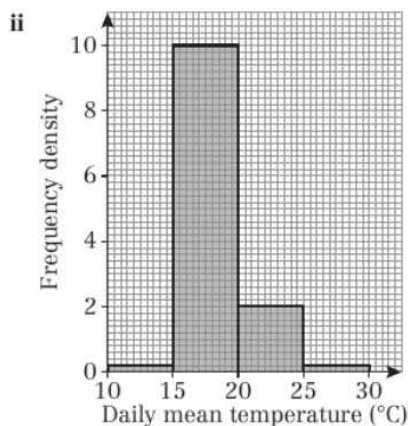
- 1 a Leuchars  
b Perth  
c ANY ONE FROM: Leeming, Heathrow, Beijing  
d ANY ONE FROM: Leuchars, Hurn, Camborne, Jacksonville, Perth  
e ANY ONE FROM: Beijing, Jacksonville, Perth
- 2 Continuous – it can take any value in the range 0 to 100
- 3 a i 10.14 hours    ii 7.6 hours  
b i 9.5 hours    ii 12.8 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
$10 \leq t < 15$	1
$15 \leq t < 20$	50
$20 \leq t < 25$	10
$25 \leq t < 30$	1



2 Students' own answer.

### Mixed exercise 1

- 1 a 9.6°C  
b Sampling frame: first 15 days in May 1987  
Allocate each date a number from 1 to 15  
Use the random number function on calculator to generate 5 numbers between 1 and 15  
c Students' own answers.  
d 10.8°C
- 2 a i Advantage: very accurate; disadvantage: expensive (time consuming).  
ii Advantage: easier data collection (quick, cheap); disadvantage: possible bias.  
b Assign unique 3-digit identifiers 000, 001, ..., 499 to each member of the population. Work along rows of random number tables generating 3-digit numbers. If these correspond to an identifier then include the corresponding member in the sample; ignore repeats and numbers greater than 499. Repeat this process until the sample contains 100 members.
- 3 a i Collection of individual items.  
ii List of sampling units.  
b i List of registered owners from DVLC.  
ii List of people visiting a doctor's clinic in Oxford in July 1996.
- 4 a Advantage – the results are the most representative of the population since the structure of the sample reflects the structure of the population.  
Disadvantage – you need to know the structure of the population before you can take a stratified sample.  
b Advantage – quick and cheap.  
Disadvantage – can introduce bias (e.g. if the sample, by chance, only includes very tall people in an investigation into heights of students).
- 5 a People not in office not represented.  
b i Get a list of the 300 workers at the factory.  
 $\frac{300}{30} = 10$  so choose one of the first ten workers on the list at random and every subsequent 10th worker on the list, e.g. if person 7 is chosen, then the sample includes workers 7, 17, 27, ..., 297.  
ii The population contains 100 office workers ( $\frac{1}{3}$  of population) and 200 shop floor workers ( $\frac{2}{3}$  of population).  
The sample should contain  $\frac{1}{3} \times 30 = 10$  office workers and  $\frac{2}{3} \times 30 = 20$  shop floor workers. The 10 office workers in the sample should be a simple random sample of the 100 office workers. The 20 shop floor workers should be a simple random sample of the 200 shop floor workers.

