

## Section Check In – 2.01 Sampling

### Questions

- David wants to generate 20 two-digit random numbers. Give one method he could use.
- Which of the following is the best definition of a simple random sample?
  - A sample created using random numbers.
  - A sample where the simplest method of sampling is used.
  - A sample where each member of the population is surveyed.
  - A sample that is representative of the population.
  - A sample where each possible sample of the given size has an equal chance of being selected.
- A company director wants to find out the views of all her employees. She chooses 20 people from those at work on Monday morning and sends them an online questionnaire to complete. Define the population in this situation.
- 40 students, sampled from a population of 1000 students, are asked about how much sleep they get. This table shows how many students from each year were sampled and the mean hours sleep for each year group.

Year group	Number in year	Number sampled ( $n$ )	Mean hours sleep per night
7	250	10	8.5
8	225	9	8.0
9	175	7	7.0
10	150	6	6.5
11	200	8	9.0

- Give the name of this type of sample.
  - Use the data to estimate the mean number of hours sleep in the whole population. Give your answer to 1 decimal place.
- A local radio breakfast programme conducts a phone-in poll between 6am and 9am to find out people's views on the frequency of rubbish collections by the council. Give two reasons why the results of this poll may not be representative of local views.
  - Mr Smith wants to create a systematic random sample of 10 houses from a street of 120 houses. Explain how he could do this.

7. A town council wants to inspect a random sample of taxicabs to check that they are meeting the council regulations. The list of registered taxicabs is shown in this table.

Licence number	Company	Licence number	Company
1	Red Cars	21	Alpha+++
2	NG Cabs	22	Red Cars
3	Castle Cars	23	Red Cars
4	Castle Cars	24	Red Cars
5	Mr A Smith	25	Red Cars
6	Red Cars	26	Prem Jarosz
7	Red Cars	27	Mr S Hussain
8	Mrs Woods	28	NG Cabs
9	Gurprit Singh	29	Alpha+++
10	NG Cabs	30	NG Cabs
11	NG Cabs	31	NG Cabs
12	NG Cabs	32	Rashid Khan
13	Rashid	33	Alpha+++
14	Castle Cars	34	Alpha+++
15	Sarah Kemp	35	Red Cars
16	NG Cabs	36	Piet Janssen
17	Alpha+++	37	Castle Cars
18	NG Cabs	38	Castle Cars
19	Alpha+++	39	Castle Cars
20	Castle Cars	40	Red Cars

Use the random digits below to construct a simple random sample of **five** of these taxicabs for the town council.

01164363183767475061263207510010431204181922891792

8. Ecologists want to survey an area of land that measures  $1 \text{ km} \times 1 \text{ km}$ .

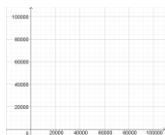
To do this, they will randomly sample one hundred  $2 \text{ m} \times 2 \text{ m}$  squares within the area and count the number of plant and animal species within each square. The sampling must satisfy the following rules:

- The squares in the sample must not overlap.
- Each square must lie entirely within the  $1 \text{ km} \times 1 \text{ km}$  area.
- **All** possible  $2 \text{ m} \times 2 \text{ m}$  squares within the area should have the same probability of being in the sample, allowing for measurements being accurate to the nearest centimetre.

The ecologists model the area of land using a coordinate system, and then generate random numbers to identify the sample squares. The coordinate system and some of the random numbers are shown below.

Model of area

Random numbers				
31980	21582	32044	21706	65194
99926	81227	65408	50360	27561
93047	44031	12173	86999	39567
48715	66690	90152	93484	01311



The ecologists use these random numbers to identify the bottom left-hand coordinates of the first three squares in their sample as follows:

1st (31980, 21582)

2nd (32044, 21706)

3rd (65194, 99926)

Evaluate this outcome and advise the ecologists appropriately.

9. A market researcher is employed by a company to conduct a survey about a new brand of chocolate. The company wants to find out whether people's preferences differ according to their **age** and **gender**.  
The researcher will select 144 participants from people shopping in a city centre on a Saturday. Each participant will taste three different chocolate products and say which one they prefer.  
Suggest a suitable sampling method and explain exactly how the method could be used to collect appropriate data.
10. A national medical organisation wants to interview ambulance drivers to find out the main causes of delays in responding to emergency calls. They can afford to use 6 researchers and want all the interviews to be conducted on a single day.  
Suggest a suitable sampling method and explain briefly how the sample might be created.

### Extension

You will need to refer to p15 of the following resource <http://www.cimt.org.uk/cmmss/S5/Text.pdf> and a means of generating random numbers.

The fish are representative of a large fish population.

The number on the body of each fish is its tag or identifier. The number at the tail of each fish is its weight in kilograms.

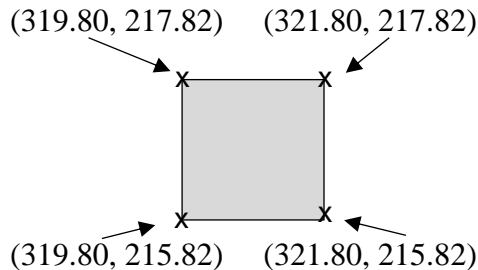
You are going to take two samples of 20 fish and find the mean weights of your samples. Because the 57 fish are representative of a large fish population, you may include the same fish twice in a sample.

- (i) Select a sample of 20 fish that you think is representative.  
Find and record the mean weight of the fish in the sample.
- (ii) Generate 20 random numbers between 1 and 57. Use them to select 20 fish.  
Find and record the mean weight of the fish in the sample.
- (iii) Calculate the true mean weight of all the 57 fish.
- (iv) Which sample gave you the best estimate of the true mean? Why do you think this might be?
- (v) Draw a grouped frequency graph representing the weights of the 57 fish.  
What do you notice?

**Worked solutions**

1. Use random number table, generate random numbers on a calculator or use two 10-sided fair dice with faces numbered 0 – 10. Or any other method that gives rise to a random sample, such as randomly choosing numbered counters from a bag.
2. E is the correct definition.  
A: does not need to be generated using numbers – could select objects from a bag.  
B: the “simple” does not refer to the method being simple. How would we identify the simplest method?  
C: this is a census, not a sample  
D: whether an individual sample is representative of the population depends on the size and nature of the sample.
3. The population is all the director’s employees. (Those in work on Monday morning are the sampling frame and those selected are the sample).
4. (i) A stratified sample: in this stratified sample, the number sampled from each stratum (year group) is the same proportion of the year group, that is,  $\frac{1}{25}$ . This is proportional stratified sampling; it would still have been stratified sampling if the sampling had not been proportional.  
(ii) A weighted mean should be calculated by finding the total hours for each year ( $n \times \text{mean}$ ), adding them up and dividing by 40. This gives an overall mean of 7.925 hours, or 7.9 to 1 d.p.
5. Not everyone will be available to phone between 6am and 9am.  
Those who phone are likely to be those who feel strongly.  
Only those who listen to the programme will be aware of the phone poll.
6. Explanation must involve selecting a random house from houses 1 to 10 and then every 12th house after that. If the first house is not randomised and we always start at the first house, most of the houses are excluded from the sample at the outset.
7. Taking the digits in pairs gives the licence numbers 1, 16, 18, 37 and 26.  
01 is a two digit representation of a single digit number.  
43, 63, 67, 47, 50, 61 are ignored as too high.

8. The scales are in cm as  $100\,000\text{ cm} = 1\text{ km}$ . Therefore  $(31980, 21582)$  represents the point in the area that is  $319.80\text{ m}$  along and  $215.82\text{ m}$  up. The resulting square will be



The second square will therefore overlap with the first square.

The third square lies outside the  $1\text{ km} \times 1\text{ km}$  area as  $999.26 + 2\text{ m} = 1001.26\text{ m} = 1.00126\text{ km}$

The ecologists need to check each sample meets the rules and ignore any samples that do not.

The first three coordinate pairs should be  $(31980, 21582)$ ,  $(81227, 65408)$  and  $(50360, 27561)$ .

The sampling method **does** meet the need for all possible squares to initially have the same probability of being in the sample. Splitting the area into a grid of  $2\text{ m} \times 2\text{ m}$  squares would not achieve this.

9. The researcher should use a quota sample. The quotas might be as shown in the table.

	18 to 30 years	31 to 50 years	Over 50 years	Total
Male	24	24	24	72
Female	24	24	24	72
Total	48	48	48	144

The researcher should ask people to take part in the survey until they have sufficient data for each category of participant. Once a quota is filled, they stop sampling people in that category.

10. They should use a cluster sample. For example, split the country into appropriate ambulance authorities and randomly sample 6 authorities. Then within each of these take a further sample of those to be surveyed. This might have to be an opportunity sample, or could perhaps be a stratified sample.

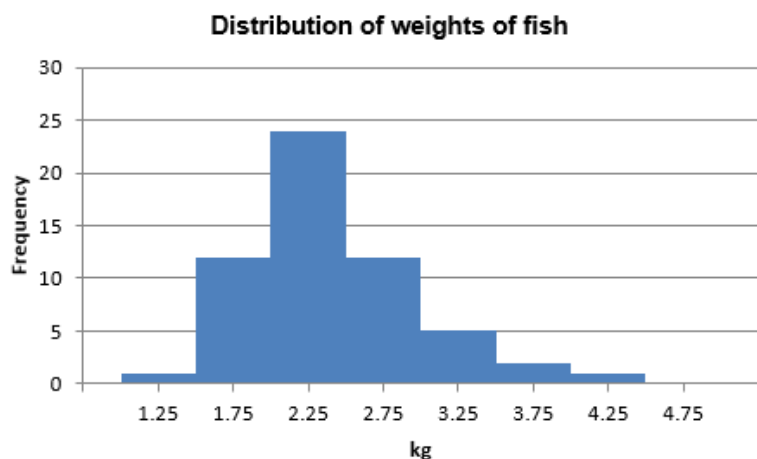
**Extension**

This is a good demonstration of the importance of random sampling.

Student responses will vary for (i) and (ii) based on their sample selections.

If the whole class does the extension activity, the results can be presented in a back-to-back stem and leaf diagram comparing the distributions of means from the two different methods. (The data will need to be grouped).

- (iii) The true mean of all the 57 fish is 2.43.
- (iv) Random samples from this data usually give an estimate that is closer to the true mean than the samples selected by eye. This is because we tend to include too many of the larger fish in the sample when selecting by eye in an attempt to be representative, as we expect the population to be normally distributed.
- (v) The graph below shows the distribution of the weights of the 57 fish. The distribution is skewed rather than symmetrical.



We'd like to know your view on the resources we produce. By clicking on '[Like](#)' or '[Dislike](#)' you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click 'Send'. Thank you.

Whether you already offer OCR qualifications, are new to OCR, or are considering switching from your current provider/awarding organisation, you can request more information by completing the Expression of Interest form which can be found here:

[www.ocr.org.uk/expression-of-interest](http://www.ocr.org.uk/expression-of-interest)

Looking for a resource? There is now a quick and easy search tool to help find free resources for your qualification:

[www.ocr.org.uk/i-want-to/find-resources/](http://www.ocr.org.uk/i-want-to/find-resources/)

### **OCR Resources: *the small print***

OCR's resources are provided to support the delivery of OCR qualifications, but in no way constitute an endorsed teaching method that is required by the Board, and the decision to use them lies with the individual teacher. Whilst every effort is made to ensure the accuracy of the content, OCR cannot be held responsible for any errors or omissions within these resources.

© OCR 2017 - This resource may be freely copied and distributed, as long as the OCR logo and this message remain intact and OCR is acknowledged as the originator of this work.

OCR acknowledges the use of the following content: n/a

Please get in touch if you want to discuss the accessibility of resources we offer to support delivery of our qualifications: [resources.feedback@ocr.org.uk](mailto:resources.feedback@ocr.org.uk)