



- 1 A bakery makes 2000 loaves of bread each day. The manager wants to check that the loaves contain the right amount of salt. Each loaf tested is crushed and analysed in a laboratory.

- State the population in this situation.
- Suggest a suitable sampling frame.
- Explain why the manager should take a sample rather than a census.

(3 marks)

a) Population = every item we want to know about

All 2000 loaves made that day

b) Sampling frame = a list of every sampling unit

Number the loaves 1 to 2000

(e.g. by batch and tray position)

c) Think about what testing does

Testing crushes the loaf

A census would destroy all 2000 loaves
so nothing would be left to sell

- 2 A leisure centre has 480 members. The manager wants a simple random sample of 30 members to ask about new opening hours.

Describe how the manager could take this sample.

(3 marks)

Sampling frame

List all 480 members

Number them 001 to 480

Random selection

Use a random number generator to give numbers 1 to 480

Ignore repeats (and any number out of range)

Keep going until 30 different numbers are picked

Sample

Ask the 30 members with those numbers

3

Name the sampling method used in each case.

- a. A dentist picks a random number between 1 and 20, then contacts that patient on her list and every 20th patient after it.
- b. A reporter stands outside a station at 8 am and interviews the first 25 people who agree to talk to her.
- c. An interviewer is told to question 15 men and 15 women aged over 60, and stops each group when it is full.

(3 marks)

a) Random start, then every k th item

Systematic sampling

b) Whoever is there and willing

Opportunity (convenience) sampling

c) Fixed numbers per group, not chosen at random

Quota sampling

4

The numbers of students in each year group at a secondary school are shown in the table.

Year group	7	8	9	10	11
Number of students	186	174	168	162	150

The head teacher wants a stratified sample of 60 students.

- Find the number of students that should be chosen from each year group.
- Describe how the students in Year 9 should then be chosen.

(4 marks)

a) Total population

$$N = 186 + 174 + 168 + 162 + 150$$

$$N = 840$$

Each stratum: $\frac{\text{size}}{840} \times 60$

$$\text{Year 7} = \frac{186}{840} \times 60 = 13.29 \rightarrow 13$$

$$\text{Year 8} = \frac{174}{840} \times 60 = 12.43 \rightarrow 12$$

$$\text{Year 9} = \frac{168}{840} \times 60 = 12$$

$$\text{Year 10} = \frac{162}{840} \times 60 = 11.57 \rightarrow 12$$

$$\text{Year 11} = \frac{150}{840} \times 60 = 10.71 \rightarrow 11$$

Check the total

$$13 + 12 + 12 + 12 + 11 = 60$$

b) Simple random sample within the stratum

Number the 168 Year 9 students 1 to 168

Use a random number generator to pick 12 different numbers

5

An accountant has 600 invoices, numbered 1 to 600, and wants to check a sample of 40 of them.

- Describe how she could take a systematic sample of 40 invoices.
- Give one advantage of systematic sampling in this situation.
- Explain why a systematic sample is not a simple random sample.

(4 marks)

a) Find the interval

$$k = \frac{600}{40} = 15$$

Pick a random number from 1 to 15 as the start, say 7

Then take invoices 7, 22, 37, 52, ... (every 15th)

This gives 40 invoices

b) Advantage

Quick and easy, and spread evenly through the list

c) Simple random: every sample of 40 equally likely

After the start, the other 39 invoices are fixed

e.g. invoices 1 and 2 can never both be chosen

6

A market researcher in a shopping centre is told to interview 40 adults under 30 and 60 adults aged 30 or over. She approaches shoppers until each group is full.

- State the name of this sampling method.
- Give one advantage and one disadvantage of this method.
- Explain what extra information she would need to take a stratified sample instead.

(4 marks)

a) Set numbers per group, not random

Quota sampling

b) One of each

Advantage: quick and cheap, no sampling frame needed

Disadvantage: not random, so it may be biased

(she chooses who to ask, and only shoppers there that day)

c) Stratified needs random choice in each group

A sampling frame: a list of the population

split into under 30 and 30 or over

7

Ravi is studying Daily Total Rainfall for Leeming in the large data set.

- State the units of Daily Total Rainfall.
- Some days are recorded as tr. State what this means.
- Explain how Ravi should deal with the tr and n/a entries before he calculates the mean Daily Total Rainfall.

(4 marks)

a)

millimetres (mm)

b) tr = trace

Less than 0.05 mm of rain fell that day

c) Clean the data first

tr: replace with 0 (it is almost no rain)

n/a: data not available, so remove those days

Then find the mean of the days that are left

8

Holly downloads two columns from the large data set for Heathrow in 2015 but forgets to copy the headings.

- Column P contains whole numbers, all between 0 and 8.
- Column Q contains values between 996 and 1031.

- Suggest which variable is in column P and state its units.
- Suggest which variable is in column Q and state its units.

(4 marks)

a) Whole numbers from 0 to 8 = eighths of the sky

Daily Mean Total Cloud Cover

units: oktas

b) Values near 1000 = air pressure

Daily Mean Pressure

units: hectopascals (hPa)

9

Owen wants to show that summer days are warmer in Perth than in the UK. He compares the Daily Mean Temperature for Perth with the Daily Mean Temperature for Hurn, using the large data set for June and July 2015.

- Using your knowledge of the large data set, explain why this comparison does not support Owen's aim.
- Suggest how Owen could change his choice of data to make a fairer comparison of summer temperatures using only the large data set.

(4 marks)

a) Where is Perth?

Perth is in Australia, in the southern hemisphere

so June and July are winter in Perth

He is comparing UK summer with Perth winter

b) Data set covers May to October only

There is no Perth summer data to use

Compare Hurn with Beijing or Jacksonville instead

(northern hemisphere, so June and July are summer too)

10

Zainab is studying Daily Total Sunshine, in hours, for Camborne in August 1987. She takes the first 8 days of the month as her sample.

Day	1	2	3	4	5	6	7	8
Sunshine (hours)	6.2	0.4	9.8	11.0	3.1	0.0	7.5	5.2

- State the name of the sampling method Zainab used.
- Calculate the mean of her sample.
- The mean Daily Total Sunshine for all 31 days of August 1987 in Camborne was 5.8 hours.
Comment on her sample mean.
- Describe how Zainab could take a simple random sample of 8 days from August 1987 instead.

(5 marks)

a) First 8 days = the easiest to take

Opportunity sampling

$$b) \text{Mean} = \frac{\sum x}{n}$$

$$\sum x = 6.2 + 0.4 + 9.8 + 11.0 + 3.1 + 0.0 + 7.5 + 5.2$$

$$\sum x = 43.2$$

$$\text{mean} = \frac{43.2}{8}$$

$$\text{mean} = 5.4 \text{ hours}$$

c) Compare with 5.8

$5.4 < 5.8$, so the sample underestimates the mean

Days in a row often have similar weather

so the sample may not represent the whole month

d) Simple random sample

Number the days 1 to 31

Pick 8 different numbers with a random number generator

Ffion wants a sample of 32 days of Daily Mean Wind Speed from the five UK stations in the large data set for 2015. After removing days recorded as n/a, the number of days left at each station is shown below.

Station	Camborne	Heathrow	Hurn	Leeming	Leuchars
Days available	175	165	160	150	150

- State the units of Daily Mean Wind Speed in the large data set.
- Find how many days Ffion should take from each station for a stratified sample of 32 days.
- Describe how she should choose the days for Hurn.
- Give one reason why a stratified sample is better here than taking all 32 days from Camborne.

(6 marks)

a)

knots (kn)

b) Total days

$$N = 175 + 165 + 160 + 150 + 150$$

$$N = 800$$

Each station: $\frac{\text{days}}{800} \times 32$

$$\text{Camborne} = \frac{175}{800} \times 32 = 7$$

$$\text{Heathrow} = \frac{165}{800} \times 32 = 6.6 \rightarrow 7$$

$$\text{Hurn} = \frac{160}{800} \times 32 = 6.4 \rightarrow 6$$

$$\text{Leeming} = \frac{150}{800} \times 32 = 6$$

$$\text{Leuchars} = \frac{150}{800} \times 32 = 6$$

Check

$$7 + 7 + 6 + 6 + 6 = 32$$

c) Random within the stratum

Number the 160 Hurn days 1 to 160

Pick 6 different numbers with a random number generator

d) Reason

Wind speed varies between stations

(Camborne is on the coast in the far south west)

so stratifying represents every station in proportion

A town council wants to know what its 24 000 adult residents think about a new cycle lane. A councillor suggests two plans.

- **Plan 1:** hand a questionnaire to every adult who walks past the town hall on a Monday morning until 200 have been completed.
- **Plan 2:** use the electoral register, which lists the adults in each of the town's 4 wards, and take a stratified sample of 200.

- Give two reasons why Plan 1 is likely to give a biased sample.
- Ward A has 7200 adults. Find how many adults from Ward A would be in the Plan 2 sample.
- Explain why Plan 2 is likely to give more reliable results than Plan 1.
- Give one practical problem with Plan 2.

(6 marks)

a) Two sources of bias in Plan 1

1. Only people near the town hall are asked

(people living further away are missed)

2. Monday morning misses people at work

so some groups of adults are under-represented

b) Proportional to ward size

$$\text{Ward A} = \frac{7200}{24000} \times 200$$

$$\text{Ward A} = 0.3 \times 200$$

$$\text{Ward A} = 60$$

c) Compare the plans

Plan 2 has a sampling frame of every adult

People are chosen at random, not by who walks past

Every ward is represented in proportion

so it is much less likely to be biased

d) A practical problem

Some people chosen may not reply (non-response)