

QQQ – Statistics Year 1 – Chapter 1 – Data Collection (Version 1)

Total Marks: 24

(24 = Platinum, 22 = Gold, 19 = Silver, 16 = Bronze)

1. A survey is undertaken to find out the most popular political party in London.

The first 1100 available people from London are surveyed.

Identify the name of this type of sampling.

Circle your answer.

[1 mark]

- simple random
- opportunity
- stratified
- quota

(Total 1 marks)

AQA June 2019 Paper 2 Q11

2. An electoral register contains 8000 names.

A researcher decides to select a systematic sample of 100 names from the register.

Explain how the researcher should select such a sample.

[3 marks]

(Total 3 marks)

AQA June 2021 Paper 3 Q12

3.

(a) Explain what you understand by a census.

(1)

Each cooker produced at GT Engineering is stamped with a unique serial number.

GT Engineering produces cookers in batches of 2000. Before selling them, they test a random sample of 5 to see what electric current overload they will take before breaking down.

(b) Give one reason, other than to save time and cost, why a sample is taken rather than a census.

(1)

(c) Suggest a suitable sampling frame from which to obtain this sample.

(1)

(d) Identify the sampling units.

(1)

(Total 4 marks)

Edexcel S2 January 2008 Q1

4. The headteacher of a school wishes to collect the opinions of the students on a new timetable structure.

To do this, a random sample of size 50, stratified by year group, will be selected.

The school has a total of 720 students.

The number of students in each of the year groups at this school is shown below.

Year group	10	11	12	13
Number of students	200	240	150	130

- (a) Find the number of students from each year group that should be selected in the stratified random sample. **[3 marks]**
- (b) State **one** advantage of using a stratified random sample. **[1 mark]**

(Total 4 marks)

5. Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

Outcome	<i>A</i>	<i>B</i>	<i>C</i>
Frequency	16	2	1

(a) State the sampling technique Helen used. (1)

(b) From your knowledge of the large data set

(i) suggest which variable was being studied,

(ii) state the name of outcome *A*. (2)

George is also studying the same variable from the large data set for Heathrow from 2015. He started with the data from 5th May and then took every 10th reading and obtained the following

Outcome	<i>A</i>	<i>B</i>	<i>C</i>
Frequency	16	1	1

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following

Outcome	<i>A</i>	<i>B</i>	<i>C</i>
Frequency	155	26	3

(c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015. (1)

(Total 4 marks)

6. A lake contains 3 species of fish. There are estimated to be 1400 trout, 600 bass and 450 pike in the lake. A survey of the health of the fish in the lake is carried out and a sample of 30 fish is chosen.

- (a)** Give a reason why stratified random sampling cannot be used. **(1)**
- (b)** State an appropriate sampling method for the survey. **(1)**
- (c)** Give one advantage and one disadvantage of this sampling method. **(2)**
- (d)** Explain how this sampling method could be used to select the sample of 30 fish. You must show your working. **(4)**

(Total 8 marks)

Solutions

1.

Scheme	Marks	AO
Circles correct answer	B1	1.2
	(1 mark)	

Typical solution

opportunity

2.

Scheme	Marks	AO
Explains how to enumerate population using valid numbering stating range used	E1	2.4
Calculates $8000 \div 100$ or 80 seen	B1	1.1b
Explains that they randomly select the first person from the first 80 people and thereafter every subsequent 80th person should be selected	E1	2.4
	(3 marks)	

Typical solution

Give each name a number from 1 to 8000

$$\frac{8000}{100} = 80$$

Randomly select a number between 1 and 80 and select every 80th person on the register from the first person selected

3.

Mark scheme (a)

Scheme	Marks
A census is when <u>every member</u> of the <u>population</u> is investigated.	B1

	(1)
--	-----

Notes

B1 Need one word from each group

(1) Every member /all items / entire /oe

(2) population/collection of individuals/sampling frame/oe

enumerating the population on its own gets B0

Mark scheme (b)

Scheme	Marks
There would be no cookers left to sell.	B1
	(1)

Notes

B1 Idea of Tests to destruction. Do not accept cheap or quick

Mark scheme (c)

Scheme	Marks
A list of the unique identification numbers of the cookers.	B1
	(1)

Notes

B1 Idea of list/ register/database of cookers/serial numbers

Mark scheme (d)

Scheme	Marks
A cooker	B1
	(1)
	(4 marks)

Notes

B1 cooker(s) / serial number(s)

The sample of 5 cookers or every 400th cooker gets B1

4.

Mark scheme (a)

Scheme	Marks	AO
Uses at least one correct ratio PI by at least one correct value	M1	3.1a
Obtains at least 2 correct values Condone values given as decimals	A1	1.1b
Obtains all 4 correct integer values Y10: 14 Y11: 17 Y12: 10 Y13: 9 CAO	A1	3.2a
	(3)	

Typical solution

$$\text{Year 10: } \frac{200}{720} \times 50 = 13.9$$

$$\text{Year 11: } \frac{240}{720} \times 50 = 16.7$$

$$\text{Year 12: } \frac{150}{720} \times 50 = 10.4$$

$$\text{Year 13: } \frac{130}{720} \times 50 = 9.0$$

Y10: 14

Y11: 17

Y12: 10

Y13: 9

Mark scheme (b)

Scheme	Marks	AO
Gives any correct advantage of a stratified sample Accept: Provides greater precision/reliability	E1	2.4

The sample is representative of the population/each (year) group Unbiased		
	(1)	
	(4 marks)	

Typical solution

Ensures each year group is included in the sample

5.

Mark scheme (a)

Scheme	Marks	AO
Systematic (sampling)	B1	1.2
	(1)	

Notes

B1 for identifying the correct sampling technique

Allow slight misspelling e.g. “sysmatic”, “sytmatic”

Do NOT allow “systemic”

Mark scheme (b)

Scheme	Marks	AO
(i) [Daily Mean] Wind Speed	B1	2.2a
(ii) Light	B1	1.2
	(2)	

Notes

(i) B1 for identifying appropriate qualitative variable. {LDS mark}

Allow “Wind speed” or “Wind strength” but NOT just “wind” or “wind direction”

(ii) B1 for realising that modal wind speed is “Light” {LDS mark}

Allow just “light” or “most light”

NB These two B marks are independent so can score B0B1 for e.g. “rainfall” and “light”

Mark scheme (c)

Scheme	Marks	AO
Variable A occurs most (around 80~90%) of the time	B1	2.2b
	(1)	
	(4 marks)	

Notes

B1 for inferring that frequency of A can be estimated fairly reliably: {underestimates B and over estimates C }

e.g. “ A is the most frequent” [can then ignore comments about B and C]

6.

Mark scheme (a)

Scheme	Marks
Sampling frame within each species of fish in the lake impossible to obtain.	B1
	(1)

Notes

‘You can’t / it’s very difficult to number all the fish’ or equivalent

Mark scheme (b)

Scheme	Marks
Quota sampling	B1
	(1)

Mark scheme (c)

Scheme	Marks
Advantages: Sample can be obtained quickly Costs are kept to a minimum Administration of survey is easy	B1
Disadvantages: Not possible to estimate sampling errors	B1

Process not random Surveyor may not be able to identify species of fish easily	
	(2)

Notes

Correct answer to (b) required. Some detail required.

Mark scheme (d)

Scheme	Marks								
<table border="1"> <thead> <tr> <th>Species</th><th>Quota</th></tr> </thead> <tbody> <tr> <td>Trout</td><td>$\frac{1400}{2450} \times 30 = 17.14$</td></tr> <tr> <td>Bass</td><td>$\frac{600}{2450} \times 30 = 7.35$</td></tr> <tr> <td>Pike</td><td>$\frac{450}{2450} \times 30 = 5.51$</td></tr> </tbody> </table> Fish are caught from the lake until the quota of 17 trout, 7 bass and 6 pike are reached.	Species	Quota	Trout	$\frac{1400}{2450} \times 30 = 17.14$	Bass	$\frac{600}{2450} \times 30 = 7.35$	Pike	$\frac{450}{2450} \times 30 = 5.51$	B1B1B1
Species	Quota								
Trout	$\frac{1400}{2450} \times 30 = 17.14$								
Bass	$\frac{600}{2450} \times 30 = 7.35$								
Pike	$\frac{450}{2450} \times 30 = 5.51$								
If a fish is caught and the species quota is full, then this is ignored.	B1								
	(4)								
	(8 marks)								

Notes

1st B1 any one correct calculation seen or implied

2nd B1 all correct to at least 1 dp

3rd B1 for 17,7,6

4th B1 accept equivalent statement. Require comment on what to do with 'extra fish'.