A company director decides to survey staff about changes to the company calendar. The company has staff in 4 different job roles

72 managers, 108 drivers, 180 administrators and 360 warehouse staff.

The director decides to take a stratified sample.

(a) Write down one advantage of using a stratified sample rather than a simple random sample for this survey.

(1)

(b) Find the number of staff in each job role that will be included in a stratified sample of 40 staff.

(3)

(c) Describe how to choose managers for the stratified sample.

(2)

(a) State two reasons why stratified sampling might be a more suitable sampling method than simple random sampling.

(2)

(b) State two reasons why stratified sampling might be a more suitable sampling method than quota sampling.

(2)

1(a)	-(accurate) estimates for each strata / job -**more** representative of the population -reflects population structure	Any 1 oe	B1
			(1)
(b)	Total staff=720	May be implied by calculations	B1
	$\text{Managers} = \frac{72}{720} \times 40 = 4$	For one correct calculation, follow through their 720.	M1
	$\text{Drivers} = \frac{108}{720} \times 40 = 6$		
	$\text{Administrators} = \frac{180}{720} \times 40 = 10$		
	$\text{Warehouse} = \frac{360}{720} \times 40 = 20$	4, 6, 10, 20 only. Must identify which job the values relate to.	A1
			(3)
(c)	Label all managers 1 – 72 o.e.	Idea of **sampling frame** or **list of managers**. Need not give the specific term.	B1
	Using random numbers in range 1-72 or 0-71 select 4 (managers).	Use of **random numbers** to select required number of managers. Must mention use of random numbers or some **random** selection process. If they are describing systematic sampling score B0.	B1
			(2)

1(a) e.g.	Analyse / find estimates for a particular **subgroup** of the population. Stratified guarantees **representation** of all **groups**, srs does not. Observe **relationships between subgroups** – srs does not guarantee equal or proportionate representation. Rare or extreme cases as part of a **small subgroups** can be represented **proportionately** in stratified i.e. stratified represents the structure of the population– srs does not allow this. Stratified typically require **large sample size** compared to srs due to lower variability within subgroups compared to entire population. Any 2 distinct reasons	B1B1 (2)
(b) e.g.	It (a stratified sample) is **not biased** as the members are chosen randomly. You can **estimate the sampling errors** (for a stratified sample) It (a stratrified sample) gives **more accurate estimates** as it is a random process. A quota sample may be (**interviewer / process**) **biased**. It's **not possible to estimate/find the sampling errors** for a quota sample (whereas you can for a stratified sample) Any 2 distinct reasons	B1B1 (2) **Total 4**

(a) Explain what you understand by a random sample from a finite population.

(1)

(b) Give an example of a situation when it is not possible to take a random sample.

(1)

A college lecturer specialising in shoe design wants to change the way in which she organises practical work.

She decides to gather ideas from her 75 students.

She plans to give a questionnaire to a random sample of 8 of these students.

(c) (i) Describe the sampling frame that she should use.

(ii) Explain in detail how she should use a table of random numbers to obtain her sample.

(3)

A gym club has 400 members of which 300 are males.

Explain clearly how a stratified sample of size 60 could be taken.

(3)

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)

1(a)	(This is a sample where) **every** (possible) **sample** (of size n) has an **equal chance** of being chosen.	B1	(1)
(b)	'When it is impossible to provide a **sampling frame**' or a correct example with an indication of sampling frame being impossible.	B1	(1)
(c)(i)	**A list/register of all** the students.	B1	
(ii)	Number the students (from 0 to 74, 1 to 75 etc.) Using the **random no. table** read off the nos. and **identify or select the students allocated those nos**.	B1 B1	(3) **Total 5**

1.	Label females 1 – 100 (or 0 – 99) and males 1 – 300 (or 0 – 299) Using *random numbers* for each group in range 1 – 100 (0 – 99) select 15 females and using 1 – 300 (or 0 – 299) select 45 males	B1 B1 B1 **[Total 3]**

#

2(a)	Sampling frame within each species of fish in the lake impossible to obtain.	B1 **(1)**
2(b)	Quota sampling	B1 **(1)**
2(c)	Advantages: Sample can be obtained quickly Costs are kept to a minimum Administration of survey is easy Disadvantages: Not possible to estimate sampling errors Process not random Surveyor may not be able to identify species of fish easily	B1 B1 **(2)**
2(d)	Fish are caught from the lake until the quota of 17 trout, 7 bass and 6 pike are reached. If a fish is caught and the species quota is full, then this is ignored.	B1B1B1 B1 **(4)**

Species	Quota
Trout	$\frac{1400}{2450}\times 30 = 17.14$
Bass	$\frac{600}{2450}\times 30 = 7.35$
Pike	$\frac{450}{2450}\times 30 = 5.51$

A college manager wants to survey students' opinions of enrichment activities.
She decides to survey the students on the courses summarised in the table below.

Course	Number of students enrolled
Leisure and Sport	420
Information Technology	337
Health and Social Care	200
Media Studies	43

Each student takes only one course.

The manager has access to the college's information system that holds full details of each of the enrolled students including name, address, telephone number and their course of study. She wants to compare the opinions of students on each course and has a generous budget to pay for the cost of the survey.

(a) Give one advantage and one disadvantage of carrying out this survey using

(i) quota sampling,

(ii) stratified sampling.

(2)

The manager decides to take a stratified sample of 100 students.

(b) Calculate the number of students to be sampled from each course.

(3)

(c) Describe how to choose students for the stratified sample.

(2)

2 A village has a population of 600 people. A sample of 12 people is obtained as follows. A list of all 600 people is obtained and a three-digit number, between 001 and 600 inclusive, is allocated to each name in alphabetical order. Twelve three-digit random numbers, between 001 and 600 inclusive, are obtained and the people whose names correspond to those numbers are chosen.

(i) Find the probability that all 12 of the numbers chosen are 500 or less. [3]

(ii) When the selection has been made, it is found that all of the numbers chosen are 500 or less. One of the people in the village says, "The sampling method must have been biased." Comment on this statement. [2]

3(a)i e.g e.g	**Quota Sampling:** Advantages: Fieldwork can be done **quick**ly, or administering the test is **easy**, or costs are kept to a minimum (**cheap**), or gives estimates for each course. or OK for large populations or sampling frame not required (o.e.) Disadvantages: **Non-random** process or not possible to estimate the sampling errors, or non response not recorded, or interviewer can introduce **bias** in sample choice. (o.e.)	B1
3(a)ii e.g. e.g.	**Stratified Sampling:** Advantages: Can give accurate estimates as it is a **random** process, or gives estimates for each course or **representative** of [BUT not "proportional" to] the whole population. (o.e.) Disadvantages: Sampling frame required, or strata may not be clear as some students overlap courses or not suitable for large populations. (o.e.)	B1 (2)
3(b)	Total enrolments=1000 (may be implied by calculations) Leisure and Sport= $\frac{420}{1000}\times 100 = 42$ Information Technology= $\frac{337}{1000}\times 100 = 33.7 = 34$ Health and Social Care= $\frac{200}{1000}\times 100 = 20$ Media Studies= $\frac{43}{1000}\times 100 = 4.3 = 4$	B1 M1 A1 (3)
3(c)	The college's information system would be used to identify each student and which course they are enrolled on. i.e. idea of **sampling frame** or **list** for **each course**. Use of **random numbers** to select required number of students **from each course**	B1 B1 (2) **Total 7**

2 (i)	Let R denote the number of choices which are 500 or less. $R \sim B(12, \frac{5}{6})$ $P(R = 12) = (\frac{5}{6})^{12}$ [=0.11216] $= \mathbf{0.112}$	M1 M1 A1	**3**	$B(12, \frac{5}{6})$ stated or implied, allow 501/600 etc p^{12} or q^{12} or equivalent Answer, a.r.t. 0.112 [SR: $\frac{500}{600}\times\frac{499}{599}\times\frac{498}{598}\times\ldots$; 0.110: M1A1] [M1 for 0.910 or 0.1321 or vague number of terms]
(ii)	Method unbiased; unrepresentative by chance	B1 B1	**2**	State that method is unbiased Appropriate comment (e.g. "not unlikely") [SR: partial answer, e.g. not necessarily biased: B1]

A telephone directory contains 50 000 names. A researcher wishes to select a systematic sample of 100 names from the directory.

(a) Explain in detail how the researcher should obtain such a sample.

(2)

(b) Give one advantage and one disadvantage of

(i) quota sampling,

(ii) systematic sampling.

(4)

1 The head teacher of a school asks for volunteers from among the pupils to take part in a survey on political interests.

(i) Explain why a sample consisting of all the volunteers is unlikely to give a true picture of the political interests of all pupils in the school. [2]

(ii) Describe a better method of obtaining the sample. [3]

4 A survey is to be carried out to draw conclusions about the proportion p of residents of a town who support the building of a new supermarket. It is proposed to carry out the survey by interviewing a large number of people in the high street of the town, which attracts a large number of tourists.

(i) Give two different reasons why this proposed method is inappropriate. **[2]**

(ii) Suggest a good method of carrying out the survey. **[3]**

(iii) State two statistical properties of your survey method that would enable reliable conclusions about p to be drawn. **[2]**

Q1 (a)	Randomly select a number between 00 and 499 (001 and 500) select every 500th person	B1 B1	(2)
(bi)	Quota **Advantage:** Representative sample can be achieved (with small sample size) Cheap (costs kept to a minimum) **not** "quick" Administration relatively easy **Disadvantage** Not possible to estimate sampling errors (due to lack of randomness) Not a random process Judgment of interviewer can affect choice of sample – bias Non-response not recorded Difficulties of defining controls e.g. social class	B1 B1	(2)
(bii)	Systematic **Advantage:** Simple or easy to use **not** "quick" or "cheap" or "efficient" It is suitable for large samples (not populations) **Disadvantage** Only random if the ordered list is (truly) random Requires a list of the population or must assign a number to each member of the pop.	B1 B1	(2) **[6]**

1	**(i)**	Biased in favour of those with strong political interest	B2 **2**	"Biased", "unrepresentative", "not indept" or equiv [but *not* "not random"] stated, with sensible reason. [SR: partial answer, B1]
	(ii)	Obtain list of all pupils Allocate numbers sequentially Choose using random numbers	B1 B1 B1 **3**	List, can be implied; number serially or randomly, not just "number pupils" Select consistently with method of numbering, not just "select randomly" [SR: systematic: List B1, every n^{th} B1, random start B1] [SR: names in a hat: B2]

4	(i)	Eg "not all are residents"; "only those in street asked"	B1 B1 **2**	One valid relevant reason A definitely different valid relevant reason *Not* "not a random sample", *not* "takes too long"
	(ii)	Obtain list of whole population Number it sequentially Select using random numbers [Ignore method of making contact]	B1 B1 B1 **3**	"Everyone" or "all houses" must be implied *Not* "number it with random numbers" unless then "arrange in order of random numbers" **SR**: "Take a random sample": B1 **SR**: Systematic: B1 B0, B1 if start randomly chosen
	(iii)	Two of: α: Members of population equally likely to be chosen β: Chosen independently/randomly γ: Large sample (e.g. > 30)	B1 B1 **2**	One reason. NB : If "independent", must be "chosen" independently, not "views are independent" Another reason. Allow "fixed sample size" but not both that and "large sample". Allow "houses"

2 A college has 400 students. A journalist wants to carry out a survey about food preferences and she obtains a sample of 30 pupils from the college by the following method.

- Obtain a list of all the students.
- Number the students, with numbers running sequentially from 0 to 399.
- Select 30 random integers in the range 000 to 999 inclusive. If a random integer is in the range 0 to 399, then the student with that number is selected. If the number is greater than 399, then 400 is subtracted from the number (if necessary more than once) until an answer in the range 0 to 399 is selected, and the student with that number is selected.

(i) Explain why this method is unsatisfactory. **[2]**

(ii) Explain how it could be improved. **[1]**

2	(i)	Not all equally likely – those in range 0 to 199 more likely to be chosen	M1 A1	**2**	Not all equally likely stated or implied Justified by reference to numbers, no spurious reasons
	(ii)	Ignore random numbers greater than 799, or 399	B1	**1**	Any valid resolution of this problem, no spurious reasons