## Question 1

**Worked Solution**

**(a) Two reasons stratified sampling is more suitable than simple random sampling (SRS)**

Any two of the following: (1) Stratified sampling guarantees representation of all subgroups, whereas SRS does not. (2) It allows accurate estimates to be made for each subgroup separately. (3) It reflects the structure of the population by sampling each stratum in proportion to its size.

**(b) Two reasons stratified sampling is more suitable than quota sampling**

Any two of the following: (1) Stratified sampling is a random process so it is not biased, whereas quota sampling can introduce interviewer bias in the selection of individuals. (2) Sampling errors can be estimated for a stratified sample but not for a quota sample.

## Question 2

**Worked Solution**

**(a) What is a random sample from a finite population?**

A random sample is one where every possible sample of size $n$ has an equal chance of being selected (equivalently, every member of the population has an equal probability of being chosen).

**(b) Example of a situation where a random sample is not possible**

When a sampling frame cannot be constructed — for example, sampling fish in a lake, or sampling people who have never visited a doctor.

**(c)(i) Sampling frame**

A complete list (register) of all 75 students on the course.

**(c)(ii) Using a random number table**

Number the students from 1 to 75 (or 0 to 74). Starting at a random position in the random number table, read off two-digit numbers. Select a student when their number appears. Ignore numbers outside the range 1–75 (or 0–74) and ignore repeats. Continue until 8 students have been selected.

## Question 3

**Worked Solution**

Company: 72 managers, 108 drivers, 180 administrators, 360 warehouse staff. Total = 720. Stratified sample of 40.

**(a) One advantage of stratified over simple random sampling**

The sample will be more representative of the population structure since each job role is represented in proportion to its size.

**(b) Number from each job role**

$$\text{Managers: } \frac{72}{720} \times 40 = 4, \quad \text{Drivers: } \frac{108}{720} \times 40 = 6$$

$$\text{Administrators: } \frac{180}{720} \times 40 = 10, \quad \text{Warehouse: } \frac{360}{720} \times 40 = 20$$

Managers: 4; Drivers: 6; Administrators: 10; Warehouse staff: 20.

**(c) How to choose managers for the sample**

Label all 72 managers from 1 to 72. Use random numbers in the range 1–72 to select 4 managers. Ignore repeats and any numbers outside the range.

## Question 4

**Worked Solution**

(Same question as Q1 — identical mark scheme applies.)

**(a) Two reasons stratified is more suitable than SRS**

(1) Stratified sampling guarantees representation of all subgroups; SRS does not. (2) It allows estimates to be made for each subgroup and reflects the population structure.

**(b) Two reasons stratified is more suitable than quota sampling**

(1) Stratified is a random process and is therefore unbiased; quota sampling can introduce interviewer bias. (2) Sampling errors can be estimated for a stratified sample but not for a quota sample.

## Question 5

**Worked Solution**

Gym club: 400 members, 300 male, 100 female. Stratified sample of size 60.

Label the 300 males from 1 to 300 and the 100 females from 1 to 100 separately. Number to select from each group:

$$\text{Males: } \frac{300}{400} \times 60 = 45, \qquad \text{Females: } \frac{100}{400} \times 60 = 15$$

Use random numbers in the range 1–300 to select 45 males, and random numbers in the range 1–100 to select 15 females. Ignore repeats and out-of-range numbers in each case.

## Question 6

**Worked Solution**

College: Leisure & Sport 420, IT 337, Health & Social Care 200, Media Studies 43. Total = 1000. Sample of 100.

**(a)(i) Quota sampling — one advantage and one disadvantage**

Advantage: Quick and cheap to carry out; no sampling frame required.
Disadvantage: Non-random process so sampling errors cannot be estimated; interviewer bias may affect who is chosen within each quota.

**(a)(ii) Stratified sampling — one advantage and one disadvantage**

Advantage: Random process giving unbiased estimates; representative of each course.
Disadvantage: A sampling frame is required; strata boundaries must be clearly defined and may be complex.

**(b) Number from each course**

$$\text{Leisure \& Sport: } \frac{420}{1000} \times 100 = 42$$

$$\text{IT: } \frac{337}{1000} \times 100 = 33.7 \approx 34$$

$$\text{Health \& Social Care: } \frac{200}{1000} \times 100 = 20$$

$$\text{Media Studies: } \frac{43}{1000} \times 100 = 4.3 \approx 4$$

Leisure & Sport: 42; IT: 34; Health & Social Care: 20; Media Studies: 4.

**(c) How to select students**

Use the college information system to obtain a list of all students enrolled on each course (this provides the sampling frame for each stratum). Use random numbers to select the required number of students from each course list separately. Ignore repeats.

## Question 7

**Worked Solution**

Lake: 1400 trout, 600 bass, 450 pike. Sample of 30.

**(a) Why stratified random sampling cannot be used**

It is impossible to construct a sampling frame for the fish in the lake (you cannot list or number all the fish).

**(b) Appropriate sampling method**

Quota sampling.

**(c) Advantage and disadvantage**

Advantage: The sample can be obtained quickly and costs are kept to a minimum.
Disadvantage: It is not a random process so sampling errors cannot be estimated; there may also be difficulty in identifying the species of each fish caught.

**(d) How to select the sample of 30**

Total estimated fish $= 1400 + 600 + 450 = 2450$.

Trout quota: $\dfrac{1400}{2450} \times 30 \approx 17$, Bass quota: $\dfrac{600}{2450} \times 30 \approx 7$, Pike quota: $\dfrac{450}{2450} \times 30 \approx 6$

Set quotas of 17 trout, 7 bass and 6 pike. Fish are caught from the lake until each quota is reached. If a fish is caught after its species quota is full, it is returned and not counted.

## Question 8

**Worked Solution**

GT Engineering: cookers in batches of 2000, unique serial numbers. Sample of 5 tested to destruction.

**(a) What is a census?**

A census is when every member of the population is investigated.

**(b) Reason (other than time/cost) for using a sample rather than a census**

Testing a cooker to find its overload threshold destroys it, so a census would leave no cookers to sell.

**(c) Suitable sampling frame**

A list of the unique serial numbers of all 2000 cookers in the batch.

**(d) Sampling units**

Individual cookers (or their serial numbers).

## Question 9

**Worked Solution**

(Identical context to Q8.)

**(a)–(d)** Answers are identical to Question 8:

(a) A census investigates every member of the population.
(b) Testing destroys the cooker, so a census would leave no cookers to sell.
(c) A list of the unique serial numbers of all 2000 cookers.
(d) Individual cookers (or serial numbers).

## Question 10

**Worked Solution**

Telephone directory: 50 000 names. Systematic sample of 100.

**(a) How to obtain the systematic sample**

Calculate the sampling interval: $50000 \div 100 = 500$. Select a random starting number between 1 and 500 (inclusive) using a random number generator or table. Then select every 500th name on the list thereafter, until 100 names are chosen.

**(b)(i) Quota sampling — advantage and disadvantage**

Advantage: Can give a representative sample at low cost; no sampling frame required.
Disadvantage: Not a random process, so sampling errors cannot be estimated; interviewer judgement may introduce bias.

**(b)(ii) Systematic sampling — advantage and disadvantage**

Advantage: Simple and easy to use, particularly for large populations.
Disadvantage: Only truly random if the ordered list has no cyclical patterns; a sampling frame (list) of the population is required.

## Question 11

**Worked Solution**

Cleaning company: 55 managers, 495 cleaners. Total = 550. Questionnaire to first 50 cleaners to leave.

**(a) Two reasons this method may produce biased results**

(1) Only cleaners are surveyed — no managers are included, so the sample does not represent all types of employee.
(2) The first 50 to leave may share the same shift or working pattern, so they may not represent the full range of views among cleaners.

**(b)(i) Systematic sample of 50**

Label all 550 employees from 1 to 550. Calculate the interval: $550 \div 50 = 11$. Select a random starting number between 1 and 11. Then select every 11th employee on the list until 50 are chosen.

**(b)(ii) Stratified sample of 50**

Total = 550.

$$\text{Managers: } \frac{55}{550} \times 50 = 5, \qquad \text{Cleaners: } \frac{495}{550} \times 50 = 45$$

Label the 55 managers from 1 to 55 and select 5 using random numbers. Label the 495 cleaners from 1 to 495 and select 45 using random numbers. Ignore repeats in each case.

**(c) Next two suitable random numbers (after 384 and 100)**

Starting from the random number tables at the top left with 8, working across, and selecting numbers between 1 and 550, the first two are 384 and 100. The next two are:

390 and 372.

*End of Worked Solutions*