2.1 Statistics: Sampling: Test Pack

EDEXCEL 9MA0 · Calculator allowed · about 60 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 A rail company wants to investigate the average delay time, in minutes, experienced by passengers across all 2.4 million passenger journeys it operated last year. The company holds a complete computer record of the delay time for every one of these journeys.

(a) State what is meant by a census. **(1)**

(b) The company could analyse the delay time for all 2.4 million journeys, i.e. take a census. Give one advantage of doing this rather than using a sample. **(1)**

(c) Give one disadvantage of the company using a census in this situation. **(1)**

(d) The company instead decides to analyse a sample of 500 journeys. Give one reason why using a sample could be preferable to a census here. **(1)**

(Total for Question 1 is 4 marks)

© Revision Library | revisionlibrary.co.uk

2 A large secondary school has 1500 students across Years 7 to 13. The head teacher wants to survey students about the quality of school lunches. The school's admissions system holds a complete, up to date list of every student, together with their year group and name.

(a) Give one reason why taking a sample, rather than a census of all 1500 students, would be appropriate here. **(1)**

(b) Recommend a suitable sampling method for the head teacher to use, and justify your choice with reference to the information given. **(3)**

(c) State one limitation of the method you recommended in part (b). **(1)**

(Total for Question 2 is 5 marks)

---

3 A council wants to survey households. It lists all 1200 households and chooses every 10th one after a random start.

(a) Name the sampling method described. **(1)**

(b) State one advantage of this method in this situation. **(1)**

(c) State one disadvantage of this method in this situation. **(1)**

(Total for Question 3 is 3 marks)

© Revision Library | revisionlibrary.co.uk

4 A library has 300 registered members. The librarian selects 60 members to ask about opening hours. The population is split as follows: Section 1: 60, Section 2: 80, Section 3: 70, Section 4: 90.

(a) Describe how the sample could be selected using stratified sampling, and find the number to be taken from each group. **(3)**

(b) State one advantage of stratified sampling over simple random sampling. **(1)**

(c) Explain how systematic sampling could be used to select a sample of 60 from the whole population of 300. **(2)**

(Total for Question 4 is 6 marks)

---

5 The 180 members of a club are numbered 1 to 180. A systematic sample of 9 members is selected, starting with member 14.

(a) Write down the sampling interval and the numbers of the first four members selected. **(2)**

(b) Find the number of the last member selected. **(2)**

(c) The club's members who are numbered with a multiple of 20 are all of one type. Explain why the sample may be biased. **(2)**

(Total for Question 5 is 6 marks)

© Revision Library | revisionlibrary.co.uk

6 A stable owner has 84 horses, numbered from 00 to 83. She wants to select a simple random sample of 4 horses to check for a particular health condition.

(a) Describe fully how the owner could use a random number generator to select this simple random sample, ensuring no horse is chosen more than once. **(3)**

(b) The random digits below were generated and grouped into pairs, reading from left to right: 51 84 07 07 62 39. Given that horses are numbered 00 to 83, identify the 4 horses selected by this method. **(2)**

(Total for Question 6 is 5 marks)

---

7 A market researcher stands outside a shopping centre. Based on past data, she knows that 40% of shoppers at the centre are men and 60% are women. She plans to stop and interview shoppers about a new energy drink until she has interviewed 20 men and 30 women.

(a) State the name of this sampling method. **(1)**

(b) Give one advantage of this method compared with using stratified sampling in this situation. **(1)**

(c) Give one disadvantage of this method. **(1)**

(d) Suggest one way the researcher could reduce the bias identified in part (c). **(1)**

(Total for Question 7 is 4 marks)

---

8 A local newspaper prints a coupon and asks readers to cut it out, fill it in, and post it back if they want to take part in a survey. State the name of this sampling method.

(Total for Question 8 is 1 mark)

---

9 A company has 250 staff: 150 full-time and 100 part-time. A stratified sample of 20 staff, proportional to employment type, is required. Calculate the number of full-time staff that should be included in the sample.

(Total for Question 9 is 2 marks)

© Revision Library | revisionlibrary.co.uk

10 A stable has 60 sheep, numbered 00 to 59. To select a simple random sample, two-digit random numbers were generated in this order: 07, 63, 42, 91, 18, 29. Any number greater than 59 must be rejected. Identify, in order, the first four valid sheep numbers selected.

(Total for Question 10 is 2 marks)

---

11 A company with 2000 employees at several sites wants to estimate average commuting time. A researcher instead surveys only the 50 employees who work in the head office building, since they are the easiest for her to reach. Identify the sampling method used, and explain why the resulting sample could give a biased estimate of the average commuting time for all 2000 employees.

(Total for Question 11 is 3 marks)

---

12 A conveyor belt produces 980 identical parts during a shift. A systematic sample of 30 parts is required. Calculate the sampling interval, to the nearest whole number, and hence find the smallest number of parts that could be sampled using this interval before the shift ends.

(Total for Question 12 is 3 marks)

---

13 A university with 24000 students wants to estimate the proportion of students who cycle to campus. Method 1 selects a simple random sample of 400 students from the central enrolment database (which lists every student) and emails them a survey. Method 2 involves standing at the three main bike racks on campus for one morning and interviewing everyone who arrives by bike, until 400 people have been interviewed. Identify one source of bias in Method 2 that Method 1 does not share, and one practical problem with Method 1 that Method 2 does not share.

(Total for Question 13 is 4 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.1 Statistics: Sampling: Test Pack

Question 1

(a) a survey/study that collects data from every member of the population, oe **B1**

(a) Answer: A survey or study that collects data from every member of the population.

(b) gives completely accurate results with no sampling error, since the whole population is used, oe **B1**

(b) Answer: It gives a completely accurate result, with no sampling error, since every journey is included.

(c) processing/analysing 2.4 million records is far more time-consuming (and costly) than using a sample, oe **B1**

(c) Answer: Processing 2.4 million individual records is far more time-consuming and expensive than analysing a sample.

(d) results can be obtained much more quickly, since far fewer records need to be processed, oe (must be a different point from part (c)) **B1**

(d) Answer: Results can be produced much more quickly, since only 500 records need to be processed rather than 2.4 million.

Question 2

(a) much quicker and cheaper to survey a sample than every one of the 1500 students, and a well-chosen sample can still give a reliable estimate of opinion, oe **B1**

(a) Answer: A sample is much quicker and cheaper to survey than all 1500 students, and a well-chosen sample can still give a reliable estimate of opinion.

(b) names an appropriate method, e.g. stratified sampling by year group **B1**

(b) justifies using the sampling frame available, which records each student's year group, so the number of students per year can be found and used to calculate proportional strata sizes **B1**

(b) further valid justification, e.g. stratified sampling ensures every year group is represented in proportion to its size, avoiding a year group being over- or under-represented as could happen with simple random sampling, oe **B1**

(b) Answer: Stratified sampling by year group is recommended, since the school's list records each student's year group, allowing proportional strata sizes to be calculated; this also guarantees every year group is fairly represented, unlike simple random sampling, which could by chance over- or under-represent a year group.

(c) it takes more time and is more complex to set up than simple random sampling, since students must first be split into strata and each stratum's sample size calculated, oe **B1**

(c) Answer: It takes more time and effort to set up than simple random sampling, since students must first be divided into strata and each stratum's sample size calculated separately.

Question 3

(a) systematic sampling **B1**

(a) Answer: systematic sampling

(b) simple and quick to carry out, and it spreads the sample evenly across the list **B1**

(b) Answer: simple and quick to carry out, and it spreads the sample evenly across the list

(c) may be biased if the list has a periodic pattern **B1**

(c) Answer: may be biased if the list has a periodic pattern

Question 4

(a) sampling fraction 60/300 = 1/5, applied to each group **M1**

(a) numbers from each group: Section 1 12, Section 2 16, Section 3 14, Section 4 18 **A1**

(a) within each group the items are selected at random (for example by random number generator) **B1**

(a) Answer: Section 1 12, Section 2 16, Section 3 14, Section 4 18

(b) every group is represented in proportion, so the sample is more representative of the population **B1**

(b) Answer: more representative of each group

(c) sampling interval 300/60 = 5 **M1**

(c) choose a random start between 1 and 5 and then select every 5th item from the numbered list **A1**

(c) Answer: every 5th item after a random start from 1 to 5

Question 5

(a) interval = 180/9 = 20 **B1**

(a) 14, 34, 54, 74 **B1**

(a) Answer: interval 20; 14, 34, 54, 74

(b) 14 + (9 - 1) x 20 **M1**

(b) 174 **A1**

(b) Answer: 174

(c) if the numbering follows a pattern with the same period as the sampling interval, the sample may contain only that type **B1**

(c) so each member does not have an equal chance of selection when the list is ordered in a periodic way **B1**

(c) Answer: possible bias from a periodic list

Question 6

(a) assigns each horse a unique two-digit number from 00 to 83, using the existing list as the sampling frame **B1**

(a) uses a random number generator to produce two-digit random numbers, rejecting any outside the range 00 to 83 **B1**

(a) also rejects any repeated numbers, continuing until 4 different horses have been selected **B1**

(a) Answer: Label the horses 00 to 83. Generate random two-digit numbers, ignoring any greater than 83 and any repeats, until 4 different horses have been chosen.

(b) correctly rejects 84 (outside the range 00 to 83) and the second 07 (a repeat) **M1**

(b) cao: horses 51, 07, 62 and 39 selected, in this order **A1**

(b) Answer: Horses 51, 07, 62 and 39.

© Revision Library | revisionlibrary.co.uk

Question 7

(a) quota sampling **B1**

(a) Answer: Quota sampling.

(b) no sampling frame of all shoppers is needed, and it is quicker/cheaper to carry out on location, oe **B1**

(b) Answer: No sampling frame of every shopper is required, so it is quicker and cheaper to carry out on the spot than stratified sampling.

(c) the researcher chooses who to approach, so the sample is non-random and interviewer bias can be introduced, e.g. avoiding people who look busy or unfriendly, oe **B1**

(c) Answer: The researcher personally chooses who to approach, so the sample is non-random and interviewer bias can occur, for example avoiding people who look busy.

(d) approach every nth shopper who passes, rather than personally choosing who to stop, oe **B1**

(d) Answer: Approach every nth shopper who walks past, rather than personally selecting who to stop.

Question 8

self-selecting sampling (volunteer sampling) **B1**

Answer: Self-selecting sampling (volunteer sampling).

Question 9

150 / 250 x 20 **M1**

12 cao **A1**

Answer: 12

Question 10

correctly rejects 63 and 91 as outside the range 00 to 59 **M1**

cao: sheep 07, 42, 18 and 29 selected, in this order **A1**

Answer: Sheep 07, 42, 18 and 29.

Question 11

opportunity sampling (convenience sampling) **B1**

identifies that head office employees may differ systematically from staff at other sites, e.g. tending to live nearer that particular location **B1**

explains the consequence: commuting times at head office may not be typical of the whole company, so the sample is unlikely to be representative and the estimate could be biased, oe **B1**

Answer: Opportunity sampling. Head office staff may live closer to that one site than staff based elsewhere, so their commuting times are unlikely to be typical of the whole company; the estimate from this sample is therefore likely to be biased.

Question 12

980 / 30 (= 32.67), rounded to interval 33 **M1**

980 / 33 (= 29.7) **M1**

cao: the smallest possible sample is 29 parts, since 33 x 30 = 990 > 980 means a late start cannot reach 30 parts **A1**

Answer: Interval = 33; the smallest possible sample is 29 parts, since 33 x 30 = 990, which is greater than 980 (a late start cannot reach 30 parts).

Question 13

Method 2 only samples people who are already cycling to campus that morning (an opportunity/self-selected sample of cyclists), so non-cyclists have no chance of selection **B1**

this would make the estimated proportion who cycle far too high, since (close to) everyone sampled by Method 2 cycles, unlike the true population of all 24000 students **B1**

Method 1 is likely to suffer from non-response, since not every emailed student will reply **B1**

if the students who choose to respond differ systematically from those who do not (e.g. cyclists feeling more strongly about the topic and being more likely to reply), this introduces non-response bias not present in Method 2's in-person approach, oe **B1**

Answer: Method 2 samples only people already cycling that morning, so it will greatly overestimate the true proportion of cyclists among all 24000 students. Method 1 instead relies on email responses, so it is vulnerable to non-response bias if the type of student who replies differs from those who do not - a problem Method 2, which interviews people in person, does not have.

© Revision Library | revisionlibrary.co.uk