A Level · Edexcel · Maths

2 hours 23 questions

1.1 Sampling & Data Collection

1.1.1 Sampling & Data Collection

Easy (7 questions)	/37
Medium (8 questions)	/39
Hard (8 questions)	/58
Total Marks	**/134**

Scan here to return to the course
or visit savemyexams.com

Easy Questions

1 The Office for National Statistics runs a compulsory census every ten years to gather information about all individuals and households in England and Wales. The information gathered helps organisations make decisions on planning and funding for public services in each area, including transport, education and healthcare.

(i) Explain what is meant by a census.

(ii) Write down two disadvantages of using a census rather than a sample.

(3 marks)

2 Computer components produced by an electronics company are each given a unique serial number. A34X processing chips are produced in batches of 2500. For quality control, the company tests a random sample of 20 A34X chips from each batch.

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

(ii) Identify the sampling units.

(2 marks)

3 (a) Continuous variables can take any value in a given range whereas discrete variables can only take specific values in a given range.

State whether each of the following variables are discrete or continuous.

(i) Perimeter of a park.

(ii) Number of workers in a factory.

(iii) Speed at which a humming bird flaps its wings.

(iv) Time in minutes spent playing CandyCrush.

(v) Colour of passing cars.

(vi) Age of a person.

(3 marks)

(b) Numerical variables or data are quantitative whereas non numerical variables or data are qualitative.

Give two examples of variables or data for each of the following

(i) quantitative

(ii) qualitative.

(2 marks)

4 (a) A fast-food chain, introducing a new vegan menu, employ a researcher to investigate people's opinions before they launch the products. The researcher decides to conduct a survey on a random sample of people to gather their opinions on vegan fast food. He decides to stand at the entrance to a busy shopping mall on a Wednesday afternoon and attempts to get shoppers to answer his questions.

Explain what is meant by the words 'population', 'sampling unit' and 'sampling frame'. You may use the context of the above scenario as an example.

(3 marks)

(b) (i) State the sampling technique the researcher is using to gather data.

(ii) Suggest one improvement the researcher could make to get a better sample and give a reason for your choice.

(3 marks)

5 (a) A flatpack furniture company, AEKI, has a testing facility where its products are put through a series of safety and quality checks.

Suggest two reasons why it would not be sensible for AEKI to test every product they produce.

(2 marks)

(b) Explain how AEKI could use a systematic sample to test their products.

(2 marks)

6 (a) A high school holds an annual summer festival to raise money for events and trips throughout the year. Before this year's festival the headteacher decided to survey the opinion of staff and students using a stratified sample.

(i) Explain why the headteacher decided to take a stratified sample of staff and students.

(ii) Suggest a suitable sampling frame and identify the sampling units.

(3 marks)

(b) Suggest a possible problem that might arise with the sampling frame when selecting the staff and students.

(2 marks)

(c) There are 360 students and 40 staff at the school.

Explain how the headteacher could take a stratified sample of size 40

(3 marks)

(d) Suggest one improvement that could be made when calculating the stratified sample to gain a more representative sample of staff and students within the school and give a reason for your answer.

(2 marks)

7 (a) N'Oréal, a hair and beauty company, release an advertising campaign for its new product called 'Face Amazifier'. The small print at the bottom of the advert says the following

'55% of 25 existing customers agree that the product makes your face feel more amazing.'

The Advertising Standards Agency get complaints that the advert is not representative.

Explain the meaning of 'bias' in relation to the choice of sample N'Oréal has used.

(2 marks)

(b) Suggest two ways N'Oréal could improve their sample to make it more representative.

(2 marks)

(c) (i) Show that 55% of 25 customers is not a valid statistic for N'Oréal to use.

(ii) State the two closest possible alternative figures that would be valid.

(3 marks)

Medium Questions

1 (a) The Office for National Statistics runs a compulsory census every ten years to gather information about all individuals and households in England and Wales. The information gathered helps organisations make decisions on planning and funding for public services in each area, including transport, education and healthcare.

Suggest one advantage and one disadvantage of the Census only being carried out every ten years.

(2 marks)

(b) A local council wants to gather opinions from residents about opening a new care home. They decide to conduct their own survey and need to pick between using a systematic or a stratified sample.

Give two reasons why the council may be better off conducting their own survey rather than attempting to use census data to obtain the desired information.

(2 marks)

(c) Explain the main differences between a systematic sample and a stratified sample.

(2 marks)

2 (a) Every week an orangutan sanctuary measures the weight of each of its orangutans.

The weights, to the nearest kg, of **ALL** their 18 adult males are listed below:

52, 57, 63, 80, 56, 66, 101, 68, 55, 96, 70, 62, 66, 64, 99, 91, 55, 92

(i) State the data collection process used to collect the above data.

(ii) Describe the type of data collected.

(2 marks)

(b) Using an opportunity sample of size six, calculate the mean weight of the male orangutans from the data set above.

(2 marks)

(c) Starting from the third data value, take a systematic sample of size six and re-calculate the mean weight of the male orangutans from the data set above.

(2 marks)

(d) Compare your results from parts (b) and (c) and state, with a reason, which sampling method is more reliable.

(2 marks)

3 (a) A supermarket wants to gather data from its shoppers on how far they have travelled to shop there. One lunchtime an employee is stationed at the door of the shop for half an hour and instructed to ask every customer how far they have travelled.

(i) State the sampling method the employee is using.

(ii) Give one advantage and one disadvantage of using this method.

(3 marks)

(b) State and briefly describe an alternative method of non-random sampling that the employee could use to obtain the required data for a sample of 30 customers.

(2 marks)

4 To check the quality of produce used in a restaurant kitchen, the head chef likes to taste one item from each box of produce as soon as it arrives at the restaurant. Some produce is then kept in cold storage and used throughout the week.

(i) Suggest a reason why the chef does not taste test every item.

(ii) Suggest one way the chef's sample taste test could be improved.

(2 marks)

5 A company wants to survey 15% of its staff to find out whether employees would like to continue working from home after the Covid-19 pandemic. The company's 580 members of staff are grouped by job as follows: 295 engineers, 11 managers, 154 office staff and 120 apprentices.

Suggest a suitable sampling method and explain how the company can use this method to obtain its sample.

(5 marks)

6 The wingspans of a random sample of flamingos in a reserve were measured to the nearest centimetre. The results are shown in the table below:

Wingspan (cm)	Number of flamingos, f
140 - 144	2
145 - 149	5
150 - 154	8
155 - 159	3
160 - 164	6
165 - 169	1

(i) State the class boundaries, midpoint and class width for the class 150 – 154.

(ii) What is the maximum possible range of wingspans for the flamingos included in the data set?

(4 marks)

7 (a) A toy shop, 'Toys 4 U', tests the battery life for a new toy by leaving a sample of the toys switched on until their batteries run out.

Give one reason why the shop decided to use a sample rather than a census.

(1 mark)

(b) Five toys out of a shipment of 5000 were tested and the battery life (in minutes) of each toy was recorded:

172 minutes 252 minutes 248 minutes 155 minutes 161 minutes

The shop decides to display a sign beside the toys saying 'Up to 4 hours battery life'. Use the sample data above to comment on the sign.

(2 marks)

(c) Suggest one way in which the shop could improve its prediction.

(1 mark)

Date (Beijing October 1987)	Daily Mean Air Temperature (° C)	Rainfall (24 hour total) (mm)
01/10/1987	20.6	0.0
02/10/1987	19.1	0.0
03/10/1987	21.1	0.0
04/10/1987	20.4	0.0
05/10/1987	19.8	0.0
06/10/1987	19.3	0.0
07/10/1987	17.1	0.0
08/10/1987	16.5	0.0
09/10/1987	18.0	0.0
10/10/1987	18.9	0.0

Date (Beijing October 2015)	Daily Mean Air Temperature (° C)	Rainfall (24 hour total) (mm)
01/10/2015	16.1	3.0
02/10/2015	19.4	0.0
03/10/2015	18.6	0.0
04/10/2015	18.4	0.0
05/10/2015	18.9	0.0
06/10/2015	20.3	0.0
07/10/2015	20.5	0.0
08/10/2015	14.5	0.0
09/10/2015	14.7	0.0
10/10/2015	14.0	0.0

8 (a) A selection of data from the large data set relating to the temperature and rainfall in Beijing for the first 10 days in October in both 1987 and 2015 is given above.

The large data set records rainfall data for 184 consecutive days in both 1987 and 2015.

(i) State why the rainfall data shown above should not be used to compare the rainfall in Beijing for these two years.

(ii) Describe how a systematic sample of 10 days could be taken to compare the rainfall data.

(3 marks)

(b) Climate activists are using temperature data to argue that temperatures are becoming less consistent due to climate change.

Using the data from the first 10 days of October given above, find the range of the daily mean air temperature for both 1987 and 2015 and state whether the data above supports the activists' claims.

(2 marks)

Hard Questions

1 (a) An online magazine which offers both free and paid for content has a large number of readers. Readers can view additional content by paying a monthly subscription fee. Based on reviews on the magazine's website, the editor of the magazine believes that an additional type of content could be introduced. Before making any changes, the editor decides to carry out a sample survey to obtain the opinions of the readers.

Define the population that would be associated with the magazine.

(1 mark)

(b) Give one advantage and one disadvantage that would have resulted from the editor using a census rather than a sample survey.

(2 marks)

(c) The editor decides to gather opinions from **only** those readers who subscribe to the additional content. A random sample of 25 subscribers is selected for the sample survey.

Suggest a suitable sampling frame for the survey and identify the sampling units.

(2 marks)

(d) State two sources of uncertainty that may arise based on the chosen sampling method and sample size.

(2 marks)

2 (a) A group of 20 sloth sanctuaries across Central and South America record data on all the sloths in their care. The sanctuaries have a central database where data for each sloth is collated. A large variety of data on each individual sloth is collected and used to help decide when sloths are ready for releasing into the wild.

(i) Explain why a single central database may be helpful for sanctuary owners to use when comparing data on the sloths in their care.

(ii) Explain the difference between qualitative and quantitative data and give an example of each in the context of the question.

(3 marks)

(b) Sanctuary owners want to look at a sample of sloths as part of a general health and well-being survey. Sloths are said to have matured into adults when they reach the age of 5; before this they are classed as juveniles. While juveniles may be treated as a single group for this survey, it is important that adult females and adult males be considered separately.

In the database there are currently 240 adult sloths, 60% of whom are male, and 64 juvenile sloths. Explain how a stratified sample of size 35 could be taken.

(3 marks)

(c) Sanctuary owners want to look at a separate sample of sloths to get an idea of how many may be suitable for release. Sloths can be released into the wild as long as they reach a heathy weight and are over the age of 3. Male sloths when fully matured are generally heavier than females, although for juvenile sloths this weight difference is negligible.

Given that a quarter of the 64 juvenile sloths in the database are over the age of 3, explain how a stratified sample of size 35 could be taken to study suitability for release.

(3 marks)

3 (a) Stephan is researching the effects a new energy drink has on the glucose levels of students aged 13 to 18. He decides to measure the blood glucose levels of 50 female students and 50 male students.

(i) State, with a reason, whether Stephan is using a census or a sample to conduct his study.

(ii) Give two advantages and one disadvantage of this method.

(4 marks)

(b) Stephan is provided with an alphabetical list of 350 male students aged 13 to 18, each of whom has agreed to supply a blood sample if asked.

Explain how Stephen could use a calculator or a random number generator to take a simple random sample from the male students aged 13 to 18.

(3 marks)

(c) Stephen has an equivalent list of 350 female students aged 13 to 18.

Explain how Stephen could take a systematic sample from the female students aged 13 to 18.

(2 marks)

4 (a) Freda needs to conduct a survey to investigate the type of ice cream people prefer. She wants a random sample of people who eat ice cream. She decides to stand in a busy high street on a Sunday afternoon and attempt to get shoppers to answer her questions.

(i) Define the word 'population' in the context of Freda's survey.

(ii) State the sampling technique Freda has used.

(2 marks)

(b) Having been unsuccessful in obtaining enough data from her previous attempt, Freda decides to look at the electoral register for her town and select a new sample of people to contact. She needs a sample of at least 50 people, so she decides to choose a person at random from the register and then use that person and every 5th person on the register after that (wrapping back around to the start of the register if necessary) to build her sample.

(i) State the sampling technique Freda has used.

(ii) Give two reasons why Freda may again be unsuccessful getting the data required using this sampling technique.

(3 marks)

(c) Suggest an alternative method for Freda to use and explain your reasons.

(2 marks)

5 (a) The CEO of Save My Exams, Jamie, wants to find out what users would like to see on the revision website in future. He notices that around 15% of those who access the site have signed up to the mailing list to get content updates.

An employee suggests that they send out an email to all those who have signed up to the mailing list with a questionnaire for them to complete and return.

(i) Give two reasons why the users who return the questionnaire would not form a random sample of users of the website.

(ii) Given the site has over 650000 users, state two problems with sending out the questionnaire in this way.

(4 marks)

(b) Jamie decides to separate users by exam board to gather more detailed opinions. A member of the Maths Content Team suggests the use a table of random numbers to select a random sample of 100 users from the 4581 IB mailing list subscribers. The first five random numbers from the table are as follows.

02743 45290 19024 24337 90044

Explain how Jamie could use these random numbers to select the first few members in the sample.

(2 marks)

6 (a) A researcher measured heights of a random sample of giraffes. The heights in metres are summarised in the table below.

Height, h (m)	Number of giraffes, g
$3.5 \leqslant h < 4.0$	2
$4.0 \leqslant h < 4.5$	7
$4.5 \leqslant h < 5.0$	24
$5.0 \leqslant h < 5.5$	18
$5.5 \leqslant h < 6.0$	3
$6.0 \leqslant h < 6.5$	6

(i) State the class boundaries, midpoint and class width for the group containing the greatest number of giraffes.

(ii) Given that all the heights had been rounded to the nearest centimetre in the data set on which the table is based, what is the minimum possible range of heights for the giraffes as represented in the data set?

(4 marks)

(b) All giraffes are allocated a unique four-digit ID number. The researcher wants to randomly select five of the giraffes to test a new tracking device. The researcher selects the first five giraffes with three zeros in their ID number.

(i) Explain why this may not be a good method for selecting the giraffes.

(ii) Describe an improved method of selecting the giraffes.

(4 marks)

7 (a) A factory produces paper fruit baskets used for fruit pickers at 'pick your own' farms. The breaking load of a paper fruit basket is the maximum load that it can carry before the basket handles break. One 'pick your own' farm purchased 15 000 paper fruit baskets but wishes to test a sample of these to establish the breaking load of the baskets.

Suggest two reasons why a census would be unsuitable for this purpose.

(2 marks)

(b) The farm tests a random sample of six paper fruit baskets. The loads required for the handles to break are shown below:

2.035 kg 2.845 kg 2.528 kg 1.998 kg 2.212 kg 2.378 kg

The factory claims that the fruit baskets can carry 2 kg of fruit without breaking.Use the sample data to comment on this claim.

(2 marks)

(c) (i) Describe any limitations to the sample the farm has used.

(ii) Suggest one way the farm could improve the reliability of its results.

(3 marks)

8 (a) The large data set provides weather data for 184 consecutive days in each of the years 1987 and 2015.

Using the large data set, Charlie selects data from the 1st and 15th of each month to compare the weather in Hurn in 1987 and 2015.

(i) Charlie says he has done a systematic sample to select the data. Explain why Charlie is wrong.

(ii) Describe how Charlie could take a systematic sample of 12 days for each year from the data for Hurn for 1987 and 2015.

(4 marks)

(b) Charlie wants to use numerical analysis to compare the daily total rainfall from his samples. Using your knowledge of the large data set, explain why Charlie's sample may not necessarily give him 12 usable data points to compare for each year.

(1 mark)