



EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2013

MODULE 1 : Collection and compilation of data

Time allowed: Three Hours

*Candidates may attempt **all** the questions.*

The number of marks allotted to each question or part-question is shown in brackets.

The total for the whole paper is 100.

A pass may be obtained by scoring at least 50 marks.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

This examination paper consists of 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 7 questions altogether in the paper.

1. (i) Explain the difference between a *census* and a *sample survey*. (2)
- (ii) In some countries it is a legal requirement to carry out a census, in order to determine the demographics and working patterns of all residents. Choose a country with which you are familiar and discuss the following aspects of such a census in that country.
- (a) How the population could be identified.
- (b) How each person could be contacted.
- (c) Examples of types of people who might be difficult to contact, or who may not be able to take part, and how these problems could be handled.
- (d) Examples of possible biases in the responses, and how these could be minimised. (17)

2. A database contains 100 records with identifiers A1–A20, B1–B40, C1–C10 and D1–D30.

You are required to draw a stratified sample. This should comprise 5 records randomly selected from each stratum, where the letter prefixes (A–D) identify the four strata.

Use the random numbers below to draw a suitable sample. You must provide sufficient detail to ensure that your procedure can be fully understood.

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 92 | 17 | 63 | 21 | 58 | 64 | 70 | 73 | 31 | 41 |
| 09 | 08 | 63 | 23 | 26 | 66 | 85 | 20 | 26 | 60 |
| 56 | 84 | 65 | 72 | 51 | 65 | 78 | 74 | 06 | 61 |
| 06 | 02 | 78 | 90 | 90 | 63 | 98 | 83 | 47 | 84 |
| 11 | 02 | 46 | 24 | 06 | 39 | 73 | 47 | 98 | 87 |
| 00 | 60 | 52 | 89 | 65 | 05 | 95 | 16 | 49 | 66 |
| 57 | 15 | 62 | 63 | 50 | 60 | 27 | 95 | 64 | 31 |
| 70 | 30 | 11 | 74 | 31 | 41 | 04 | 94 | 60 | 18 |
| 63 | 02 | 79 | 09 | 44 | 01 | 91 | 75 | 88 | 50 |
| 95 | 99 | 09 | 98 | 74 | 47 | 71 | 77 | 46 | 93 |

(8)

3. A survey is to be carried out on makers of organic cheese in England. The aim is to assess their training needs in relation to their knowledge of food health and safety laws. Cheese makers are a very busy group of people who start work early in the morning and often work right through the day. They work in controlled environments, widely scattered throughout England, to which the general public are not normally allowed access and which they cannot leave for long. You have been given the names, addresses and telephone numbers of the people in the sample, together with the questions to be asked. Some questions are closed and some are open. You have a fixed budget and a short timescale.
- (i) Outline the advantages and disadvantages of each of the following methods for obtaining the required information in this survey.
 - (a) Interview at a visit.
 - (b) Postal questionnaire.
 - (c) Telephone interview.

(18)
 - (ii) State which method you would recommend. For your chosen method provide details of how the cheese makers would be contacted and how the information would be acquired. Justify your decisions.

(7)
4. An audit has been commissioned in a hospital. The aim is to estimate the frequency of occurrence of a particular adverse event for the whole hospital over a 12-month period. The people carrying out the audit have been advised that they do not need to visit every ward, but that they can select wards and also months in which to carry out the audit.
- (i) Identify the type of sampling used, giving reasons for your answer.

(3)
 - (ii) Discuss whether you think this is an appropriate method of sampling for this audit.

(8)

5. The following extract appeared in a report of a survey carried out in India.

"The sampling was carried out as follows. In the first stage, 2000 clusters were chosen from 3 strata: 400 from urban areas, 200 from the tea estate sector and 1400 from rural areas. The second stage involved systematic selection of 10 houses from each cluster. In this way we assured that the results of the survey were representative of the whole country."

(i) Define the terms *clusters* and *strata*. (4)

(ii) What other information would you require before you could comment on how representative the results are likely to be? (7)

6. A sports club offers a gym, a swimming pool, aerobics and a café. It is possible to buy annual membership. The manager wants to survey a sample of 100 annual members to find out their views of the club. He has designed the following questionnaire.

| |
|---|
| <u>XX Sports Club</u> |
| Dear Member, please complete this short questionnaire so that we can improve facilities. Put it in the box at reception. Thank you for your time. |
| What is your date of birth? |
| Are you male or female? |
| For how many years have you had annual membership at this club? |
| Which facilities do you use most? |
| Do you consider annual membership value for money and will you renew it? |

(i) Identify problems with this questionnaire. (10)

(ii) Re-design the questionnaire to obtain similar information, but addressing the problems you identified. (5)

7. (i) Briefly describe the problems caused by missing data in questionnaires returned in a sample survey. (2)

(ii) A questionnaire about library usage contains the following two questions.

A. Currently the library is open late on three days. Which suits you best? Please tick only one answer.

| | |
|---------------|--|
| Monday | |
| Wednesday | |
| Thursday | |
| No preference | |

B. Which of the following facilities do you use? Tick all that apply.

| | |
|--------------|--|
| Photocopier | |
| Computers | |
| Reading room | |
| DVDs | |

(a) Suggest a reason for including the answer option 'No preference' in question A.

(b) Explain why it would be impossible to identify all missing data for question B, and re-design the question to overcome this problem.

(c) Describe how answers to the two questions could be coded in a statistical worksheet, paying attention to how you would code missing values. Justify your decisions.

(9)

BLANK PAGE

BLANK PAGE

BLANK PAGE