



EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2015

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) Briefly explain the difference between a census and a sample survey. (2)
- (ii) Briefly describe three ways in which a sampling frame might be inadequate. What major problem might arise from having an inadequate sampling frame? (4)
- (iii) Briefly describe a sample survey in a country of your choice. Identify the sampling frame used and describe two steps that could be taken to maximise its adequacy. (5)

2. A village in a remote country consists of 36 houses built around a central area. A researcher wants to survey people in this village by carrying out six lengthy face-to-face interviews. In order to do this she wants to select six of the 36 houses at random.

Someone has suggested the 'spin the bottle' method. This involves standing at the centre of the village and spinning a bottle on the sandy ground. The house whose direction is nearest to the direction in which the neck of the bottle points is selected, provided that it is not already in the sample.

- (i) Describe a village layout where this method might give approximately random sampling. (1)
- (ii) Describe a village layout where it would be difficult to obtain a random sample using this method. (1)
- (iii) Explain what is meant by a table of random numbers in the range 00 to 99. (2)
- (iv) Describe a more scientific method of obtaining a simple random sample of six of the 36 houses in this village, and explain why your method would work equally well for each of your village layouts described in parts (i) and (ii). (7)

3. A team of university researchers is investigating the experiences and opinions of nurses working in nursing homes for elderly people. They specifically want to compare responses across three areas of the United Kingdom: England and Wales, Scotland and Northern Ireland. This is because there are variations in the regulations relevant to their research questions between these three areas. The research is to be led by three universities, one in each geographical area. There are many more nursing homes for elderly people in England and Wales than in the other two areas.

The researchers have decided to select 30 nursing homes in each of the three areas. These are to be near to the participating universities and known to the researchers. It is believed that this will increase the response rate of the selected participants in the research. Within each selected nursing home, three nurses will be chosen at random from the nursing staff at the home.

- (i) Explain why the proposed survey uses stratified sampling with the three areas of the United Kingdom as strata. (1)
- (ii) Within each stratum, what sampling method has been used to select the homes? Comment on and criticise the method used. (2)
- (iii) Comment on the claim that the method of selecting nursing homes is likely to increase the response rate. (3)
- (iv) The researchers claim that they will get a random sample of nurses working in nursing homes in the United Kingdom. Comment on this claim. (5)
- (v) Comment on the appropriateness of aiming for equal numbers of nurses from each of the three areas. (3)

4. A manager of a restaurant wants to survey a sample of customers after they have had a meal. He has decided to use a paper-based questionnaire to be given out with a free coffee after the meal and to be handed in at the reception desk.

(i) One of the questions he wants to ask is:

How satisfied were you with the range of choice on the menu?

Would this question, as worded, be open or closed? Discuss the advantages and disadvantages of open versus closed questions in such a questionnaire.

(5)

(ii) As well as the range of choice, the manager also wants to ask about the following.

The quality of the service.

The acceptability of the environment.

The value for money.

Design a suitable questionnaire for this purpose. Restrict your questions to these four issues, and choose open and closed questions carefully, justifying your decisions.

(8)

5. As part of a sample survey a researcher has collected the following data on 8 individuals.

<i>Age (years)</i>	<i>Sex M=male F=female</i>	<i>Number of children</i>	<i>Age of oldest child (years)</i>	<i>Marital status M=married or cohabiting S=single D=no longer with partner</i>
12	M	0	No answer	S
25	F	2	4	M
21	F	1	0	No answer
10	F	2	6	D
26	M	3	6	M
31	M	1	12	M
14	G	0	0	S
27	F	0	3	D

- (i) Explain the differences between missing values and errors in a dataset like this and describe the problems each can cause. (6)
- (ii) Identify all the problems in these data. (7)
- (iii) Explain briefly one way in which you could set about resolving the problems. (1)
- (iv) Assuming that the problems cannot be resolved, explain how you would code the data ready for entry into a computer-based statistical package. Pay particular attention to coding of errors and missing values, justifying your decisions. (6)

6. (i) Explain the key difference between an experimental study and an observational study. (2)
- (ii) Why is an experimental study generally preferred to an observational study when trying to establish which of two treatments is the more effective? (5)
- (iii) A researcher wants to investigate the safety of a procedure that is carried out on women whose babies get 'stuck' and who are therefore in distress during labour in childbirth. In particular he wants to compare the outcomes for both mother and baby from this procedure with those for another, more common, procedure that is used in similar situations. He has listed the outcomes of interest. He intends to collect data over the next 12 months from mothers who attend a particular large maternity hospital and who experience one of the two procedures during routine care.
- (a) Is this an experimental or observational study? Justify your answer. (1)
- (b) Describe any variables, other than outcome variables, that you think would be important in this study, explaining why you believe them to be important. (4)
- (c) Describe any problems you would anticipate in collecting the data. (4)
- (d) Briefly state the steps that would be necessary in processing the data ready for analysis. (4)
7. A survey is to be carried out on the alcohol consumption of full-time students attending a particular university. Many of the questions are sensitive in nature. All such students have an email account and ready access to a computer which they are required to use in their courses. Each student's term-time address is recorded centrally by the university. It has been decided that it would not be appropriate to carry out face-to-face interviews or phone interviews for this population.
- (i) Comment on the decision not to use face-to-face interviews or phone interviews. (4)
- (ii) The team is considering the use of either questionnaires sent by post or an internet survey. What would be your advice and why? (7)

BLANK PAGE

BLANK PAGE