EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2014

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).

1. (i)   Define *strata* in the context of sampling.

(2)

A database comprises 1000 records of three different types with the prefix A, B or C. The records are labelled as follows.
- A1–A500
- B1–B100
- C1–C400

A sample of 100 records is required.

(ii) Explain why a simple random sample is not appropriate in this situation.

(2)

(iii) The following sample is drawn.

A10 A20 A30 A40 … A470 A480 A490 A500 (sample of size 50)

B10 B20 B30 B40 … B70 B80 B90 B100 (sample of size 10)

C10 C20 C30 C40 … C370 C380 C390 C400 (sample of size 40)

How does this sample appear to have been obtained? Describe any problems with this method.

(3)

(iv) What would be your advice about the best way to select an appropriate sample of 100 records?

(3)

2.   In a survey of dog owner attitudes and their experiences in managing diabetic dogs, questionnaires were distributed to a sample of 36 veterinary clinics. At each clinic, the questionnaire was given to all dog owners whose dogs were undergoing treatment for diabetes.

(i)   Which method of sampling was used in this study? Discuss the merits of this approach for this survey.

(5)

(ii)   Some owners had more than one dog undergoing treatment for diabetes, and these owners received one questionnaire for each dog. Discuss the implications for the results of the survey if some owners completed several questionnaires.

(2)

(iii)   93 questionnaires were returned. What extra information would you need in order to be able to assess whether or not this was a good response rate?

(3)

(iv)   The developers of the questionnaire tried to ensure that it was straightforward to read and complete. Describe the features of the questionnaire which you think would be important in order for it to be straightforward to read and complete.

(8)

(v)   The questions were modified following responses from a pilot study of ten dog owners.

(a)   What is the purpose of *piloting* in a sample survey?

(b)   Contrast the following two methods of piloting the questionnaire.

• Asking dog owners to complete it and then asking them how straightforward it was

• Talking to owners as they complete the questionnaire

(c)   What other aspects of this survey would you consider important to pilot? Justify your answer.

(9)

3. (i) Distinguish between *open* and *closed questions*, stating the advantages and disadvantages of each method of obtaining data in a survey.

(7)

The manager at a garage wants to obtain views of customers who have had their cars serviced at the garage. He wants to ask about the following.

- The facilities for arranging the service, communicating with the garage staff, and paying for the work

- The quality of the work carried out

- Any other comments the customer may have

(ii) Justifying all your decisions, design a questionnaire that could be used for this purpose, and describe how you would administer this questionnaire.

(10)

4. People designing sample surveys often ask the question "How many replies do I need for my sample to be representative?"

(i) Explain what is meant by "representative" and why it is desirable for a sample to be representative.

(2)

(ii) Outline the advantages of receiving a large rather than a small number of responses in a sample survey.

(1)

(iii) Draft an answer to the original question.

(3)

5. A researcher wants to carry out a postal survey of people who have made use of a counselling service. Some of the questions are of a sensitive nature, and so the researcher believes that it would be best to keep the responses totally anonymous, thus ensuring that the identities of respondents cannot be linked to any of their answers. Comment on whether you think this is sensible or not.

(4)

6. (i)   Distinguish between *cross-sectional* and *longitudinal* surveys. Give an example of a question that could be investigated by a longitudinal survey but not by a cross-sectional survey.

(3)

(ii) Describe the main problems caused by dropout in longitudinal studies.

(2)

(iii) A tutor wants to investigate how her students' understanding develops over the three years of the course that she teaches. She has developed a test paper of 50 questions assessing key knowledge. She gave the test to all her students at the end of 2011 and collected in the answers. The average mark for the 43 students in their first year was 55%, the average mark for the 35 students in their second year was 60%, and the average mark for the 20 students in their third year was 65%. The tutor claims that these results show how students' scores tend to increase over the three years of the course.

(a)   Explain why this is a cross-sectional study.

(b)   Explain why the tutor's conclusion is incorrect and why such a deduction cannot be made from a cross-sectional study.

(3)

(iv) The tutor decides to use a longitudinal design and studies a single cohort of students on her course. In 2011 the average mark for the 43 students was 55%. Of these students, 40 completed the second year, with an average mark of 52% in 2012. Of these students, 12 completed the third year and their average mark in 2013 was 70%. The students had no access to the questions between examinations.

What conclusions can the tutor draw from these results? What other information would be helpful in interpreting the results?

(4)

5

7. Based on experience in her clinic, a doctor believes that women who report eating a lot of a certain foodstuff during pregnancy tend to have shorter labours than women who eat little or none of it. The doctor is proposing doing a more formal investigation into this phenomenon. She plans routinely to ask all pregnant women attending her clinic about their consumption of this foodstuff and then to relate this to the details of their subsequent labours. Her objective is to prove that eating it tends to reduce length of labour.

   (i) Is this proposed investigation *observational* or *experimental*? Justify your answer.

   (2)

   (ii) What advice would you give the doctor about her ability to meet her objective?

   (4)

8. (i) Distinguish between *accuracy* and *precision* when estimating features of a population from a sample.

   (2)

   (ii) Researchers often want to estimate the proportion of people who agree with a simple statement. Respondents are required to answer either 'yes' or 'no'. For each of the following scenarios discuss, with reasons, the likely accuracy and precision of the results.

   (a) The question is posted on an open-access website. The target population is English-speakers world-wide. A total of 3000 responses is obtained.

   (b) An anonymised postal survey is carried out of members of a small society. Of the 20 members, 18 reply.

   (c) From a society of 1200 members, 100 attend a meeting and indicate their view by a show of hands.

   (9)

9. A company has a large number of employees. All employees have work email addresses and easy access to computers, but a sizeable minority of them are not required to use computers as part of their job. The company routinely distributes email messages that are of interest to employees but not essential to their work. It is known that some employees never read these emails. The management wants to carry out a survey of the staff to establish how employees could be motivated to read emails regularly. They have a very low budget for this activity. What advice would you give them about how to do this?

   (7)

BLANK PAGE

BLANK PAGE