EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2012

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 5 printed pages, **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1.	Answer the following questions in the context of a real survey of which you are aware.

   (i)	Briefly describe the survey in terms of its purpose, the target population, the sampling procedure, and the methods of collecting and processing the data.

   (5)

   (ii)	When carrying out a survey there are many *sources of error*. Identify the possible sources of error in your chosen survey and describe the strategies used to try to minimise them. State, giving reasons, whether you think these strategies were adequate and suggest any modifications.

   (9)

The manager of a hotel wants to use a survey to find out clients' opinions about the hotel. Questions 2–5 relate to this survey.

2.	The manager needs to know the following.
   - The age and sex of the client
   - The client's reason for needing hotel accommodation
   - The client's satisfaction with the booking procedure, the room and the meals
   - What the client liked best about the stay
   - What the client liked least about the stay
   - Any other comments that the client may have.

   Using closed questions where possible, design a self-completion questionnaire to elicit this information.

   (12)

3.	Use examples from the questionnaire you designed in question 2 to illustrate your answers to the following.

   (i)	Explain why it is generally easier to code the answers from closed questions than from open questions.

   (6)

   (ii)	The hotel manager ran the survey in 2011. In his questionnaire he used several open questions, but found it very difficult to analyse these. He decides that he will repeat the survey in 2013 in order to compare results for the two years. Explain how he could replace open questions with closed questions where possible. Discuss the advantages and disadvantages of what you recommend.

   (6)

2

Turn over

4. The manager is undecided about ways of administering the questionnaire. In particular he is unsure whether to

   – leave questionnaires in the hotel rooms or post them to clients' home addresses

   – give the questionnaire to all clients who book in during a 2-month period, or to a proportion of clients who book in during a full year.

   Discuss the advantages and disadvantages of these approaches.

   (6)

5. It is suggested that people often do not like self-completion questionnaires and so clients should be interviewed, either face-to-face or by telephone.

   Discuss the advantages and disadvantages of self-completion questionnaires, face-to-face interviews and interviews by telephone for this particular survey. What would your advice be in this case?

   (6)

6. (i) Two market research companies have run similar surveys on comparable populations. They used the same questionnaire and cluster sampling.

      Company A sent out 5000 questionnaires and received 1000 responses.
      Company B sent out 1500 questionnaires and received 750 responses.

      Company A claims that their survey is better because they have a larger number of responses thereby enabling them to calculate better estimates from the survey data.

      Comment on whether you think this claim is justified, giving your reasons.

      (3)

   (ii) A researcher undertaking a postal survey has been advised that she needs 100 responses in order to estimate mean values with the required precision.

      (a) Explain what is meant by *precision*, and why you might expect it to be lower if fewer than 100 responses were obtained. (3)

      (b) The researcher believes that the response rate to the questionnaire will be about 60%. She does not know whether to send out 167 questionnaires in order to get about 100 back, or to send out fewer questionnaires, followed by reminders to people who do not respond to the first questionnaire.

         Discuss the relative merits of these two approaches. (5)

      (c) Discuss briefly other strategies that the researcher could use to try to minimise non-response to the questionnaire. (6)

3

Turn over

7. A health organisation wants to run a survey to determine what foods mothers are feeding to infants less than one year old.

Discuss the practical problems of obtaining this information in a country with which you are familiar and how they could be overcome for different types of location, such as a large town or a rural area.

(8)

8. An organisation wants to gather information on its employees, 90% of whom are male. It has chosen to use interviews, which will be quite long, and will be conducted during working hours. Some of the questions are sensitive, and people can decline to answer any subset of them.

The average salary of the women is lower than that of the men, so it is more costly to interview the men. It is estimated that an interview with a male employee will cost £50 on average, and that for a female employee £30. The total budget for the interviews is £1000. The organisation has been advised to use stratified sampling.

(i) Explain what is meant by *stratified sampling* and why it might be important here.

(3)

(ii) How many male and female employees could be interviewed if a uniform sampling fraction is used?

(2)

(iii) It is suggested that a uniform sampling fraction might not be the most appropriate method, and that instead the following formula (in the usual notation) should be used:

$$\frac{n_i}{N_i} \propto \frac{S_i}{\sqrt{C_i}} \qquad i = 1, 2.$$

(a) Explain what is meant by each of the terms in the formula. (2)

(b) No one in the organisation knows the values of S_i. What advice would you give? (4)

(c) Assuming that $S_1 = S_2$, use the given formula to calculate the number of male and female employees that could be interviewed. (4)

(iv) Further advice is that, if the mean values of some of the answers for men and women are to be compared, there should be approximately similar numbers of each in the sample. How many men and women could be interviewed using this rule?

(2)

(v) Overall what would your advice be about the numbers of men and women to interview in this study?

Turn over