

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2010

MODULE 1 : Data collection and interpretation

Time allowed: One and a half hours

*Candidates should answer **THREE** questions.*

Each question carries 20 marks.

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

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).

The notation \log denotes logarithm to base e .

Logarithms to any other base are explicitly identified, e.g. \log_{10} .

Note also that $\binom{n}{r}$ is the same as nC_r .

This examination paper consists of 3 printed pages **each printed on one side only**.

This front cover is page 1.

Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. (a) In reviewing a draft questionnaire, the following questions are thought to need amendment. For each question, explain why an improvement is needed and rewrite the question in a better form. The questions will be separated by other questions in the questionnaire, and those numbered 3 to 6 below will be in a section to be answered only by persons currently in employment.

1. Please state the year in which you left full-time education.

day month year

2. Please state your employment status.

Employed full-time Employed part-time (< 40 hours) Student

3. Do you really enjoy your present job?

4. What is your income?

5. How many employees are there where you are employed?

10 or fewer 10-50 more than 50

6. Rate your job satisfaction out of a score of 10.

(14)

- (b) Explain, with examples, why (i) sensitive questions, (ii) hypothetical questions, are likely to be troublesome in questionnaires.

(6)

2. The display below shows the heights in centimetres at ages 6 and 10 years of six girls born to short mothers and six girls born to tall mothers. The girls were selected at random using available medical records, and all were from different families. Write a short report describing the main features of these data, illustrating it with summary statistics. You should show your calculations, but should make clear that they are not part of the report.

Age	Short mother		Tall mother		
	6	10	6	10	
Girl 1	111.0	130.5	Girl 7	120.4	140.7
2	110.0	131.4	8	120.2	146.5
3	113.7	136.0	9	118.9	144.0
4	114.0	134.0	10	120.7	146.0
5	114.5	135.0	11	121.0	144.0
6	112.0	135.2	12	115.9	141.1

(20)

3. (i) Distinguish carefully between stratified sampling, cluster sampling and quota sampling, stating the benefits and drawbacks of each method. (12)
- (ii) A country is divided into regions and in each region there are both urban and rural areas. A survey is to be undertaken in which adults are to be interviewed. Devise a sampling scheme which is a combination of stratified, cluster and quota sampling. (4)
- (iii) A village has 506 inhabitants, listed by name and address in a register. Explain in detail how to take a systematic sample of about 50 inhabitants from this list. State the benefits and drawbacks of this method. (4)
4. (i) Refusal to take part by selected sample units and failure to contact sample units in a face-to-face social survey can both contribute to survey non-response bias. Discuss this statement and suggest ways in which the bias might be reduced. (14)
- (ii) One thousand people were selected to participate in a face-to-face survey. Of the thousand, 725 were at home at the first attempt to contact them but 52 of these refused to take part. One call-back was made to each person missed at the first attempt, and 30 of these people were not at home at the time of call. Of those that were at home, 49 refused to take part in the survey. Calculate the response rate at the first contact, the response rate at the second contact and the overall response rate. (6)