EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2007

(Modular format)

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 $\begin{pmatrix} n \\ r \end{pmatrix}$ *is the same as* $^{n}C_{r}$.

This examination paper consists of 6 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.   The questions **on the next page** are taken from a draft of a questionnaire that is to be sent by mail to members of an environmental organisation X. The information obtained is to be used in attempts to increase membership, but current members are not to be told the true purpose of the survey.

    (i)   Draft a short introduction to the questionnaire, bearing in mind that a high response rate is desirable.

(8)

    (ii)   Draw attention to any features of the draft questionnaire that you feel are particularly worth commending.

(2)

    (iii)   For each question that you feel could be improved, explain what the problem is and suggest how that question might be reworded. Your suggestions could include asking supplementary questions as well as making changes to the wording.

(10)

The extract from the questionnaire is on the next page

MEMBERSHIP SURVEY

Membership matters

Q1a.   How strongly were you influenced by the following factors when you decided to join X? Please rate each factor from 1–5 where 1 means it was not a strong influence and 5 means it was a very strong influence.

To support the environment
1   2   3   4   5

To learn more about the environment
1   2   3   4   5

As a way of meeting like-minded people
1   2   3   4   5

To take advantage of membership special offers
1   2   3   4   5

Q1b.   Were there any other factors that particularly motivated you to join X?

………………………………………………………………………………..

Q2a.   When you first joined X how did you rate the membership scheme in terms of value for money?

Excellent ☐   Very good ☐   Good ☐   Fair ☐

Q2b.   How do you rate the membership scheme in terms of value for money now?

Excellent ☐   Very good ☐   Good ☐

Q3.   Rate the programme of members' activities out of a score of 10. ………..

Q4.   How frequently do you look at the members' handbook?

Frequently ☐   Occasionally ☐   Never ☐

Q5.   If you never look at the members' handbook, can you tell us why? …………….

Background information

Q6.   Sex
Male ☐   Female ☐

Q7.   Age
Under 24 ☐   25-44 ☐   45-64 ☐   65+ ☐

Q8.   Marital status
Married ☐   Single ☐   Widowed ☐   Separated ☐
Divorced ☐   Other partnership ☐

Q9.   What are your main interests apart from the environment? Please write in.
…………………………………………………………………………

Q10.   Which of the following types of charity do you support? Please tick all that apply.
Environmental ☐   Third world ☐   Animal welfare ☐   Health research ☐
Other (please write in)
………………………………………………………………………………..

Turn over

2. The table below shows, for a cross-sectional postal survey of male Gulf War veterans, the numbers of questionnaires sent out at three successive mailings, and the numbers of responses, non-responses, refusals, and undelivered questionnaires at each mailing. At subsequent mailings questionnaires were not sent again to subjects who had responded or had refused to respond following earlier mailings.

	Mailing 1	Mailing 2	Mailing 3
Total questionnaires sent	**4822**	**2721**	**1964**
Responses	2099	701	483
Non-responses	2544	1797	1239
Refusals	2	56	110
Undelivered questionnaires	177	167	132

(i) Write a short report discussing the main points shown in this table. Your report should include the results of any calculations you make from the data, and could contain a flow diagram representing the information in the table.

(8)

(ii) Give two reasons why survey practitioners follow up non-respondents, and discuss briefly how useful a follow-up can be.

(4)

(iii) Discuss whether other courses of action instead of, or as well as, those taken might have increased the response rate in this survey even more. You should mention any disadvantages associated with the suggestions you make.

(8)

Turn over

3.   A local authority is investigating various aspects of the usage of its public library.

(a)   All 85 students chosen in a simple random sample from the 987 secondary school students in the local authority's area have been interviewed. The display below summarises the responses to a question asking how many times the student had visited the library during the last four weeks.

Number of visits	Number of students
0	16
1	21
2	29
3	19
Total	85

(i)   Estimate the total number of visits made to the library during the last four weeks by the 987 students in the school population, and obtain the estimated variance of your estimator.

(7)

(ii)   Estimate the percentage of the 987 students in the school population who visited the library at least once during the last four weeks, and obtain the estimated variance of your estimator.

(5)

(b)   An estimate of the total number of visits made to the library during the previous four weeks is needed for all adults in the area who are registered with the library. The postal addresses at time of registration are known. Two ways of obtaining information about the number of visits have been suggested. One is to stratify by sex and post a short questionnaire to a random sample of adults selected from each stratum using proportional allocation. The other is to take a systematic sample of adults leaving the library on one particular day and ask them to complete a short questionnaire at the time. Discuss the advantages and disadvantages of each method.

(8)

Turn over

4. The display below shows some information for each of 12 tablets taken from a batch of a drug product, and for each of 10 human subjects administered tablets from the same batch. The "in vitro" figures are the percentages of drug dissolved half an hour after the tablet was introduced into a dissolution medium. The "in vivo" figures are the percentages of drug absorbed into the blood stream after half an hour, expressed as percentages of the 24-hour value and estimated by a technique known as deconvolution.

Dissolution medium (in vitro)
5.94 6.22 5.09 6.22 5.37 5.51 4.10 5.51 4.52 4.52 4.10 4.81

Human subjects (in vivo)
3.92 1.55 2.52 0.96 0.84 5.01 2.20 2.51 2.84 0.96

(i) Find the median and quartiles for each of the in vitro and in vivo figures, explaining your method.

(8)

(ii) Represent the in vitro and in vivo figures as boxplots on a single diagram.

(3)

(iii) Calculate the mean and standard deviation for each set of figures.

(4)

(iv) On the basis of your answers to parts (i), (ii), and (iii), compare and contrast the in vitro and in vivo figures. You are <u>not</u> expected to perform any significance tests or to find confidence limits.

(5)