

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



ORDINARY CERTIFICATE IN STATISTICS, 2002

Paper I

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 silent, cordless, non-programmable electronic calculators.

*Where a calculator is used the **method** of calculation should be stated in full.*

Read the following three paragraphs before answering any of questions 1, 2, 3

In some parts of England, local government consists of two levels. There are county councils responsible for large areas, and within these there are borough councils, each of which covers a small section of the county area. Borough councils are responsible for some public services.

One such borough contains 43 342 households and 103 456 residents. It is partly rural and partly suburban, with two main township centres. It publishes a "newsletter" two or three times a year, and this is distributed free to each household in the borough.

In June 2000, the newsletter contained a questionnaire about highway services (roads, street signs and nameplates, covered shelters for passengers waiting for buses, etc) for which the borough council is responsible. People were asked to return the questionnaire to the council offices by post, or to telephone its Highway Services Unit with their response. A total of 52 people did so, and the results appeared in the April 2001 newsletter.

Questions 1, 2, 3 concern the responses to this survey.

1. A borough councillor has questioned whether the results can be taken as a good representation of the views of residents in the borough.

Explain the concept of *bias* in surveys, and discuss ways in which the results of this survey may be biased.

(8)

2. *Reliability* of results means that if the survey were to be repeated in exactly the same way, similar results could be expected.

Discuss whether the results of this survey are likely to have reliability in this sense.

(5)

3. For this borough, explain how you would carry out a survey of residents' opinions by each of the following methods. State the advantages and disadvantages of each method.

(i) A simple random sample using the register of voters and a postal survey. (8)

(ii) A stratified random sample using the register of voters and personal interviewers. (8)

(iii) A random sample selected from the local telephone directory, interviewed by telephone. (8)

4. A survey is to be conducted to estimate the average daily time it takes for the residents of an area to travel to and from work. A pilot survey has been carried out, with the area divided into rural and urban parts. Data from the pilot survey were as follows.

Stratum, i	Population stratum size, N_i	S.D. (estimate), s_i
Rural, $i = 1$	12 000	11
Urban, $i = 2$	80 000	8

The sampling costs, c_i , are thought to be about £18 per response in the rural stratum and £5 per response in the urban stratum.

- (i) Using *proportional allocation*, the sample sizes n_i in each stratum are in the same ratio as the stratum sizes N_i in the population. Find the numbers to be sampled from each stratum by this method if the total sample size is to be 500. Find also the cost of taking this sample. (5)

- (ii) An alternative *optimum allocation* method minimises the variance of the estimate, and in order to obtain this n_i must be taken proportional to $N_i s_i / \sqrt{c_i}$. Also the total budget is £3600, of which £500 will be required for fixed costs. Using the results from the pilot survey, find the sample sizes n_1, n_2 needed by this method. (9)

- (iii) Do you consider the optimum method sufficiently better than proportional allocation to make it worth using for this survey? Justify your answer briefly. (3)

5. Discuss reasons (**other than** variance estimation as used in question 4) for carrying out a pilot survey. (7)

6. An education authority undertakes a survey of 700 parents, to determine their views on the availability of out-of-school facilities (such as sports and other leisure activities) and holiday projects for children in the area. A computer data-analysis program is available to enter and analyse the responses to the survey questionnaire.
- (i) Explain:
- (a) how the questionnaire should be designed so that data may easily be transferred to the computer;
 - (b) how the computer database should be set up;
 - (c) how the accuracy of the data input might be checked;
 - (d) how missing observations might be dealt with.
- (11)
- (ii) Discuss briefly the advantages and disadvantages of computer analysis compared with analysis by hand of the data from this survey.
- (4)
7. The UK Index of Retail Prices is used as a measure of price inflation, and other countries have similar measures. For the Index of Retail Prices or a similar index of your own choice (**you should say which index you are using**):
- (i) explain the way in which the general categories of expenditure are grouped, and how weights are allocated to them;
 - (ii) discuss the problems involved in collecting data on prices;
 - (iii) mention any difficulties in drawing conclusions from such an index when considering expenditure in different types of household – for example those with children, those in various types of work, the elderly.
- (12)
8. For **EACH** of the two methods *quota sampling* and *cluster sampling*, give an example where the method is suitable. State clearly what populations are being studied, what sampling frame is being used, and why the method is better than alternatives.
- (12)