

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



ORDINARY CERTIFICATE IN STATISTICS, 2005

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 calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

This examination paper consists of 4 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.

Questions 1 to 3 relate to the following situation.

A country is divided into 50 administrative districts. In 2006, the Education Ministry plans to undertake a sample survey into the careers and career plans of those who left secondary school during 2003, and wishes to survey about 10% of the relevant population.

The following information will be available about each school in the country: size (number of pupils on 31 January 2006); district; names and contact details of students who left the school in 2003, arranged in order of leaving date.

1. Your answers to this question should relate to the proposed survey.
 - (i) Suggest two different two-stage sampling methods of selecting a suitable sample. Each of these methods should use a different combination of *simple random*, *stratified*, *systematic* and *cluster* sampling, and each of these four methods of sampling should be used once only. You may assume that sampling frames suitable for your purpose exist. (6)
 - (ii) For each of the sampling methods that you have suggested, comment briefly on any advantages and disadvantages it has for this survey. (6)
2. Draft a short self-completion questionnaire to obtain the following information from the members of the sample.
 - Month, year and age when secondary school was left.
 - For each of 2004 and 2005, the occupation and the district in which this was undertaken. (If a respondent had more than one occupation during the year, you should direct him or her to give information about the main occupation, and the dates spent in it. Note that being a student is to be regarded as an occupation; a respondent should be asked to specify the main subject studied.)
 - Long-term career plans. (12)
3. The information which is to be obtained from the respondents, as detailed in question 2, is to be stored in electronic form, together with contact details of the respondents. Indicating which, if any, responses are to be coded:

either design a database suitable for this purpose. Your answer should list the fields, giving appropriate field names, together with field types and suitable widths;

or design a spreadsheet suitable for this purpose. Your answer should list the variables, giving appropriate variable names, together with their types and suitable cell widths. (12)

Questions 4 to 8 relate to the following proposed design for a survey whose main objective is to determine the variation of prices of fruit and vegetables over time and in different regions of a small country where a large proportion of the population experience rural poverty.

The country is divided into 7 regions, 4 mountainous and 3 coastal. One mountainous region and one coastal have been selected as being typical. The survey is to be of a number of markets taking place in rural areas in those two regions. The markets are considered to belong to one of two strata: large markets and small markets. A sample of markets is to be taken from each stratum in each region. Local representatives at each market are to be asked for background information on production and supply.

Each market in the sample is to be visited once a month over a period of four months. At each visit, prices of a selection of fruit and vegetables on sale that day are to be collected by interviews with a number of merchants. The merchants present in a market and the produce on sale are likely to change from day to day.

4. (i) Suggest two advantages and two disadvantages of using a non-random method, such as that described here, to select two regions. (6)
- (ii) Discuss the difficulties, in the proposed survey, of
- (a) selecting merchants to interview,
- (b) deciding for which fruit and vegetables prices should be obtained.
- In each case, suggest how these difficulties might be addressed. (10)
5. (i) Background information, about the production and supply of produce for sale, times of openings and numbers of merchants who have stalls, is required for each market. Local representatives are to be approached to obtain this. What advice would you give as regards to choosing suitable local representatives, and in collecting the information needed from them? (6)
- (ii) Comment briefly on the difficulties of obtaining accurate details, such as last year's prices of products, in a survey undertaken this year. (4)

6. Explain what is meant by *interviewer bias*. Give examples to show how this might occur in the collection of prices in this survey, and suggest how these biases might be reduced.

(10)

7. A study of various aspects relating to a widely eaten staple vegetable was done in another region of the country last year. In the table below, costs and prices are measured in the local currency of the country. The table shows the standard deviations, s_i , of the price per kilo of this vegetable, as found in the previous study. It also shows the numbers N_i of large and small markets in the coastal region to be used in the proposed survey, and the expected costs c_i (in the local currency) of sampling one market of each type.

Stratum	Number of markets (N_i)	SD of price of vegetable (s_i)	Cost of sampling one market (c_i)
Large markets	27	0.05	15
Small markets	53	0.07	12

- (i) Calculate the required number of markets in the sample from each of the strata using a uniform sampling fraction to achieve an overall sample of about 8 markets. What is the total cost of sampling markets using these sample sizes?

(4)

- (ii) The *optimal allocation* method of choosing sample sizes n_i to estimate a mean minimises the variance of the estimate and, in order to obtain this, the n_i must be taken proportional to $N_i s_i / \sqrt{c_i}$. Calculate the required number of markets in the sample from each stratum using this method if the total cost of sampling markets in this region is to be no more than 105 in the local currency.

(8)

- (iii) Which of the two methods of calculating sample sizes used in parts (i) and (ii) do you prefer here, and why?

(4)

8. The information to be collected in the survey is to be used to obtain a price index for fruit and vegetables. By making reference to an official index of prices with which you are familiar, or otherwise, suggest how this might be done. Your answer should indicate what decisions need to be made in defining the index in detail, whether other information needs to be collected in order to calculate it, and how it might be updated over time.

(12)