



EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2014

MODULE 8 : Survey sampling and estimation

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 ${}^n C_r$.

This examination paper consists of 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. The formula for the variance of the estimator of a population mean based on a stratified (random) sample is

$$V = \sum_{h=1}^L W_h^2 \frac{S_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right).$$

Define the terms N_h , S_h , n_h and W_h in the above formula. Explain the conditions under which stratified sampling may be superior to simple random sampling.

(4)

The Chief Education Officer for a region wishes to estimate the total number of children who have played truant in the past week (that is, who have been absent from lessons without good reason). The region is divided into four education authorities (strata) and a random sample of ten schools is taken from each education authority. The results are as follows.

<i>Education authority h</i>	<i>Total number of schools</i>	<i>Number of children who have played truant (y) in schools selected</i>	<i>Sample mean</i>	<i>Sample standard deviation</i>
1	141	4, 8, 10, 0, 1, 4, 0, 12, 1, 0	4.0	4.50
2	471	5, 15, 6, 9, 8, 15, 17, 10, 6, 16	10.7	4.62
3	256	23, 26, 11, 23, 14, 17, 33, 0, 6, 22	17.5	9.92
4	1499	2, 3, 3, 3, 4, 0, 3, 1, 2, 3	2.4	1.17

- (i) Estimate the total number of children who have played truant in the past week and obtain an approximate 95% confidence interval for this total.

(7)

- (ii) The Officer wishes to report estimates of the total number of children who have played truant in the past week for each of the four education authorities, as supporting information. Obtain a point estimate and an approximate 95% confidence interval for this total number for education authority 1.

(4)

- (iii) The Officer is planning a new survey, and is intending to sample an equal number of schools from each authority in the region. Giving reasons, suggest another allocation method that might be preferred to compute the sample sizes in each authority. Use this method to compute the stratum sample sizes for a sample of 40 schools.

(5)

2. In a certain country, local government consists of two levels. There are regional councils for large areas, and within these there are district councils, each of which covers a small section of the area. District councils are responsible for some public services.

One such district 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 district.

In June 2011, the newsletter contained a questionnaire about highway services (roads, street signs and nameplates, covered shelters for passengers waiting for buses, etc) for which the district 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 2012 newsletter.

- (i) A district councillor has questioned whether the results can be taken as a good representation of the views of residents in the district.

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

(5)

- (ii) For this district, 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.

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

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

(c) A simple random sample selected from the local telephone directory, interviewed by telephone.

(15)

3. A forest resource manager is interested in estimating the number of dead fir trees in a 200-hectare area. Using an aerial photograph, he divides the area into 200 one-hectare plots and estimates the number of dead fir trees in each plot. The total number of dead fir trees in the 200 plots estimated from the photograph count is 4200. He then selects a simple random sample of 10 plots and counts the exact number of dead trees on the ground in each of these. The data for these 10 plots are given below.

<i>Plot Sampled</i>	<i>Photograph Count</i>	<i>Ground Count</i>	<i>Plot Sampled</i>	<i>Photograph Count</i>	<i>Ground Count</i>
1	12	18	6	30	36
2	30	42	7	12	14
3	24	24	8	6	10
4	24	36	9	36	48
5	18	24	10	42	54

- (i) Using the mean of a simple random sample, estimate the total number of dead fir trees in the 200-hectare area based on the ground count, and the standard error of your estimator. (6)
- (ii) The manager wants to know whether the supplementary information from the photograph could be used to improve the estimation of the total number of dead fir trees.
- (a) Using the above sample data, plot the ground counts against the photograph counts. Use this diagram to comment on the suitability of a ratio estimator. (6)
- (b) You are given that the ratio estimate of the total number of dead fir trees in the 200-hectare area is 5492.31 and its estimated standard error is 214.22. Giving a reason, state whether this estimate is better than that calculated in part (i), and use it to construct an approximate 95% confidence interval for the true total number of dead fir trees in the 200-hectare area. Explain what this confidence interval shows. (4)
- (iii) For a future survey, the forest manager is wondering whether to use simple random sampling or systematic sampling. Explain the term *systematic sampling*. Give a reason why systematic sampling might be preferred. What might be its drawbacks? (4)

4. (a) A survey is planned to study family income in a mixed urban and rural population. Discuss any practical difficulties that might arise in defining "income", in defining "family", and in combining information from rural and urban areas.

(5)

- (b) A survey organisation defines the 'true level of business confidence' for a particular sector of economic activity as the proportion of managing directors of all companies in that sector who expect prospects for their company to improve in the next six months.

In a pilot survey in the light engineering sector, the managing directors of 67 out of a random sample of 125 companies stated that they expected prospects for their company to improve in the next six months.

- (i) Using this information, obtain an approximate 95% confidence interval for p , the proportion of companies expecting an improvement. Explain what this confidence interval shows. You may assume that N , the number of companies in this sector, is much larger than 125.

(5)

- (ii) A business analyst wants to calculate an approximate 95% confidence interval for p . What sample size would be required to produce a 95% confidence interval which has a width of 0.08?

(4)

- (c) 'Smart meters' are to be installed in every home in the UK by 2020. These devices will show exactly how much gas and electricity is used at any specific time, which could help households to be more energy efficient and cut their bills. There is limited evidence of how much and for how long British consumers' behaviour might change after installing these meters. You have been asked to design a survey to assess this. Explain briefly the difference between *longitudinal* and *cross-sectional* surveys, and how the distinction might be relevant in this situation. What might be the drawbacks of a longitudinal survey here?

(6)

BLANK PAGE

BLANK PAGE

BLANK PAGE