

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2010

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 nC_r .

This examination paper consists of 5 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. In a study of the advertising expenditure of small businesses in Kingstown, a stratified random sample of small businesses gave the following results.

<i>Employment size-band</i>	<i>Number of businesses</i>	<i>Achieved sample size</i>	<i>Mean advertising expenditure in 2008 (£000's)</i>	<i>Standard deviation of advertising expenditure</i>
h	N_h	n_h	\bar{y}_h	s_h
Fewer than 5	10 000	100	3.5	3
Between 5 and 9	6 000	60	10.0	5
Between 10 and 20	2 000	15	35.0	15
TOTAL	18 000	175		

- (i) Construct a 95% confidence interval for the mean advertising expenditure of all businesses that employ fewer than 5 employees. (3)
- (ii) Explain carefully what this confidence interval means. (2)
- (iii) In order that this confidence interval shall be no greater in width than £500, what achieved sample size would have been necessary? (4)
- (iv) Construct a 95% confidence interval for the mean advertising expenditure for all small businesses in Kingstown. (6)
- (v) The sample for this study was selected from a detailed sampling frame of local businesses. It has been suggested that this frame might enable ratio estimation to be used in future years.

Give an example of a supplementary variable x that a detailed frame might contain and which could help to adjust the prediction of the mean advertising expenditure \bar{y} using ratio estimation.

Give reasons for your answer, including an explanation of the qualities needed from a supplementary variable.

(5)

2. A proposed new road round the community of Chiptown has provoked some debate. On the one hand it should reduce traffic levels in the town but on the other hand it will cut through an area of unspoilt beauty.

The local newspaper has conducted a survey of its readers to determine their views on the proposed road. The sample was selected by asking newspaper readers to text a simple 'yes' or 'no' vote to the question "Are you in favour of the proposed road to divert dangerous traffic away from Chiptown?"

The results from the survey were that 110 out of an achieved sample of 150 said 'yes'. The population of the town is 25 000.

- (a) A local academic has written to the newspaper expressing serious misgivings about the validity of the survey process. He criticised the sampling frame, the data collection methodology and the question being posed. He concluded that the survey results were very likely to be biased.

(i) Explain what is meant by bias. (2)

(ii) Outline what you believe his specific criticisms of the survey process might have been. You should discuss how these factors might influence bias, but not limit your discussion solely to the issue of bias. (12)

- (b) Another resident has written in, pointing out that Chiptown is a mixture of areas built up at various times in the past 100 years. He says that a cluster sample should have been used. Discuss the possible advantages of this approach, and suggest how clusters should be constructed and sampled. (6)

3. The Learning and Skills Office (LSO) has the role of ensuring that a particular country has the skills necessary for the future. A random sample has been taken in 2009 from students in secondary education. They have been asked the question "Do you intend to study for a university degree?" In one region there were 12 000 students altogether on these courses, 8000 in schools in areas classified as urban and 4000 in rural schools. The results in this region were as follows.

School location h	Population of secondary students N_h	Achieved sample size n_h	Proportion intending to study for a university degree p_h
Urban	8 000	300	0.3
Rural	4 000	150	0.6
TOTAL	12 000		

- (i) Identify two purposes of stratification. How might these be relevant in this survey? (3)
- (ii) Construct 95% confidence intervals for the proportions of students in schools in each type of area, urban and rural, who intend to study for a university degree. (3)
- (iii) Test whether the proportions of students who intend to study for a university degree differ between the urban and rural schools. (4)
- (iv) Estimate the total number of students in this region who intend to study for a university degree. Construct a 95% confidence interval for this total number. (4)
- (v) The LSO plans to run a similar survey next year, with the sample size in this region being 500. The Statistics Officer has been recommended to use either *proportional allocation* or *optimal allocation* when designing the survey. Explain what these terms mean, and find the stratum sizes that each method would give using the information you have from the 2009 sample. (6)

4. The Southern England Chamber of Commerce set out to assess business confidence in the region, in the light of a global recession. In particular they were interested in employment prospects in the region.

A survey was carried out on a simple random sample of businesses that were members of the Chamber of Commerce. Of these businesses, 100 responded, and 40% anticipated increasing their workforce in the next 12 months.

This led to the following headline in a local newspaper: "Figures released by the Southern England Chamber of Commerce show that most businesses will be shedding jobs in the next 12 months".

- (i) What reservations would you have about the newspaper headline, from the viewpoints both of the wording and of what you know about the survey upon which it was based?
(7)
- (ii) Outline a possible survey methodology that the Chamber of Commerce could use to answer the research question more robustly. Your discussion should include issues such as identifying the population, devising a suitable sampling frame, conducting the survey and analysing the results.
(13)