

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2013

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 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. (a) A survey is being planned into the cost of house repairs in a certain district with 43 000 households. The budget for the survey is \$20 000. Two possible designs have been suggested.

Design 1: a simple random sample, at a cost of \$4000 plus \$10 per house sampled.

Design 2: a stratified random sample using 5 classes of house (I to V) as the strata, at a cost of \$10 000 plus \$10 per house sampled.

Rough estimates of the means and standard deviations of the cost of house repairs in the previous year, in each stratum, are as follows.

	I	II	III	IV	V
Percentage of houses in class	10	10	20	30	30
Mean cost of repairs (\$)	1350	1100	850	600	400
Standard deviation of cost (\$)	600	400	300	200	160

- (i) For each design, determine the number n of households that would be sampled on the total budget of \$20 000.

For Design 2, compute the sample sizes in each class that give optimal allocation, and explain why such allocation might be beneficial in this instance.

(7)

- (ii) For each design (using optimal allocation for Design 2), calculate an estimate of the variance of the estimated mean cost of house repairs for this district. You may assume that an estimate S^2 of the overall variance is 182 205.

(7)

- (iii) Giving a reason, state which design you would recommend.

(1)

- (b) Suggest a suitable sampling frame for selecting a simple random sample of households for a survey. In what way does this sampling frame differ from the target population?

(5)

2. A simple random sample of 10 hospitals was selected from a population of 33 hospitals that had received state funding to upgrade their emergency medical services. Within each of the selected hospitals, the records were examined for all patients hospitalised in the past 12 months for trauma, which is defined as body wounds or shock produced by sudden physical injury due to accident or violence. The numbers of patients hospitalised for trauma, and the numbers of patients with trauma who were discharged dead, for the selected hospitals are given below.

<i>Hospital</i>	<i>Number of patients hospitalised for trauma</i>	<i>Number with trauma discharged dead</i>
1	560	4
2	190	4
3	260	2
4	370	4
5	190	4
6	130	0
7	170	9
8	170	2
9	60	0
10	110	1

- (i) Explain why this design may be considered as a cluster sample. What are the first-stage and second-stage units? (3)
- (ii) (a) Estimate the total number of people hospitalised for trauma for the 33 hospitals. State a property of your estimator. Give an approximate 95% confidence interval for this total number. (7)
- (b) Let p be the true proportion of people suffering from trauma discharged dead. Explain why an estimator of p based on the above cluster sample is a ratio estimator. You are given that the 95% confidence interval for p based on use of a ratio estimator is (0.006, 0.021). Explain what this confidence interval shows. (4)
- (iii) Give reasons why, for this survey, cluster sampling might be preferred to stratified random sampling using each of the 33 hospitals as a stratum. What might be the drawbacks of cluster sampling?

Discuss any improvements you might make if another survey was being planned on the same topic.

(6)

3. A national bus company sets out to determine whether people are changing the way they travel in a country where fuel prices have been rising.

An international internet-based market research organisation is commissioned to conduct a survey. Responses are obtained from an invited group of individuals selected from its large panel of members. A person either registers as a member by visiting the organisation's website or is proactively recruited by the organisation. Each panel member completes a detailed profiling questionnaire at registration. This information is used to select a sample of panel members each time a survey is conducted. All members receive a small incentive for each survey they complete.

This survey leads to the following article in a local newspaper: "Motorists turn to public transport as fuel price bites. More than three in five drivers are turning to public transport due to high fuel prices, a survey has revealed. The survey by a national bus company has found 61 per cent of car users are either definitely or probably considering using public transport because of the rise in prices at the pumps. The bus company has said that the survey, of more than 2000 people, shows 'clear evidence' that people are changing the way they travel."

- (i) What reservations do you have about the newspaper article? (5)
- (ii) Define what is meant by a *sampling frame*. What is the sampling frame from which this sample of respondents was drawn? Indicate typical problems likely to be associated with using this type of sampling frame. (6)
- (iii) What further information do you need to assess whether the method of selecting the sample in this survey can lead to bias? Discuss briefly whether or not this information could be obtained by additional questions in the survey. (6)
- (iv) It has been suggested that it would be easier to achieve a representative sample of respondents with a telephone survey, because almost everyone has a telephone. How would you respond? (3)

4. An auditor is to examine the accounts of a certain company. The company reports that the total amount of money due at a specific time is \$172 500, and the auditor wishes to check this total. A simple random sample of 20 accounts is selected from the 500 accounts with money due to the company. The amounts due, and whether the underlying documents are in compliance with stated procedures, are recorded for each account. The data are as follows.

<i>Account</i>	<i>Amount</i> \$	<i>Compliance</i>	<i>Account</i>	<i>Amount</i> \$	<i>Compliance</i>
1	398	Yes	11	270	No
2	275	Yes	12	304	No
3	444	Yes	13	132	Yes
4	135	No	14	80	Yes
5	123	Yes	15	204	Yes
6	480	Yes	16	53	Yes
7	444	No	17	267	No
8	416	Yes	18	317	Yes
9	317	Yes	19	314	No
10	241	Yes	20	437	Yes

You are given that the sum of the amounts due is 5651 and the sum of squares of the amounts due is 1 918 589.

- (i) Using these data, obtain a point estimate and an approximate 95% confidence interval for the total amount of money due to the company.

The company reported a figure of \$172 500 as the total amount due. With reference to your 95% confidence interval, state whether the auditor should accept this total amount as valid.

(8)

- (ii) A secondary objective of sampling was to check whether the underlying documents were in compliance with the audit controls set up by the company. Using the data given on the compliance checks, obtain a point estimate and an approximate 95% confidence interval for the proportion of unpaid accounts not in compliance with stated procedures. Explain what this confidence interval shows.

(5)

- (iii) Explain what is meant by a *simple random sample* of 20 accounts. What might be the advantage of the auditor using systematic sampling to select a sample of 20 accounts? With 500 unpaid accounts, comment on whether a 1-in-25 systematic sample would be an appropriate scheme. Explain how you would carry out this sampling scheme, and comment on whether it would be reasonable to use the simple random sampling formula you used in parts (i) and (ii) when calculating confidence intervals.

(7)

BLANK PAGE

BLANK PAGE

BLANK PAGE