

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2009

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. The consultancy business 'Evalutec' specialises in evaluating projects in the public sector. It has recently been evaluating the work of the Weston-on-Sea Development Agency (WOSDA), in part by examining a sample of WOSDA's projects. One of the criteria examined was the question "Has the project been completed within the allocated budget?" The results of the study are shown in the following table.

<i>Size of project</i> h	<i>Number of projects completed</i> N_h	<i>Sample size</i> n_h	<i>Proportion of projects completed within the allocated budget</i> p_h	<i>Uncorrected standard error of the proportion p_h</i> $SE(p_h)$	<i>Corrected standard error</i> $SE_c(p_h)$
Large	60	30	0.4	0.0894	
Medium	200	40	0.6	0.0775	0.0693
Small	400	80	0.7		0.0458
ALL	660	150			

- (i) The uncorrected standard errors in the fifth column are those obtained from the standard formula

$$SE(p_h) = \sqrt{\frac{p_h(1-p_h)}{n_h}}$$

Use this formula to calculate the uncorrected standard error for the proportion of small WOSDA projects completed within the allocated budget.

(1)

- (ii) The sixth column in the table shows the standard errors after adjusting the corresponding variances using the finite population correction factor $(1-f)$, where f is the sampling fraction.

- (a) Explain the circumstances under which the finite population correction factor should be used. Why is its use to be recommended in this example?

(4)

- (b) Calculate the corrected standard error for the proportion of large WOSDA projects completed within the allocated budget.

(2)

- (c) Calculate a 95% confidence interval for the proportion of large WOSDA projects completed within the allocated budget.

(2)

- (d) Calculate a 95% confidence interval for the proportion of all WOSDA projects completed within the allocated budget.

(7)

- (iii) Discuss briefly the advantages and disadvantages of stratification. How do these apply in this study?

(4)

2. Suppose that you have recently joined a company that carries out surveys for manufacturing, business, government and medical units. It now wishes to tender for a wider range of work than it has been doing. You are asked to run a training course which provides a refresher in the commonly used sampling methods and also introduces some more advanced ones.

Specific requests from staff are for information on the topics listed below. Make brief notes on each of these topics, suggesting simple examples that could be used for illustration. The mark allocation for your answer to each topic is shown against the topic.

- how to relate study populations to target populations (3)
- how to obtain suitable sampling frames in the areas mentioned above (4)
- when and when not to use non-probability methods such as quota sampling (5)
- how to use supplementary information, in the form of a measurement x , to adjust the prediction of the target measurement y by ratio and regression methods (5)
- when to recommend a longitudinal sample survey and when a cross-sectional one (e.g. in medical work) (3)

3. In 2008 the Students' Union at the University of Kingswood carried out a survey of students in its three faculties. One of the survey questions was "In a typical week, how many hours do you study? (Include lecture/seminar attendance as well as study in your own time)".

The results were as follows.

<i>Faculty</i> h	<i>Number of students</i> N_h	<i>Achieved sample size</i> n_h	<i>Mean number of hours studied</i> y_h	<i>Standard deviation of hours studied</i> s_h
Technology	800	40	45	15
Social Sciences	2000	50	40	10
Medicine	600	40	55	10
	3400	130		

- (i) (a) Construct a 95% confidence interval for the mean number of hours studied by all students within the Faculty of Social Sciences. (3)
- (b) The President of the Students' Union would have liked the interval in part (a) to be no greater in width than 4 hours. What achieved sample size within the Faculty would have been necessary to meet this requirement with at least 95% confidence? (4)
- (c) Construct a 95% confidence interval for the mean number of hours studied by all the students. (6)
- (ii) The Students' Union is planning to make this an annual survey. In 2009 the survey will be repeated with a sample size of 200.

Explain what is meant by each of the terms *proportional allocation* and *optimal allocation*.

Using the 2008 data, find the necessary stratum sizes for each Faculty using each of these methods.

Explain which of these allocation methods is likely to be preferable for this survey. (7)

[**Note.** Optimum allocation has stratum sample sizes proportional to $N_h s_h$ (in the usual notation as used above).]

4. A large city in England is currently looking at a number of schemes to reduce traffic congestion in the city centre. One possible scheme includes removing a popular cycle track and replacing it with a tramline.

- (i) The City Council has set up a consultation survey on its website. This survey shows that 60% of respondents are against the proposed tramline. What reservations would you have about this figure? (3)
- (ii) Outline a better approach to estimating the popularity or otherwise of the council's proposals for removing the cycle track and replacing it with a tramline.

Your answer should include consideration of the population of interest, how a sampling frame might be constructed, how the survey could be conducted, what sampling methodology should be used and how bias could be minimised. (17)

[**Note.** Credit will be given for all valid points made in well thought out answers.]