

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2007

(Modular format)

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. (a) An academic researcher conducted a survey of the 800 students within her faculty. The objective of the study was to examine student debt within the faculty. A simple random sample of 50 of these 800 students was selected, and a questionnaire was sent to each of them. Questionnaires were completed and returned to the researcher by 25 of these students.

In these responses, the mean size of student loan was £8000 and the standard deviation was £3000.

- (i) Calculate an approximate 95% confidence interval for the mean student loan amongst students in the faculty.

The finite population correction factor is $1 - f$, where f is the sampling fraction. Explain the circumstances under which it should be used. Would it make much difference in this situation?

(7)

- (ii) The researcher was hoping to estimate the mean student loan to within $\pm£1000$. Estimate the smallest achieved sample size that would have been necessary to do so with 95% confidence.

(5)

- (b) At the same time as this research was being carried out, an academic colleague conducted a similar survey in a large London college. In a bid to increase response rates, he offered all respondents a £10 gift token for a local store. He received 40 completed questionnaires out of 50.

One of the key variables considered was the proportion of students who had taken out a student loan. From the 40 respondents, 25 had taken out a student loan.

Use a suitable approximation to calculate a 95% confidence interval for the proportion of students at the London college who had taken out a student loan.

(4)

- (c) Discuss any concerns you might have about the planning or conduct of these surveys, especially regarding possible biases.

(4)

2. The owners of a small provincial airport have applied to the local authority for permission to expand and upgrade it so that it becomes a major international airport.

The owners of the airport claim that 70% of the public are in favour of the scheme, as measured in a poll on the airport website.

The local council are not so convinced, and wish to test local opinion themselves.

- (i) Outline some factors which might result in bias in the airport owners' claim. (5)
- (ii) Outline a possible survey methodology for the local council to use to test public opinion more rigorously. You should discuss problems such as how to identify the relevant population, how to devise a suitable sampling frame, and how you might carry out the sampling. (15)

3. (a) In 2005, researchers in a Government Department wished to examine the degree of absenteeism amongst employees. As it was time consuming to interrogate the Department's personnel system, the researchers took a stratified random sample of personnel records, with the stratification being by pay grade (A is the highest grade). The results are summarised below.

<i>Pay grade</i>	<i>Number of employees</i>	<i>Sample size</i>	<i>Mean number of days of absence</i>	<i>Variance</i>
h	N_h	n_h	\bar{y}_h	s_h^2
A	1200	40	4.0	1.5
B	5000	40	6.5	2.5
C	3800	40	8.0	3.0
Total	10000	120		

- (i) Estimate the mean number of days of absence for all employees. (2)
- (ii) Estimate the standard error of the mean number of days of absence for each of the three pay grades. (3)
- (iii) Estimate the standard error of the mean number of days of absence for all employees. Hence calculate an approximate 95% confidence interval for the mean number of days of absence for all employees. (8)
- (b) In 2007, researchers are to repeat the investigation to determine whether the absenteeism rate has improved. It has been suggested that the stratification for the 2007 survey could be improved by using optimal allocation. This would use information from the 2005 study, and a total sample size of 120.
- (i) In optimal allocation, stratum sample sizes n_h are proportional to $N_h s_h$ (using the notation of part (a)). Find the values of n_h that give optimal allocation, and explain why such allocation might be beneficial in this instance. (4)
- (ii) Find the values of n_h given by proportional allocation and comment on whether the precision of the estimated mean is likely to be much affected by which method is used. (3)

4. A Government Department carries out a regular attitude survey of its employees. One of the key questions of interest is "Are you actively seeking employment outside the department?" Because the Finance Division in particular is experiencing retention problems, the population of employees has been stratified into two strata – Finance Division staff and Others.

The results for this key question in 2006 are summarised below. The stratum population and sample sizes are N_h , n_h ; and p_h is the sample proportion actively seeking work outside.

<i>Division</i> <i>h</i>	N_h	n_h	p_h
Finance	1000	100	0.4
Other	15000	100	0.2
Total	16000		

- (i) Identify two key purposes of stratification. How do these apply in this survey?
(4)
- (ii) Construct approximate 95% confidence intervals for the proportions of staff in the Finance Division, and in the other divisions, who are actively seeking work outside. From these confidence intervals, what might be said about the Finance Division employees as compared to the other employees?
(4)
- (iii) Estimate the total number of employees in the Finance Division who are actively seeking work outside, and construct an approximate 95% confidence interval for this total.
(3)
- (iv) The Head of the Finance Division looks at the confidence intervals found in part (ii) and says that what he really wants to know is whether his Division is any different from the others. Carry out a test to compare the two proportions, and comment on the information it gives.
(5)
- (v) In previous years, the attitude survey has always been conducted through the use of an anonymous questionnaire. For administrative reasons, it is being suggested that respondents put their names on the questionnaire. What are the advantages and disadvantages of doing so?
(4)