

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2008

(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. (i) A tourism board for a county in the south west of England has conducted a survey of small hotels (defined as hotels with 12 or fewer rooms). One of the key objectives was to establish the extent to which small hotels used the local tourist information centre to obtain bookings. The board's research officer has used a simple random sample of hotels.

Of the 60 responses received, 45 hotels used the local tourist information centre to obtain bookings.

- (a) Calculate an approximate 95% confidence interval for the proportion of all small hotels in the county which use the local tourist information centre to obtain bookings. Explain what this confidence interval means in language which the head of the local tourism board would understand.

(6)

- (b) The research officer has been told about something called the "finite population correction factor", defined as $1 - f$, where f is the sampling fraction. Explain to the research officer the circumstances under which it should be used. If there are around 6000 small hotels in the county, should the finite population correction be used? Justify your answer.

(4)

- (ii) Within the same survey, hotel keepers were asked about their room prices. The average price of a single room was £50 per night and the standard deviation of price was £15.

- (a) Calculate a 95% confidence interval for the mean single room price for all small hotels in the county at the time of the survey.

(3)

- (b) Without carrying out any detailed calculations, explain whether a 99% confidence interval for the mean single room price for all small hotels in the county would be wider or narrower than a 95% confidence interval.

(2)

- (c) The research officer was hoping to estimate the mean room price to within $\pm£3$. Estimate the smallest achieved sample size that would have been necessary to do so with at least 95% confidence.

(5)

2. In 2007, a large organisation surveyed its staff in three different divisions as part of a "Work/Life Balance" initiative.

Among the questions asked was "How many hours did you spend at work last week?" The results are shown below.

<i>Division</i>	<i>Number in division</i>	<i>Sample size</i>	<i>Mean number of hours worked</i>	<i>Standard error of mean number of hours worked</i>
h	N_h	n_h	\bar{y}_h	$SE(\bar{y}_h)$
Information Management	300	50	41	1.5
Health Statistics	400	50	40	2.0
Internal Communications	100	20	46	2.5
ALL	800	120		

- (i) (a) Construct a 95% confidence interval for the mean number of hours worked in the Internal Communications Division. (2)
- (b) The organisation has a target that the mean working week should be 40 hours. Based upon your confidence interval, is this target being met in that Division? (2)
- (ii) (a) Construct a 95% confidence interval for the mean number of hours worked by all employees. (5)
- (b) Based upon your confidence interval, is the organisation's target of a 40-hour mean working week being met for all employees? (1)
- (iii) Comment on the results found in parts (i) and (ii). (1)
- (iv) Identify two key purposes of stratification. How do these apply in this survey? (3)
- (v) The survey will be repeated in 2008, and work has started on the sampling methodology. The total sample size is to be 120. The researcher responsible has heard the terms *proportional allocation* and *optimal allocation*. Explain what these terms mean, and find the stratum sample sizes each method of allocation would give using the 2007 data. (6)

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

3. A large supermarket company plans to build a new store in a town which is situated in a popular retirement area. The local Chamber of Commerce has conducted a survey of local residents to determine their views on this proposed store.

The survey was conducted by questioning shoppers on the town's main street. The key question asked was "Do you approve of the proposed supermarket scheme?" The results for this question are summarised in the table below.

<i>Sex</i>	<i>Population of each sex in the town</i>	<i>Achieved sample size</i>	<i>Number who approve of the supermarket scheme</i>
Female	28000	100	32
Male	21000	100	65

- (i) (a) Construct 95% confidence intervals for the proportions of residents of each sex who approve of the proposed supermarket scheme. From these confidence intervals, what might be said about the views of females compared to those of males? (5)
- (b) Estimate the proportion of the total population of the town who approve of the supermarket scheme. Express this estimate as a 95% confidence interval. (5)
- (ii) A local academic has claimed that the extent of the difference identified between the views of the two sexes is misleading, and that much of the difference can be attributed to differences in opinions of people of different ages. The supermarket chain is also unhappy with the survey results, claiming that the method of selecting the sample was biased.
- (a) Explain what is meant by bias. (2)
- (b) What assumptions is the academic making? To what extent do you think his claim is valid? (3)
- (c) What further information would be needed to assess whether the method of selecting the sample could lead to bias? Discuss briefly whether or not this information could be obtained by asking additional questions in the survey. (5)

4. The Faculty of Mathematics and Engineering at the University of Bathford has recently invested heavily in web-based learning materials. This enables lecturers to develop new ways of presenting material, and gives students the opportunity of studying from home.

The web development team is keen to find out what staff and students think of the website itself, and of the materials within the website. A research officer has been asked to develop a survey to elicit these views.

- (i) An obvious and easy way of setting up the survey would be to mount it on the website. What would be the key advantages and disadvantages of using this approach alone?

(4)

- (ii) Outline a possible survey methodology that the web team could use to elicit the opinions of staff and students. You should discuss factors such as identification of the population, constructing a sampling frame, the sampling methodology and how the survey should be conducted.

(16)

[Note. As is the case with most real-life surveys, there is not a single "best" approach, but credit will be given for well thought out answers.]