



MASENO UNIVERSITY
UNIVERSITY EXAMINATIONS 2016/2017

**THIRD YEAR FIRST SEMESTER EXAMINATIONS FOR THE
DEGREE OF BACHELOR OF SCIENCE WITH INFORMATION
TECHNOLOGY**

MAIN CAMPUS

MAS 307: THEORY OF SAMPLING TECHNIQUES

Date: 3rd December, 2016

Time: 8.30 - 11.30 am

INSTRUCTIONS:

- Answer question ONE and any other TWO questions.
- Start each question on a fresh page.
- Indicate question numbers clearly at the top of each page.
- Observe further instructions on the answer booklet.



EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE
SIRIBA CAMPUS

MAS 307: THEORY OF SAMPLING TECHNIQUES

TIME: 2 HOURS

Instructions:

The paper consists of five questions

Answer **Question 1** and any other two questions

Observe further instructions on the answer booklet

Question 1 (30 Marks)

- a) Define the following concepts as used in statistics
- i) Estimator
 - ii) Sampling distribution of a statistic
 - iii) Error of estimation
 - iv) Confidence Interval
- (8 Marks)
- b) State two situations when
- i) the ratio estimator is more efficient than mean per unit
 - ii) you would use a regression estimator instead of a ratio estimator
- (2 Marks)
(2 Marks)
- c) In a survey to determine the amount of crop yield due to an air pollutant, a simple random sample of $n = 20$ plots was selected from $N = 1000$ in the population. The summary statistics on yield y_i (in weight) and level of pollutant x_i (in parts per million) were $\bar{y} = 10$, $\bar{x} = 6$, $\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y}) = -60$, $\sum_{i=1}^{20} (x_i - \bar{x})^2 = 30$ and $\sum_{i=1}^{20} (y_i - \bar{y} - bx_i)^2 = 80$. The mean pollutant level is $\mu_x = 5.0$
- i) Estimate the mean yield for the population with a linear regression estimate
 - ii) Estimate the variance of the linear regression estimate in (i) above
 - iii) Predict the yield on a plot in which the pollutant level is $x_i = 4$
- (3 Marks)
(4 Marks)
(3 Marks)

The following results were obtained from a stratified random sample:

Stratum 1: $N_1 = 100, n_1 = 50, \bar{y}_1 = 10, s_1^2 = 2800$

Stratum 2: $N_2 = 50, n_2 = 50, \bar{y}_2 = 20, s_2^2 = 700$

Stratum 2: $N_2 = 300, n_2 = 50, \bar{y}_2 = 30, s_2^2 = 600$

- i) Estimate the mean for the whole population.
ii) Give a 95% confidence interval for the mean.

(3 Marks)

(5 Marks)

Question 2 (20 Marks)

a) A local authority is investigating various aspects of the usage of its public library. All 85 students chosen in a simple random sample from the 987 secondary school students in the local authority's area have been interviewed. The display below summarizes the responses to a question asking how many times the student had visited the library during the last four weeks.

Number of visits	Number of students
0	16
1	21
2	29
3	19
Total	85

- i) Estimate the total number of visits made to the library during the last four weeks by the 987 students in the school population, and obtain the estimated variance of your estimator.
ii) Estimate the percentage of the 987 students in the school population who visited the library at least once during the last four weeks, and obtain the estimated variance of your estimator.

(7 Marks)

(5 Marks)

b) An estimate of the total number of visits made to the library during the previous four weeks is needed for all adults in the area who are registered with the library. The postal addresses at time of registration are known. Two ways of obtaining information about the number of visits have been suggested. One is to stratify by sex and post a short questionnaire to a random sample of adults selected from each stratum using proportional allocation. The other is to take a systematic sample of adults leaving the library on one particular day and ask them to complete a short questionnaire at the time. Discuss the advantages and disadvantages of each method.

(8 Marks)

Question 3 (20 Marks)

There are three wards for patients requiring intensive nursing care in a particular hospital. Wards A and B are in the main hospital building, and ward C is at a remote site. Samples of patients are to be taken from each ward and interviewed about various aspects of their stay. The table below shows the number of patients (N_i) in each ward, and the estimated cost (c_i) in dollars of interviewing a patient in that ward. It also shows the standard deviation (s_i) of the length of time (in minutes) that patients waited to see a nurse on admission to the ward, obtained from a survey of patients who were in that ward in May last year (none of these patients is still in the hospital).

Ward	Number of patients (N_i)	Cost of sampling one patient (c_i)	SD of length of time waited (s_i)
A	50	5	1.81
B	25	5	3.23
C	10	10	2.18

- i) Using a uniform sampling fraction, calculate the required number of patients to sample from each of the wards to achieve an overall sample of about 36 patients. What is the total cost of sampling patients using these sample sizes? (6 Marks)
- ii) The optimum allocation method of choosing sample sizes n_i to estimate a mean minimizes the variance of the estimator, and, in order to obtain this, the n_i must be taken proportional to $N_i s_i / \sqrt{c_i N_i}$. Calculate the required number of patients in the sample from each ward using this method if the total cost of sampling patients is to be no more than 200 dollars. (10 Marks)
- iii) Which of the two methods of estimating sample sizes used in parts (i) and (ii) do you think is better for selecting the sample of patients, and why? (4 Marks)

Question 4 (20 Marks)

In a city of 72,500 people, a simple random sample of four households is selected from the 25,000 households in the population to estimate the average cost on food per household for a week. The first household in the sample had 4 people and spent a total of \$150 in food that week. The second household had 2 people and spent \$100. The third, with 4 people, spent \$200. The fourth, with 3 people, spent \$140.

- a) Identify the sampling units, the variable of interest, and any auxiliary information associated with the units. (3 Marks)
- b) Describe two types of estimators for estimating the mean expenditure per household for a week's food in the city. Summarize some properties of each estimator. (4 Marks)
- c) Estimate mean expenditure using the first estimator, and estimate the variance of the estimator. (6 Marks)
- d) Estimate mean expenditure using the other estimator, and estimate the variance of the estimator. (6 Marks)
- e) Based on the data, which estimator appears preferable in this situation? (1 Mark)

Question 5 (20 Marks)

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 summarized below. The stratum population and sample sizes are N_h , n_h ; and p_h is the sample proportion actively seeking work outside.

Division h	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 Marks)
- 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? (6 Marks)