EASTERN UNIVERSITY, SRI LANKA

THIRED EXAMINATION IN SCIENCE 2001 / 2002

(APRIL' 2002)

FIRST SEMESTER

ST 302 - SAMPLING THEORY

Answer All Questions

TIME : THREE HOURS

Q1. In a sample survey to study the yield of mango trees, a simple random sample of 10 of 150 villages in a district was selected and the number of mango trees, y, and the area under mango, x, were recorded for each.

Village	Total number of mango trees, (y)	Area (in hectares) under mango trees (x)
1	49	1.2
2	101	1.5
3	71	1.1
4	127	2.8
5	189	3.5
6	78	1.6
7	29	0.5
8	80	1.6
9	78	1.6
10	62	2.3

The total area under mango trees in this district is 88 hectares. Estimate the total number of mango trees in the district using

(i) The simple random sample mean,

(ii) The ratio estimator,

(iii) The regression estimator,

Explain the differences between these estimates.

Q2. Two dentists A and B make a survey of the state of the teeth of 200 children in a village. Dr. A selects a simple random sample of 20 children and counts the number of decayed teeth for each child, with the following results.

Number of decayed Teeth / child	0	1	2	3	4	5	6	7	8	9	10
Number of children	8	4	2	2	1	1	0	0	0	1	1

Dr. B, Using the same dental techniques, examines all 200 children, recording merely those who have no decayed teeth. He finds 60 children with no decayed teeth.

Estimate the total number of decayed teeth in the village children.

(a) Using A's results only.

(b) Using both A's and B's results.

(c) Are the estimates unbiased?

(d) Which estimate do you expect to be more precise?

Q3. (a) If variates y_i and x_i are measured on each unit of a simple random sample of size n, assumed large, show that the variance of $\hat{R} = \frac{\bar{y}}{\bar{x}}$ is approximately

$$\frac{1-f}{n\bar{X}^2} \sum_{i=1}^{N} \frac{(y_i - Rx_i)^2}{N-1}$$

where $R = \frac{\bar{Y}}{\bar{X}}$ is the ratio of the population means and $f = \frac{n}{N}$.

(b) From a list of 468 small two - year colleges a simple random sample of 100 colleges was drawn. The sample contained 54 public and 46 private colleges. Data for number of students *(y)* and number of teachers *(x)* are shown below.

	n	Σy	Σx	Σy^2	Σxy	Σx^2
Public	54	31,281	2,024	29,881,219	1,729,349	111,090
Private	46	13,707	1,075	6,366,785	431,041	33,119

(a) For each type of college in the population estimate the ratio (number of students) / (number of teachers).

(b) Compute the standard errors of your estimates.

(c) For the public colleges, find 90 % confidence limits for the student / teacher ratio in the whole population.

Q4. (a) Distinguish between proportional and optimal allocation in stratified sampling.

(b) An investigator proposes to take a stratified random sample with two strata. He estimates the relevant quantities for the two strata as follows:

Stratum(h)	W_h	S_h	C_h (Rs)
1	0.4	10	4
2	0.6	20	9

where $h =$ Stratum

$W_h = \dfrac{N_h}{N}$ the stratum weight

$S_h =$ Population variance of y, the item of interest
$C_h =$ Cost of sampling unit from stratum h
$n_h =$ Number of units randomly selected from stratum h.

He assumes that his total field cost will be of the form $C_1 n_1 + C_2 n_2$. Find the values of n_1/n and n_2/n that minimize the total field cost for a given value of $V(\bar{y}_{st})$ where n is the total number units selected and $\bar{y}_{st}$ is the usual estimate of the population mean from a stratified sample. Find the total number of units, n, required, under this optimum allocation, to make $V(\bar{y}_{st}) = 1$. Assume that the finite population correction is negligible. How much will the total field cost be?

Q5. In order to estimate the total cattle population in a district consisting of 1238 villages, a simple random sample of 16 villages was selected. The number of cattle recorded in the survey, together with the most recent census figures, are given below:

	Number of cattle	
Village	Survey	Census
1	654	623
2	696	690
3	530	534
4	315	293
5	78	69
6	640	842
7	692	475
8	210	161
Total	3815	3687

	Number of cattle	
Village	Survey	Census
9	292	371
10	555	298
11	2110	2045
12	592	1069
13	707	706
14	1890	1795
15	1123	1406
16	115	118
Total	7384	7808

The census showed that there were 680,900 cattle in the 1238 villages. Estimate the total cattle population from the survey data, using

(i) The ratio estimator,

(ii) The regression estimator,

Also estimate and compare the efficiencies of these estimators relative to an estimator based on the survey information alone.
Select one of these estimators and construct an approximate 95 % confidence interval for the number of cattle in the 1238 villages.

Q6. (a) For stratified random sampling (without replacement), the variance of the estimated proportion, P_{st}, of units in a population possessing a certain attribute is

$$\sum_{h=1}^{L} \frac{W_h^2 P_h (1 - P_h)}{n_h} (1 - \frac{n_h}{N_h})$$

Explain the terms W_h, N_h, P_h, n_h, and L

(b) The cost (in suitable units) of data collection in a stratified sample survey is given by the function

$$C = C_o + \sum_{h=1}^{L} C_h n_h$$

where C_h is the cost per individual observation in stratum h and C_o is the fixed cost of the survey.

(i) Show that the sample size allocation that minimizes $V + \lambda C$, where V is the variance of the estimated proportion (P_{st}) and λ is a positive constant, is given by

$$n_h = W_h \sqrt{\frac{P_h(1-P_h)}{\lambda C_h}}$$

(ii) Show how to choose λ so that the optimal allocation minimize the total cost of sampling for fixed variance V.

(c) A survey is to be conducted to determine the proportion of households living in rented houses in a city. The 2026 households in the city are divided up into four strata.

The following data are given below:

Stratum	Population size	Estimated Proportion renting	Sampling cost per household
1	1190	0.75	9
2	523	0.50	9
3	215	0.20	16
4	98.	0.12	16

For the above data evaluate $n_h = W_h \sqrt{\frac{P_h(1-P_h)}{\lambda C_h}}$ for $h = 1, 2, 3, 4$

in terms of λ