EASTERN UNIVERSITY, SRI LANKA

THIRD EXAMINATION IN SCIENCE 2004/2005

SECOND SEMESTER (Oct./Nov.'2006)

ST 302 - SAMPLING THEORY

Answer all questions

Time: 3 hours

---

1. Considering all possible simple random samples drawn without replacement of size 2 from a population having 6 units with values 8, 3, 1, 11, 4, 7.

    (a) Show that the sample mean is an unbiased estimator of population mean.

    (b) Show also that the sample variance is an unbiased estimator of population variance.

    (c) Calculate the variance of the sample mean and show that it agrees with the formula $(1-f)\dfrac{S^2}{n}$, where S^2 is the population variance, f is the sample fraction and n is the sample size.

2. An industry has 36000 employees. A random sample of 1000 employees were asked to state the number of days they were absent from work in the previous six months. The results were as follows:

Days absent	0	1	2	3	4	5	6	7	8
Number of employees	451	162	187	112	49	21	5	11	2

(a) Estimate the average number of days absent by employees in the industry and find a 95% confidence interval for the average number of days absent.

(b) Find a 95% confidence interval for the proportion of employees absent for more than 3 days.

3. Total income from 19 items sold by a company in January 2001 is Rs. 674,000. Income (in Rs. 1000) from a sample of 6 of the 19 items sold in January 2001 and April 2001 are as follows.

Item	January 2001	April 2001
A	21	26
B	63	91
C	35	47
D	60	70
E	16	17
F	50	76

(a) Estimate the total income for April 2001 using the following

    i. a simple random sample estimator,

    ii. a ratio estimator,

    iii. a regression estimator.

(b) Which of the three methods (i), (ii) or (iii) above, do you consider to be most appropriate in this case? Why?

(c) Estimate and compare the relative efficiencies of your estimators.

4. The 2026 households in a city are divided up into four strata based on income. Simple random samples of households are selected from within strata and the number of households renting the house they live in is found. The results are as follows.

Stratum based on income	Stratum population size	Stratum sample size	Number renting
< 50	1190	40	30
50-100	523	35	18
100-200	215	35	7
200 <	98	40	5

(a) Show that the sample proportion $p_{st} = \dfrac{1}{N} \sum_{h=1}^{L} N_h p_h$ is an unbiased estimator for the proportion of households living in rented houses.

(b) Derive expression for the variance of p_{st}.

(c) Estimate the proportion of households living in rented houses and its standard error.

(d) Is the sample allocation reasonable? If not, suggest a more suitable allocation of a total sample of 150 households.

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

$$C = c_0 + \sum_{h=1}^{L} c_h \sqrt{n_h}$$

where c_h is the cost per individual observation in stratum h, c_0 is the fixed cost of the survey, n_h is the sample size in stratum h and L is the number of strata. It is required to estimate the population mean using a stratified, simple random, sample (without replacement). The variance of this estimate is

$$\sum_{h=1}^{L} W_h^2 \left[1 - \frac{n_h}{N_h} \right] \frac{S_h^2}{n_h}$$

where W_h is the proportion of the total population in stratum h and S_h^2 is the variance in stratum h.

Find the optimum allocation for the sample if this variance is to be minimized for a given total cost.

3

(b) If the amount remaining from the portion of the survey budget allocated to data collection after deduction of the fixed cost is 100 units, determine the total sample size and the optimum strata allocation under the following conditions.

Stratum	W_h	S_h	c_h
1	0.4	4	1
2	0.4	5	2
3	0.2	6	4

6. A pilot survey is taken from a population of 8000 farms. The cultivated area y of each farm (in hectares) and whether or not the farmer keeps animals are recorded. Using standard notation, with p_i equal to the proportion of farms keeping animals for stratum i, $i = 1, 2, 3$, the results are as follows.

Stratum	1	2	3
N_i	1000	3000	4000
n_i	2	6	8
$\bar{y}_i$	8.0	2.6	1.0
s_i	0.2	1.2	0.6
p_i	1.0	0.8	0.33

In the main survey the same strata are to be used. Recommend possible sample sizes to estimate the mean area of farm to within 0.2 hectares and the proportion of farmers keeping animals to within 0.1, allowing only a 5% chance for each that the requirement is not achieved.

4