EASTERN UNIVERSITY, SRI LANKA

THIRD EXAMINATION IN SCIENCE - 2007/2008

SECOND SEMESTER(December/January, 2008/2009)

ST 302 - SAMPLING THEORY

(SPECIAL REPEAT)

Answer all Questions | Time: Three hours

Q1. (a) Define a "Sampling Unit" in terms of various context: Enumeration, Recording, Analysis and Presentation. [20 marks]

(b) Describe six advantages and disadvantages of using Sampling Techniques. [30 marks]

(c) What is meant by "Sampling Errors" and "Non Sampling Errors"? Describe six important circumstances, where "Non Sampling errors occur in a sample survey. [50 marks]

Q2. (a) Prove that in Simple Random Sampling without replacement (SRSWOR) the sample mean is an unbiased estimator of the population mean and the variance of the estimator $\overline{y}$ is given by,

$$\mathrm{Var}(\overline{y}) = \left[1 - \frac{n}{N}\right]\frac{S^2}{n}, \text{ where } S^2 = \frac{1}{N}\sum_{i=1}^{N}\left[Y_i - \overline{Y}\right]^2.$$

[60 marks]

(b) An industry has 36,000 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 off	0	1	2	3	4	5	6	7	8
Number of Employees	451	162	117	112	49	21	5	11	2

(i) Estimate the average number of days "Days off" taken by workman in t industry and 95% confidence interval.

(ii) Find a 95% confidence interval for the proportion of employees absent f more than 3 day.

[40 mar

Q3. (a) Prove that, in Stratified random sampling, the variance of the estimator $\overline{y}_{st}$ given by,

$$\text{Var}(\overline{y}_{st}) = \sum_{h=1}^{L} \left[\frac{1}{n_h} - \frac{1}{N_h}\right] w_h^2 S_h^2,$$

where $w_h = \dfrac{N_h}{N}$ is the proportion of the total population in stratum h, $S_h^2 = \dfrac{1}{N_h - 1} \sum_{i=1}^{N_h} \left[Y_{hi} - \overline{Y}_h\right]^2$ is the variance in stratum h, n_h is the sample si in stratum h, L is the number of strata and assume the samples are take independently from each stratum and in simple random sampling.

If the sampling fraction $f_h = \dfrac{n_h}{N_h}$ are negligible in all strata then show that

$$\text{Var}(\overline{y}_{st}) = \sum_{h=1}^{L} \frac{w_h^2 S_h^2}{n_h}.$$

[50 marks

(b) With two strata, a sampler would like to have $n_1 = n_2$ for administrative con- venience instead of using the values given by *Neyman's* allocation. If $\text{Var}(\overline{y}_{st})$ and $\text{Var}(\overline{y}_{st})_{opt}$ denote the variances given by $n_1 = n_2$ and *Neyman's* allocatio respectively, show that the fractional increase in the variance is,

$$\frac{\text{Var}(\overline{y}_{st}) - \text{Var}(\overline{y}_{st})_{opt}}{\text{Var}(\overline{y}_{st})_{opt}} = \left[\frac{r-1}{r+1}\right]^2,$$

where $r = \dfrac{(n_1)_{opt}}{(n_2)_{opt}}$ as given by *Neyman's* allocation and ignore the sampling fraction.

[50 marks]

Q4. (a) Define a "Linear Systematic Sample" and show that its sample mean is an unbiased estimator of the population mean. Show also that the variance of the estimated mean $Var(\overline{y}_{sys})$ is given by,

$$\text{Var}(\overline{y}_{sys}) = \left[\frac{N-1}{N}\right] S^2 - \left[\frac{(n-1)k}{N}\right] S^2_{wsy},$$

where $S^2_{wsy} = \frac{1}{k(n-1)} \sum_{r=1}^{k} \sum_{i=1}^{n} \left[Y_{ri} - \overline{y}_r\right]^2$ is the sum of squares among units which lie within the same systematic sample, $S^2 = \frac{1}{(N-1)} \sum_{r=1}^{k} \sum_{i=1}^{n} \left[Y_{ri} - \overline{Y}\right]^2$ and $\overline{Y}$ is the population mean.

[40 marks]

(b) The data in following table are small artificial population which exhibits a fairly steady rising trend. Each column represents a systematic sample and the rows are the strata. Compare the precision of systematic sampling, random sampling and stratified sampling.

Data for 10 systematic samples with $n = 4$, $k = 10$, $N = nk = 40$.

Strata	1	2	3	4	5	6	7	8	9	10	Total
I	0	1	1	2	5	4	7	7	8	6	41
II	6	8	9	10	13	12	15	16	16	17	122
III	18	19	20	20	24	23	25	28	29	27	233
IV	26	30	31	31	33	32	35	37	38	38	331

[60 marks]

Q5. If there are two strata and ϕ is the ratio of actual $\frac{n_1}{n_2}$ to the *Neyman* optimum allocation $\frac{n_1}{n_2}$, show that whatever be the values of N_1, N_2, S_1, and S_2, the ratio $\frac{Var(\overline{y}_{st})_{min}}{Var(\overline{y}_{st})}$ is never less than $4\phi(1+\phi)^{-2}$.

(where finite population correction factors are negligible)

[100 marks]

Q6. (a) In stratified random sampling , Let the cost function be ,

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

where c_h is the cost per individual observation in stratum h and c_0 the fix cost of survey. Show that The variance of the estimated mean

$$\text{Var}(\bar{y}_{st}) = \left[\frac{\sum_{h=1}^{L} w_h^2 S_h^2}{n_h} - \frac{\sum_{h=1}^{L} w_h^2 S_h^2}{N_h} \right]$$

is minimum when n_h is proportional to $\frac{N_h S_h}{\sqrt{c_h}}$.

[40 marks

(b) A survey has been planned to obtain information about the age of universit academic staff at a well known university. The information is collected by in terview and the cost incurred varies according to the status of the person bein interviewed. The following Table gives the relevant. information.

	Number	Standard Deviation of Age	Cost per inter
Professors	80	5 years	Rs 400
Assoc. Professors	60	4 years	Rs 256
Senior Lecturers	320	2 years	Rs 100
Lecturers	200	1 year	Rs 100

A total cost, C of Rs 25,000 has been assigned for the above survey while the capital cost, c_0= Rs 5000.

Find the number of persons in each academic category that should be called for the interview.

[60 marks]