EASTERN UNIVERSITY OF SRI LANKA
DEPARTMENT OF MATHEMATICS
THIRD EXAMINATION IN SCIENCES -2008/2009
SECOND SEMESTER (Sep. / Oct., 2010)

ST 302 – SAMPLING THEORY

Answer **All** Questions. Time: **Three** Hours

1. a) i) What are the advantages of the sampling method?

   ii) Define the "target population" and the "study population": (20 Marks)

   b) A simple random sample of size n is selected without replacement from a population of size N and then the population mean, $\overline{X}=\sum_{i=1}^{N} X_i \Big/ N$ and the population variance, $S^2=\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2 \Big/ (N-1)$, are estimated by the sample mean, $\bar{x}=\sum_{i=1}^{n} x_i \Big/ n$ and the sample variance, $s^2=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 \Big/ (n-1)$

   Show the following:

   i) $\bar{x}$ is an unbiased estimator for $\overline{X}$;

   ii) $V(\bar{x})=\left(\frac{N-n}{N}\right)\frac{S^2}{n}$;

   iii) s^2 is an unbiased estimator for S^2. (80 Marks)

2. a) Show that $n \geq \frac{N}{1+NV}$, where N is Population size and $V=\left(\frac{d}{Z_\alpha S}\right)^2$; d is the margin of error, S is population standard deviation and α is the level of significance.

   (20 Marks)

   b) In a particular sector of industry a survey is conducted in an attempt to investigate the extent of absenteeism not connected with illness or official holidays. A random sample of 500 men out of a total workforce of 36000 are asked how many days they have taken off from work, in the previous six months as 'casual holidays' and the results were as follows.

Days off	1	2	3	4	5	6	7	8	9	10
No. of men	157	192	90	31	18	5	2	4	0	1

The objective is to estimate the average number of casual holidays taken by workmen in the industry.

i. Find an estimate for the population variance.

ii. What is the margin of error, if the sample size to estimate population mean is 500? (Consider level of significance as 5%)

iii. How large a sample is needed to estimate the average number of casual holidays taken to within 10% of the correct figure with 95% assurance?

(50 Marks)

c) In a private library the books are kept on 150 shelves of similar size. The numbers of books on 15 shelves picked at random were found to be 28, 23, 25, 33, 18, 22, 29, 29, 30, 22, 26, 20, 21, 28, and 25. Compute confidence interval for total number of books in the library at 5% significance level.

(30 Marks)

3. a) The values of two variables X and Y are observed on a simple random sample of size n from a population of size N. Let $\bar{x}$ and $\bar{y}$ be the sample means and $\bar{X}$ and $\bar{Y}$ be the population means respectively. For sufficiency large n, prove each of the following:

i) Ratio of the sample means r is an unbiased estimator for that ratio of the population means R. Where $r = \frac{\bar{y}}{\bar{x}}$ *and* $R = \frac{\bar{Y}}{\bar{X}}$

ii) $V(r) = \frac{(1-f)}{n\bar{Y}^2} \sum_{i=1}^{N} \frac{(Y_i - RX_i)^2}{N-1}$

(40 Marks)

b) The following figure gives the information on weekly expenditure of food(y), the number of persons (x_1) and weekly family income (x_2) in a simple random sample of 33 middle income families.

$\sum y = 907.2$ $\qquad \sum x_1 = 123$ $\qquad \sum x_2 = 2394$

$\sum y^2 = 28{,}224$ $\qquad \sum x_1^2 = 533$ $\qquad \sum x_2^2 = 177{,}254$

$\sum yx_1 = 3595.5$ $\qquad \sum yx_2 = 66{,}678$

Neglecting sample fraction, estimate each of the following:

i) The mean weekly expenditure on food per family;

ii) The mean weekly expenditure on food per person;

iii) The percentage of the income of that spent on food.

In each case, compute the standard error of the estimator.

(60 Marks)

4. a) Show that:

$$V(\bar{x}_{st})_{Prop} \geq V(\bar{x}_{st})_{Neymann} \geq V(\bar{x})_{SRS},$$

where $V(\bar{x}_{st})_{Neymann} = \frac{1}{n}\left(\sum_{i=1}^{k} W_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} W_i S_i^2$

$$V(\bar{x}_{st})_{Prop} = \frac{(1-f)}{n}\sum_{i=1}^{k} W_i S_i^2$$

$$V(\bar{x})_{srs} = \frac{(1-f)}{n} S^2 \quad and \quad W_i = \frac{N_i}{N}$$

(30 Marks)

b) A stratified population has five strata. The stratum sizes N_i, and means $\bar{X}_i$ and variance S_i^2 of some variable X are as follows:

Stratum	N_i	$\bar{X}_i$	S_i^2
1	117	7.3	1.31
2	98	6.9	2.03
3	74	11.2	1.13
4	41	9.1	1.96
5	45	9.6	1.74

i) Calculate overall population mean and variance.

ii) For a stratified random sample size 80, determine the appropriate stratum sizes under proportional allocation and neymann allocation.

iii) Compare the precisions of these methods with that of simple random sampling.

(70 Marks)

5.a) If the total cost is given by $C = C_0 + \sum_{i=1}^{k} n_i c_i$ where Co is the overhead cost, C_i is the within stratum cost and $W_i = \frac{N_i}{N}$, show that $n_i = \frac{W_i S_i / \sqrt{C_i}}{\sum_{i=1}^{k}\left(W_i S_i / \sqrt{C_i}\right)} n$ when the variance and the cost is minimum.

(30 Marks)

b) A group of students, as part of their group project, has decided to take a stratified sample to estimate the proportion of families with TV sets in an area with three towns. Rough estimates of the total number of families, the proportion with TV sets, and the cost of surveying one family are given in the following table. Treating the three towns as strata, obtain the total optimum sample size and its allocation to the three towns, if the total cost (excluding overheads) is fixed as Rs.4000/=.

Town	Number of families(N_i)	Rough proportion having TV(Pi)	Cost per family (Ci)(in Rs.)
1	14000	0.10	22.00
2	30000	0.25	10.00
3	16000	0.30	25.00

(70 Marks)

6. a) Define a "Circular Systematic Sampling" and show that its sample mean is an unbiased estimator of the population mean.

(30 Marks)

b) Verify that the given statement for the populations size N=11 and the number of systematic sampling n=3.

" Circular systematic sampling mean is an unbiased estimator for population mean"

(20 Marks)

c) Show that the variance of the estimated mean $V(\bar{x}_{sy})$ is given by:

$$V(\bar{x}_{sy}) = \left(\frac{N-1}{N}\right)S^2 - \left[\frac{(n-1)k}{N}\right]S^2_{wsy}$$

Where, $S^2_{wsy} = \frac{1}{k(n-1)}\sum_{r=1}^{k}\sum_{i=1}^{n}(X_{ri} - \bar{x}_r)^2$ is the sum of squares among units which lie within the same systematic sample, $S^2 = \frac{1}{(N-1)}\sum_{r=1}^{k}\sum_{i=1}^{n}(X_{ri} - \bar{X})^2$ and $\bar{X}$ is the population mean?

(20 Marks)

d) Show that the mean of a systematic sample $(\bar{x}_{sy})$ is more precise than the mean of simple random sample $(\bar{x})$ if and only if $S^2_{wsy} > S^2$

(30 Marks)