INDIAN STATISTICAL INSTITUTE

CHENNAI CENTRE

M.STAT I. 2014-15 Semester I

Sample Surveys Final Examination

Total Points 60.

Duration: 3 hours

8 September 2014

*Note: Question no. 1 is **COMPULSORY**. The rest of the paper carries 54 marks and you can answer as much as you can from this part. The maximum you can score from this part is 44. Marks are given in brackets ( ).*

1. The no. of students (x) in a computer course of 5 colleges in a locality of Chennai and the expenditure on peripherals (y) are shown below:

college sl.no.	1	2	3	4	5
size x	80	100	70	50	40
expenditure y(in'000 Rs.)	?	15.7	?	7.5	?

Suppose that a sampler has selected colleges with serial numbers 2 and 4 by Probability Proportional to Size and Without Replacement (PPSWOR) Sampling Scheme.

a) Calculate the sampler's estimate of the total Y for the 5 colleges using Horvitz and Thompson (HT) method. Write down the estimate of variance due to Sen-Yates-Grundy and verify whether it is non-negative in this case.

b) If, by mistake, the sampler had selected these units by PPSWR technique, calculate the HT estimate.

c) If, further, it is known that college no. 4 is selected first in the sample followed by college no. 2, calculate an unbiased estimate t of Y and an unbiased estimate $v(t)$ of Var(t).

(6+2+1) + (3) + (3+1) = (16)

(Continued on reverse)

2a) Describe how you would draw a sample of size *two* from a population of size *N* using (i) Lahiri-Midzuno-Sen scheme,(ii) Brewer's scheme and (iii) Durbin's scheme.

b) Verify whether each of these methods results in a πPS(inclusion probability π_i of i-th. unit Proportional to Size)scheme. If not, suggest modifications/conditions.

c) In each case, show that the Sen-Yates-Grundy variance estimator of Horvitz- Thompson estimator is non-negative.

(6) + (6) + (6) = (18)

3a) From a population of Departmental stores in Chennai a large sample of n_1 stores is selected by Simple Random Sampling (SRS) design for which data on number of customers *x* during a month is collected. Next a subsample of size n_2 is selected by SRS at the second phase on which the amounts spent by the customers using credit cards *y* during the month are noted. Construct a double sampling regression estimator for the *population mean* of *y*.

b) Under certain conditions to be stated by you, derive the variance of the estimator suggested in a) above.

c) If the cost function is given by $n_1 + 10n_2 = 2000$, verify whether *optimum* use of double sampling is always better than a single sample.

(4) + (6) + (8)= (18)

4a) How do you use Warner's Randomized Response Model based on a related question for unbiasedly estimating the population proportion of respondents possessing a sensitive character. Compare this model (w.r.t. the Variance criterion) with the Direct Response Model.

b) Discuss briefly the sources of non-sampling errors in sample surveys and their assessment and control.

(4 +6) + (8) = (18)