INDIAN STATISTICAL INSTITUTE

CHENNAI CENTRE

M.STAT. I YEAR: SECOND SEMESTRAL EXAMINATIONS

Duration: 3 hours SAMPLE SURVEYS Date: 22 February 2016

Note: **Question No. 1 is COMPULSORY. The rest of the paper carries 53 marks and you can answer as much as you can from this part. The maximum you can score from this part is 43.**

1. The size and damage (Rs. hundred) of 6 households(hhs.) in a backward area of a village in Tamil Nadu are given below:

hh. sl. no.	1	2	3	4	5	6
size(x)	8	10	7	5	12	4
damage(y)	-	15.7	-	-	-	7.6

For a quick assessment, a sampling statistician has selected hhs. with serial numbers 2 and 6 by Probability Proportional to Size x and Without Replacement from this population.

(a). Calculate the statistician's estimate of the total Y for the 6 hhs. as suggested by Horvitz and Thompson. Also verify whether the estimate of variance of this estimate due to Sen-Yates-Grundy is non negative.

(b). Further, if it is known that hh. number 6 is selected first in the sample followed by hh. number 2, calculate an unbiased estimate t of Y and of Var. (t) and verify whether this is non negative .

(8+3)+ (2+2+2) = (17)

2(a). For the sampling design corresponding to Lahiri-Midzuno-Sen scheme, verify whether π_i, the inclusion probability of unit U_i is proportional to its size. If not, suggest a modification.

(b). Explain **EITHER** for Brewer's scheme **OR** for Durbin's scheme, how π_i is made proportional to the size of the unit U_i. Also verify whether you can obtain a non negative variance estimator for the chosen scheme.

4+(6+1+3) = (14)

3(a). From a population of class 12 school children, a large sample n' is selected by Simple Random Sampling(SRS) scheme, on which information on the amount spent on coaching classes x is collected. Next, a subsample of size n is selected at the second phase at which the score y in the final examination is noted. Build up a double sampling regression estimator for the population mean of y.

(b). Under suitable assumptions, to be stated by you, derive the variance of the estimator suggested in (a) above.

(c). Further, assuming a suitable cost function, obtain a condition under which optimum use of double sampling is better than a single phase.

(3+7+9) = (19)

4. *In each of the applications of sampling techniques, provide briefly the necessary mathematical justification:*

(a). Areas of N square sheets are available as A_i, $i = 1,2,...,N$.. It is desired to draw a sample of n sheets from N (assumed to be very large) with probability of selection proportional to the side (length) of the sheet. Suggest a simple method.

(b). Mahalanobis's IPNS technique takes into account not only non sampling errors, but also sampling errors. Explain how.

(c). Neyman's optimum allocation of sample size to strata involves population parameters. How do you use this allocation in practice? If in a study of 4 strata, the calculated allocation for a sample of size 12 turns out to be 3.6, 1.3, 5.1 and 2.0, how many units do you select in each stratum?

(d). If from the sampling design: s_1: (U_1, U_2, U_3, U_4), $p(s_1) = 0.3$; s_2: (U_2, U_3, U_5), $p(s_2) = 0.3$; s_3: (U_1, U_2, U_5), $p(s_3) = 0.2$; s_4: (U_2, U_3), $p(s_4) = 0.1$; s_5 = (U_4, U_5), $p(s_5) = 0.1$; a field statistician has selected s_2 : (U_2, U_3, U_5), what is his one-by-one drawing mechanism which resulted in s_2 ?

(4 X 5) = (20)