Indian Statistical Institute, Bangalore
B. Math (III)
Second Semester 2013-2014
Mid-Semester Examination : Statistics (V)
Sample Surveys and Design of Experiments.

Date: 05-03-2014 **Maximum Score 60** **Duration: 3 Hours**

1. Consider cluster sampling set up with N clusters. Let y_{ij} be the $y-$value of the j^{th} unit in the i^{th} cluster, $1 \leq j \leq M_i$ and $1 \leq i \leq N$, where M_i is the number of units in the i^{th} cluster, $1 \leq i \leq N$. Define $\overline{M} = \frac{1}{N}\sum_{i=1}^{N} M_i$. Let $\overline{Y}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} y_{ij}$ be the mean of the i^{th} cluster, $1 \leq i \leq N$. Suppose we want to estimate the population mean $\overline{Y} = \frac{1}{N\overline{M}}\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} = \frac{1}{N}\sum_{i=1}^{N}\frac{M_i\overline{Y}_i}{\overline{M}}$. Suggest an estimator for $\overline{Y}$, based on a sample of n clusters drawn using *simple random sampling without replacement (SRSWOR).* Is your estimator unbiased? Find and estimate the *mean squared error (MSE)* of your estimator.
[2 + 3 + 3 + 4 = 12]

2. Consider stratified set up with L strata. Let y_{hi} be the $y-$value of the i^{th} unit in the h^{th} stratum, $1 \leq i \leq N_h$ and $1 \leq h \leq L$, where N_h is the number of units in the h^{th} stratum, $1 \leq h \leq L$. Suppose we use *simple random sampling with replacement (SRSWR)* in all L strata. Suggest an unbiased estimator for the population mean $\overline{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{L} W_h\overline{Y}_h$. Here $W_h = \frac{N_h}{N}$, and $\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$, mean for the h^{th} stratum, $1 \leq h \leq L$. Show that your estimator unbiased. Obtain its variance. Obtain an allocation that minimizes the variance subject to the cost constraint $C = c_0 + \sum_{h=1}^{L} c_h n_h$, where c_0 is the fixed cost, n_h is the number of units to be sampled from the h^{th} stratum and c_h is the cost per unit in the h^{th} stratum $1 \leq h \leq L$, and C denotes the total budget. How would you 'interpret' the optimal allocation? What happens when $c_1 = c_2 = \cdots = c_L$?
[2 + 3 + 4 + 5 + 2 + 2 = 18]

3. Consider the following estimator of the *population mean* $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$

$$\widehat{\overline{Y}} = \left| \begin{array}{ll} \overline{y} + c & \text{if the sample contains unit 1} \\ \overline{y} & \text{otherwise,} \end{array} \right.$$

where $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean and $c \geq 0$ is a constant. The value y_1 is known to be an unusually low y-value, this having come to light after the *SRSWOR* sample is drawn. Is $\widehat{\overline{Y}}$ unbiased for $\overline{Y}$? Show that $V(\overline{y}) \geq MSE(\widehat{\overline{Y}})$ if and only if $2\frac{N-n}{n(N-1)}\left\{\overline{Y} - y_1\right\} \geq c \geq 0$.
[4 + 10 = 14]

4. Explain *Lahiri's method* and show that it indeed gives rise to *probability proportional to size (PPS) sampling.*
[4 + 6 = 10]

5. Consider finite population with $N = 90$. Consider sampling design p, say, that first takes a linear systematic sample of size 9 and then takes an additional unit from the remaining units in the population with equal probability. Is p a fixed sample size design? Find π_i and π_{ij}, $1 \leq i \neq j \leq N$, for the design p. Define *Horvitz-Thompson estimator* e_{HT} for the *population total* $Y = \sum_{i=1}^{N} y_i$ based on the design p. Is the variance of the strategy (p, e_{HT}) estimable? Substantiate.
[4 + 5 + 2 + 2 + 3 = 16]