Indian Statistical Institute, Bangalore
B. Math (III)
Second Semester 2008-2009
Backpaper Examination : Statistics (V)
Sample Surveys and Design of Experiments.

Date: 21-07-2009 **Maximum Score 60** **Duration: 3 Hours**

1. **Explain** how you would estimate the population total $Y = \sum_{i=1}^{N} y_i$ using *probability proportional to size sampling with replacement* $(PPSWR)$. The size of the ith unit is given by $x_i > 0$, $1 \leq i \leq N$. **Prove** that your estimator is unbiased and **obtain** its variance. How would you **estimate** the variance of your estimator?

   $[2 + 3 + 2 + 3 = 10]$

2. For estimating the population mean $\overline{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}$ **obtain** optimal allocation in the stratified sampling set up that employs *simple random sampling without replacement (SRSWOR)* within each stratum, where y_{hi} is the $y-$value of the ith unit in the hth stratum, $1 \leq i \leq N_h$, $1 \leq h \leq L$ and $N = \sum_{h=1}^{L} N_h$. Take $B = c_0 + \sum_{h=1}^{L} c_h n_h$, as the cost function, where B is the given budget, c_0 is the overhead cost, n_h is the number of units to be sampled from the hth stratum, c_h is the cost per unit of of sampling in the hth stratum, $1 \leq h \leq L$. **Interpret** the result. What happens in the special case when $c_1 = c_2 = \cdots = c_L = c$ (say)?

   [14]

3. **Estimate** the population mean $\overline{Y} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^{N} M_i}$ based on $SRSWR$ sample of n clusters, where y_{ij} is the $y-$value of the jth unit in the ith cluster, N is the number of clusters and M_i is the number of units in the ith cluster, $1 \leq j \leq M_i$, $1 \leq i \leq N$. Is your estimator unbiased? **Obtain** and **estimate** its variance.

   $[2 + 2 + 6 = 10]$

4. How would you estimate the unknown proportion π_A of a sensitive attribute A using Warner's *model*? Is your estimator unbiased? **Obtain** its variance.

   $[4 + 2 + 3 = 9]$

[PTO]

5. A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering suggests that the tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentration of practical interest is between 5% and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5, 10, 15 and 20%. They decide to make up six test specimens at each concentration level, using a pilot plant. All 24 specimens are tested on a laboratory tensile tester, in a random order. The data from this experiment are shown in the following table.

Hardwood	Observations					
Concentration(%)	1	2	3	4	5	6
05	07	08	15	11	09	10
10	12	17	13	18	19	15
15	14	18	19	17	16	18
20	19	25	22	23	18	20

Carry out *ANOVA* to **test** the null hypothesis that different hardwood concentration levels do not affect the mean tensile strength of the paper.

[20]

6. Consider the following statistical model for factorial design with two noninteracting factors and n replicates.

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk} \; ; \; 1 \leq i \leq a \, , 1 \leq j \leq b \, , 1 \leq k \leq n.$$

Obtain least squares estimators *(LSE)* for the model parameters.

[8]