Indian Statistical Institute
Bangalore Centre
B.Math (Hons.) III Year 2012-2013
Second Semester
Sample Survey and Design of Experiments

Semestral Examination Date : 10.05.13

Answer as many questions as possible. The maximum you can score is 112.
All the notation have their usual meaning. State clearly the results you use.

1. Suppose a sample of size n is to be drawn from N units using SRSWOR scheme. Let Y denote the variable under study and X denote an auxiliary variable.

   (a) Define ratio estimator ($\bar{Y_R}$) in this context and explain when it is useful.

   (b) Obtain the expressions for the following. (i) $V[\bar{y}]$ and (ii) $Cov[\bar{x}, \bar{y}]$.

   (c) Show that $\bar{Y_R}$ is biased for the population mean. Obtain an expression for the bias, correct upto order $(1/n)$, in terms of S_x^2, S_y^2 and S_{xy}.

   (d) Suggest a sampling scheme for which the ratio estimator ($\bar{Y_R}$) is unbiased for the population mean. Justify. [3 + 5 x 2 + 8 + (2+8) = 31]

2. Consider a general sampling scheme with fixed sample size(n), where units are selected without replacement. If π_i denote the probability that the i^{th} unit is in the sample, show that $\sum_{i=1}^{N} \pi_i = n$. [6]

3. (a) Define systematic sampling. Provide an unbiased estimator Y_{sy} of the population mean obtained by using systematic sampling. Find its variance.

   (b) Is it possible to estimate the variance of Y_{sy} ? Justify. Show that the variance of Y_{sy} can be estimated by taking more than one systematic sample. [(2 + 2 + 4) + (2 + 7) = 16]

4. After the decision to take a simple random sample had been made, it was realized that y_1 would be unusually low and y_N would be unusually high. Therefore, the following estimator of $\bar{Y}$ was suggested. Take a constant c. Then, define

$$\begin{aligned} \bar{Y_S} &= \bar{y} + c \quad \text{if the sample contains } y_1 \text{ but not } y_N \\ &= \bar{y} - c \quad \text{if the sample contains } y_N \text{ but not } y_1 \\ &= \bar{y} \quad \text{for all other samples} \end{aligned}$$

   Prove that $\bar{Y_S}$ is unbiased and its variance is $(1 - n/N)\left[\frac{S^2}{n} - \frac{2c}{(N-1)}(y_N - y_1 - nc)\right]$. [8]

5. An experiment was to be carried out to compare the effectiveness of v medicines for a particular illness. Since the condition of a patient may also depend on the hospital, patients were chosen from b different hospitals. Each patient was given a medicine for a fixed period and the response (Y) after this period was noted.

   (a) Write an appropriate linear model.

(b) Assuming that k patients were chosen from the each hospital, derive the reduced normal equations for the effects of the medicines in the form

$$C\hat{\tau} = Q. \tag{1}$$

(c) Show that (i) rank of C is $\leq v-1$, (ii) $E(Q) = C\tau$, (iii) $COV(Q) = \sigma^2 C$ and (iv) a linear function $l'\tau$ is estimable whenever l is in the column space of C.

(d) Suppose $\hat{\tau}$ is a solution of (1). Consider the statement "$\sigma^2 C^-$ **acts as** the covariance matrix of $\hat{\tau}$". Is $\sigma^2 C^-$ really the covariance matrix of $\hat{\tau}$? What is the meaning of the phrase **acts as** in the statement above ? [Here A^- is a generalized inverse of A].

[3 + 7 + (3 + 3 + 3 + 3) + 6 = 28]

6. (a) Define a BIBD. If N is the incidence matrix of a BIBD (v, b, r, k, λ) show that $NN' = (r-\lambda)I_v + \lambda J_v$. [Here J_v is the $v \times v$ all-one matrix.]

(b) Suppose an experiment is conducted using a BIBD.

(i) Show that the C-matrix is of the form $a(I - J/v)$. Find a in terms of the parameters of the BIBD.

(ii) Consider an l satisfying $l'1_v = 0$. [Here 1_v is a $v \times 1$ vector of all ones]. Show that $l'\tau$ is estimable and the BLUE of $l'\tau$ is $p.l'Q$. Express p in terms of the parameters of the BIBD.

(c) Consider an Abelian group (V,+) with v elements. Consider an array $A = ((a_{ij}))$, $1 \leq j \leq k$, $1 \leq i \leq h$ with elements from V. Suppose A satisfies the property that every non-zero element of V appear λ times in δ, where δ is a multiset defined as follows. Let δ_i be the multiset : $\delta_i = \{a_{ij} - a_{il}, j \neq l,\ j,l = 1, \cdots k\}$. and $\delta = \bigsqcup_{i=1}^{h} \delta_i$. [Here $\sqcup$ denotes union counting multiplicities].

Then, show that $\exists$ a BIBD (v, b=hv,k, r=hk,λ).

(d) Construct a BIBD with $v = b = 7, r = k = 3, \lambda = 1$.

[(3 + 3) + (2 + (2 + 2)) + 5 + 4 = 21]

7. In a paper manufacturing factory, the percentage of hardwood concentration in raw pulp and the cooking time of pulp are being investigated for their effects on the strength of the paper.

(a) Suppose two concentrations (C) and two cooking times (T) are used and two observations are taken for each level combination. Explain what is meant by (i)the main effects of C and T and (ii) the interaction effect CT and how one can determine them.

(b) Construct a 2^4 experiment on blocks of size 4, so that no main effect is confounded. Find out the effects that are confounded in your design with justification.

[(3 + 4) + 6 = 13]