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

Date: 08-05-2009 Maximum Score 80 Duration: 3 Hours

1. Consider the following linear model

$$y = X\beta + \varepsilon$$

where y, $n \times 1$, is the vector of observations; β, $p \times 1$, is the vector of parameters; X, $n \times p$, is the design matrix and ε, $n \times 1$, is the vector of error terms. Further $\varepsilon \backsim N(0, \sigma^2 I_n)$. Let $rank\,(X) = r$ and $(X'X)^-$ be a fixed g-inverse of $X'X$. Let $\widehat{\beta} = (X'X)^- X'y$ be the least squares estimator *(LSE)* for β. Now consider matrix L and vector z such that the row space of L is contained in the row space of X and the system of equations $Lv = z$ is consistent.

Show that *LSE* for β subject to the constraint $L\beta = z$ is given by

$$\widetilde{\beta} = \widehat{\beta} - (X'X)^- L' \left(L\,(X'X)^- L' \right)^- \left(L\widehat{\beta} - z \right)$$

Derive an F-test for testing the following hypotheses

$$H_0 : L\beta = z \ vs \ H_1 : L\beta \neq z.$$

[20]

2. Consider the set up for completely randomized design *(CRD)* with a treatments and n replicates.

(a) **Explain** the following statistical model **stating** clearly the assumptions that you make.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \ ; \ 1 \leq i \leq a \ , 1 \leq j \leq n.$$

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

(c) **Write down** the analysis of variance *(ANOVA)* table for the following testing problem.

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \ \ vs \ H_1 : \tau_i \neq 0 \ for \ at \ least \ one \ i, \ 1 \leq i \leq a. \qquad (1)$$

(d) **Prove** that sum of squares due to treatments *(SSTreat)* and sum of squares due to errors *(SSError)* are independent.

(e) **Explain** the notion of a contrast.

(f) Consider the contrast defined by $\Gamma = \sum_{i=1}^{a} c_i \tau_i \ ; \ \sum_{i=1}^{a} c_i = 0$. For testing the hypotheses $H_0 : \Gamma = 0 \ \ vs \ \ H_1 : \Gamma \neq 0$, **derive** the distribution of the *test statistic*, under the null hypothesis and **spell out** the test procedure.

$[2 + 4 + 3 + 6 + 2 + (6 + 2) = 25]$

3. If the assumption of normality of errors, ε_{ij}, is not justified in Que 2(a), then **obtain** a nonparametric test for the testing problem given in (1).

[12]

4. An experiment was performed to determine the effect of four different chemicals on the strength of a fabric. These chemicals are used as a part of the permanent press finishing process. Five fabric samples were selected and a randomized complete block design *(RCBD)* was run by testing each chemical type once in a random order on each fabric sample. The data are shown below.

	Fabric Sample (Block)				
Chemical Type↓	1	2	3	4	5
1	1.3	1.6	0.5	1.2	1.1
2	2.2	2.4	0.4	2.0	1.8
3	1.8	1.7	0.6	1.5	1.3
4	3.9	4.4	2.0	4.1	3.4

Carry out *ANOVA* and test whether different chemical types produce the same effect. Also test whether the fabric samples differ significantly.

[20]

5. State an important use of 2^k factorial designs. For $k = 2$, let A and B be two factors and let there be n replications. Define main and interaction effects. For this set up write down and explain the regression model. If β_1 is the parameter in the regression model corresponding to factor A then derive an F-test for testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$. Write down the *ANOVA* table.

[20]