Introduction to Linear Models and Regression

(Compulsory Back Paper)

Full Marks: 50 Time : 3 hrs

Group A is mandatory. From Group B, answer as many as you want but maximum you can score in this group is 30

[Bold-faced letters are used to denote vectors]

Group A

1. Following are the yields (in kg) per plot for a particular crop. There are three varieties of seeds for the crop.

Variety 1	Variety 2	Variety 3
77	109	46
70	106	70
63	137	71
84	79	65
95	134	61
81	78	40
88	126	47
101	98	73

Analyze the data and give your comment(s). Consider $F_{0.05, 2, 21} = 3.465$, $F_{0.05, 3, 21} = 3.06$, $F_{0.05,2,22} = 3.44$. $F_{0.05, 3, 22} = 3.05$, $t_{0.025, 21} = 2.080$, $t_{0.05, 21} = 1.721$. (12)

2. Consider a dataset with the variables temperature (X_1), corn yield (X_2), rainfall (X_3). The corresponding simple correlation coefficients are $r_{12} = 0.59$, r13 = 0.86, r23 = 0.77.
   i) Do you think, a linear model will fit satisfactorily for regressing corn yield on temperature and rainfall?
   ii) Compute the correlation between temperature and corn yield eliminating the effect of rainfall. Give your comments. 4+4=8

Group B

3. Let **X** be a random p-component vector with mean vector **μ** and dispersion matrix Σ. Show that for any λ >0,

$$P\left[(\boldsymbol{X}-\boldsymbol{\mu})'^{\Sigma^{-1}}(\boldsymbol{X}-\boldsymbol{\mu}) > \lambda\right] < \frac{p}{\lambda} \qquad 10$$

4. Under the usual notations of general linear model **Y** = X**β** + **ε**, show that a linear function of the observations belongs to the error space iff its coefficient vector is orthogonal to the columns of X. Hence or otherwise, prove that, the covariance between any linear function belonging to the error space and any BLUE of **β** is zero. (5+5 = 10)
5. While developing a linear model **Y** = X**β** + **ε**, it is observed that some off-diagonal elements of the dispersion matrix of **ε** are non-zero. Describe how will you estimate **β.** (10)
6. Consider the following data.

Age (xi)	No. of patients with high SBP (systolic blood pressure)	Total number of patients
30	7	32
35	4	16
40	10	25
45	14	32
50	20	48
55	28	60

Fit logistic regression model to the data. (10)