8. (a) Develop a suitable test for testing for the significance of regression in multiple linear regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$. Further, show that the test statistics is equivalent to

$$F_0 = \frac{R^2(n-k-1)}{(1-R^2)k}$$

(b) Show that, for any linear model $Y = X\beta + \varepsilon$,

$$\sum_{i=1}^{n} V(\hat{Y}_i)/n = trace\frac{\{X(X'X)^{-1}X'\}\sigma^2}{n} = p\sigma^2/n.$$

Suppose that this model contains a β_0 term in the first position, and $\underset{\sim}{1}$ is an $n \times 1$ vector of ones. Show that $(X'X)^{-1}X'\underset{\sim}{1} = (1\ 0\ \ldots\ 0\ 0)'$ and that $\underset{\sim}{1}'X(X'X)^{-1}X'\underset{\sim}{1} = n$. (6½,6)

[This question paper contains 6 printed pages.]

Your Roll No..............

Sr. No. of Question Paper : 4691 **E**

Unique Paper Code : 32371402

Name of the Paper : Linear Models

Name of the Course : **B.Sc. (H) STATISTICS under CBCS (LOCF)**

Semester : IV

Duration : 3 Hours Maximum Marks : 75

College Library

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.

2. Attempt **six** questions in all, selecting **three** from each Section.

3. Use of a simple scientific calculator is allowed.

SECTION A

1. For a given model $\underset{\sim}{Y}_{n\times1} = X_{n\times p}\underset{\sim}{\beta}_{p\times1} + \underset{\sim}{\varepsilon}_{n\times1}$ with $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}, V(\underset{\sim}{\varepsilon}) = \sigma^2 I$ and $\rho(X) = p < n$, prove that the least squares estimator of $\underset{\sim}{\beta}$ is BLUE. Also, obtain an unbiased estimator of σ^2. (12½)

P.T.O.

2. (a) If $\underset{\sim}{Y}' = (y_1, y_2, \ldots y_n)$ be a vector of n independent standard normal variates. Let $Q_1 = \underset{\sim}{Y}' A_1 \underset{\sim}{Y}$ and $Q_2 = \underset{\sim}{Y}' A_2 \underset{\sim}{Y}$ be distributed as χ^2 with n_1 and n_2 degrees of freedom respectively. Show that the necessary and sufficient condition for Q_1 and Q_2 to be independently distributed is $A_1 A_2 = 0$.

(b) Let x, y, and z denote three independent standard normal variables. Stating appropriate theorem, find the distribution of homogenous equation of second degree, $0.71x^2 + 0.86xy + 0.36y^2 + 0.93z^2 + 0.42yz - 0.28xz$. (8½,4)

3. Consider a study examining the relationship between exercise and blood pressure in middle-aged men. In this study, the participants are classified into three groups based on their exercise habits, that are, those who exercise 6 days a week, those who exercise 3 days a week, and those who do not exercise at all. The age of each participant is also recorded along with their blood pressure. Present the complete statistical analysis to analyze the relationship between exercise and blood pressure, while controlling for the effects of age. (12½)

A: Take one observation at each of $(X_1, X_2) = (-1, -1)$ and $(1, 1)$ and take three observations at each of $(-1, 1)$ and $(1, -1)$.

B: Take two observations at each of the four sites.

If a model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ is to be fitted by the least squares but it is feared there may be some additional quadratic curvature expressed by the extra $\beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$, evaluate the anticipated biases in the estimated coefficients β_0, β_1, β_2 for each design.

(b) Discuss the problem of testing for lack of fit in the simple linear regression model. (6½,6)

7. Write notes on **any two** of the following :

(a) Orthogonal Columns of the design matrix

(b) Stepwise Regression

(c) Orthogonal Polynomials (6½,6)

4. (a) Suppose X_i, Y_i, Z_i, $i = 1, 2, 3$ are 9 independent observations with common variance σ^2 and expectations $E(X_i) = \theta_1$, $E(Y_i) = \theta_2$, $E(Z_i) = \theta_1 - \theta_2$, $i = 1, 2, 3$. Find (i) the BLUEs of θ_1, θ_2, (ii) $\operatorname{cov}\left(\hat{\theta}_1, \hat{\theta}_2\right)$, and (iii) the BLUE of $\theta_1 + \theta_2$.

(b) Fill in the missing entries of the partially completed two-way ANOVA table :

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	Variance Ratio
Factor A		231.5054	115.7527	
Factor B		98.5	14.0714	
Error		573.75		
Total	23			

(i) How many classes/levels of factor A are being compared?

(ii) How many observations are analyzed?

(iii) At 0.05 level of significance, can one conclude that the classes of factor A have different effects? Why? (Use $F_{0.05[2,23]} = 3.42$, $F_{0.05[7,14]} = 2.76$, $F_{0.05[2,14]} = 3.74$, $F_{0.05[2,7]} = 4.74$) (6½,6)

SECTION B

5. (a) Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ given that $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$, ε's are uncorrelated.

   (i) Obtain the least squares estimates of β_0 and β_1.

   (ii) Verify the bias and variance properties of $\hat{\beta}_0$ and $\hat{\beta}_1$.

   (iii) Show that $\text{cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\sigma^2 \tilde{x}}{S_{xx}}$.

   (b) Develop for the simple linear regression model with intercept a confidence interval on the mean response at x_0 and a prediction interval for the future observation y_0 corresponding to a specified level x_0 of the regressor variable x. Which of the two intervals is wider? (6½,6)

6. (a) Discuss the problem of bias in regression estimates? Eight experiments are to be done at the coded levels (±1, ±1) of two predictor variables X_1 and X_2. Two experimenters A and B suggest the following designs :