INDIAN STATISTICAL INSTITUTE, BANGALORE CENTRE
B.MATH - Second Year, 2023-24
Statistics - III, Semesteral Examination, May 2, 2024
Time: 3 Hours **Total Marks: 50**

You may use any of the results stated and discussed in class, by stating them explicitly.

1. For $n \geq 4$ let $Z_i, 1 \leq i \leq n$ be independent $N(0, \sigma^2)$ random variables. Consider $0 < \alpha < 1$. Define $X_1 = Z_1$, $X_2 = \alpha Z_1 + \sqrt{1-\alpha^2} Z_2$, $X_3 = \alpha Z_1 + \sqrt{1-\alpha^2} Z_3$ and $X_4 = Z_4$. Let $\mathbf{X} = (X_1, \ldots, X_4)'$.
(a) Find the probability distribution of $\mathbf{X}$.
(b) Find the partial correlation coefficients $\rho_{12.3}$ and $\rho_{12.34}$ (between elements of $\mathbf{X}$). [4+8]

2. Consider the model $\mathbf{Y} = X\beta + \epsilon$, where $X_{n\times p}$ has rank $r \leq p$, and $\epsilon \sim N_n(\mathbf{0}, \sigma^2 I_n)$. Let $(X'X)^-$ denote a generalized inverse of $X'X$. Also let $\hat{\beta}$ be a least squares solution of β. Suppose $A\beta$ is estimable where $A_{q\times p}$ has rank q.
(a) Show that $A(X'X)^- X'X (X'X)^- A' = A(X'X)^- A'$;
(b) Find the probability distribution of $A\hat{\beta}$. [4+5]

3. Consider the model $\mathbf{Y} = X\beta + \epsilon$, where $X_{n\times p}$ has rank p. Suppose $\epsilon \sim N_n(\mathbf{0}, \sigma^2 D)$, where D is a diagonal matrix with positive diagonal entries. Let $\hat{\beta}_{ls}$ denote the ordinary least squares solution which minimizes $(\mathbf{Y} - X\beta)'(\mathbf{Y} - X\beta)$, and $\hat{\beta}_{mle}$ denote the maximum likelihood estimate of β.
(a) Find $\hat{\beta}_{ls}$ and $\hat{\beta}_{mle}$. Are they different?
(b) When is a linear parametric function $\mathbf{a}'\beta$ estimable under the above model?
(c) What is the BLUE of such an estimable $\mathbf{a}'\beta$? [8+2+5]

4. The response time in milliseconds was determined for three different types of circuits that could be used in an automatic valve shutoff mechanism. The results were the following:

Circuit	Response time (ms)				
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

(a) Describe the methodology for determining whether the response times for the different circuit types significantly differ. Numerical computations are not needed.
(b) What is meant by a linear contrast in an experiment like this?
(c) What is the relation between the ANOVA null hypothesis and the hypotheses to check various linear contrasts? [6+1+2]

5. Explain the similarities and differences between Least squares, Ridge and LASSO regression [5]