Indian Statistical Institute
Bangalore Centre
B.Math (Hons.) III Year 2010-2011
First Semester
Statistics III

Backpaper Examination — Date : 28.12.11

Answer as many questions as possible. The maximum you can score is 100
All symbols have their usual meaning, unless stated otherwise. State clearly the results you use.

1. Consider the linear model

$$Y(n \times 1) = X(n \times p)\ \beta(p \times 1) + \varepsilon(n \times 1).$$

Here $\varepsilon \sim N_n(0, \sigma^2 I_n)$.

(a) Suppose l is in R^p. When is $l'\beta$ said to be estimable? Obtain the condition on l so that $l'\beta$ is estimable.

(b) Define estimation space and obtain it in terms of the column space of X.

(c) Define 'BLUE' of a linear parametric function. Consider a vector a in the estimation space. Show that $a'Y$ is the BLUE of its expected value.

(d) Derive the distribution of SS_E, the error sum of squares.

(e) Suppose $l'\beta$ is estimable.

(i) Show that $l'\hat{\beta}$ is independent of SS_E.

(ii) Show how you can find a 95% confidence interval for $l'\beta$.

(f) Consider the hypothesis $H_0 : H'\beta = 0$

Assuming that $H'\beta$ is estimable and that $\Sigma_H = Cov(H'\hat{\beta})$ is positive definite

(i) derive the distribution of $SS_H = (H'\hat{\beta})'(\Sigma_H)^{-1}H'\hat{\beta}$ under H_0 and

(ii) show that SS_H is independent of SS_E.

[(1+2) + (2 + 3) + (1 + 4) + 5 + (4 + 4) + (8 + 6) = 40]

2. An mill owner wants to study whether the strength of the fibre produced in the mill depends on the percentage of cotton. A linear regression model was to be fitted. Stating the required condition on the data set explain how the lack of adequacy of the linear model can be tested.

[8]

3. Consider a random vector $X = (X_1, \cdots X_p)'$ with covariance matrix Σ.

(a) Obtain the 'best predictor' of X_1 among all **linear** functions of $X_2, \cdots X_p$ and denote it by $X_{1.2\cdots p}$.

(b) What is the 'best predictor' of X_1 among **all** functions of $X_2, \cdots X_p$? Fill in the blank in the following statement with justification.

"When X follow - distribution, 'the best predictor' here coincides with that of Q(a)".

(c) Let $R_{1.2\cdots p} = X_1 - X_{1.2\cdots p}$. Show that $R_{1.2\cdots p}$ is uncorrelated with every $X_j, j = 2, \cdots p$.

(d) Define multiple correlation coefficient between X_1 and $X_2, \cdots X_p$. What does it measure ?

(e) Define partial correlation coefficient ($\rho_{12.3\cdot p}$) between X_1 and X_2, when $X_3, \cdots X_p$ are eliminated. Show that

$$\rho_{12.3\cdot p} = -\sigma^{12}/[\sqrt{(\sigma^{11}.\sigma^{22})}],$$

where σ^{ij} is the (i,j)th entry of Σ^{-1}.

[6 + (2 + 3) + 4 + (2 + 1) + 8 = 26]

4. The effects of **different catalysts** on the **time of production** of a chemical is being studied. The experimenter also suspects that the **raw material from different batches** may not be identical.

Assume that there were v catalysts, b batches of raw material and each batch was used to make k units of the chemical. Further assume that the effects of catalysts as well as raw material are constants.

(a) Write an appropriate linear model.

(b) Derive the reduced normal equations (say $C\hat{\tau} = Q$) for the effects of the catalysts.

(c) Show that $l'\tau$ is estimable if and only if l is in $\mathcal{C}(C)$.

(d) Find $Cov(Q)$.

(e) If $l'\tau$ is estimable, find the variance of $l'\hat{\tau}$.

(f) Consider $SS_{cat}(adj) = \sum_{i=1}^{n} Q_i\hat{\tau}_i$.

Justify its use in the procedure for testing whether different catalyst have different effects.

[Hint : Look at the expected value].

[3 + 6 + 6 + 4 + 3 + 8 = 30]