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

Semestral Examination Date : 7.12.11

Answer as many questions as possible. The maximum you can score is 120. All symbols have their usual meaning, unless stated otherwise. State clearly the results you use.

1. Consider the linear model

$$Y(n \times 1) = X(n \times p)\, \beta(p \times 1) + \varepsilon(n \times 1), \quad \text{where } \varepsilon \sim N_n(0, \sigma^2 I_n).$$

(a) If $l'\beta$ is estimable then show that $l'\hat{\beta}$ is independent of SS_E.

(b) What condition $H(k \times p)$ must satisfy so that the hypothesis $H_0 : H'\beta = 0$ can be tested ? Justify.

(c) Assuming that H satisfies the condition of Q(c), and that $\Sigma_H = Cov(H'\hat{\beta})$ is positive definite show that

$$SS_H = (H'\hat{\beta})'(\Sigma_H)^{-1} H'\hat{\beta} \text{ is independent of } SS_E.$$

Explain how this fact is utilized in finding a test procedure for H_0. [The proofs of the relevant results are not required].

[6 + 3 + (7 + 4) = 20]

2. (a)State Fisher-Cochran theorem.

(b) Suppose $X \sim N_n(0, I_n)$. Show that the quadratic form $Q = X'AX$ follows χ^2 distribution if and only if A is idempotent. Further, the degrees of freedom of Q is the same as rank(A).

(c) Consider three quadratic forms Q, Q_1, Q_2 of X, such that $Q = Q_1 + Q_2$. Further, suppose $Q \sim \chi^2(a)$, $Q_1 \sim \chi^2(b)$ and Q_2 is non-negative definite. Show that $Q_2 \sim \chi^2(a - b)$. [3 + 7 + 7 = 17]

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

(a) Find the 'best predictor' of X_1 among all **linear** functions of $X_2, \cdots X_p$.

(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) Denote 'the best predictor' of X_1 obtained in (a) by $X_{1.2\cdots p}$. Let $R_{1.2\cdots p} = X_1 - X_{1.2\cdots p}$.

(i) Find variance of $X_{1.2\cdots p}$.

(ii) Find the covariance between X_1 and $X_{1.2\cdots p}$.

(iii) Show that $R_{1.2\cdots p}$ is uncorrelated with every $X_j, j = 2, \cdots p$.

(d) Define multiple correlation ($\rho_{1.2\cdots p}$) of X_1 with $X_2, \cdots X_p$. Show that

$$1-\rho^2_{1.2\cdots p} = det\rho_{11}/det\rho_{22}, \text{ where } \rho_{tt} = ((corr(X_i,X_j)))_{t\le i,j\le p},\ t=1,2.$$

[7 + (4 + 3) + (3 + 3 + 5) + 8 = 33]

4. An industrial engineer is investigating the effect of four assembly methods (A,B,C,D) on the assembly time for a television component. Four different machines are used, operated by four operators of varied skills. The engineer uses the following design.

	Operators			
Machines	O_1	O_2	O_3	O_4
1	C	D	A	B
2	A	B	C	D
3	B	A	D	C
4	D	C	B	A

(a) Assuming that the effects of assembly methods, machines as well as operators are constant, write an appropriate linear model.

(b) Obtain the BLUE of the difference between the effects of two assembly methods.

(c) Obtain the sum of squares for testing whether the effects of the assembly methods are significantly different.

(d) In another investigation, the operators were selected from a large number of skilled operators and so the effect of the operators were assumed to be a random variable with mean zero and constant variance. Show how this variance can be estimated.

[3 + 7 + 9 + 10 = 29]

5. Consider the linear model

$$Y(n\times 1) = X(n\times a)\,\alpha(a\times 1) + Z(n\times b)\,\beta(b\times 1) + \varepsilon(n\times 1),$$

where α is a vector of constants, but β is a vector of random variables with $E(\beta) = 0$ and $Cov(\beta) = \sigma^2_b\, I_b$. Assumptions on ε are as usual.

(a) Derive the covariance matrix of Y.

(b) Define error space and estimation space and obtain them.

(c) Derive the system of normal equations.

(d) The effects of **v given catalysts** on the **time of production** of a chemical is being studied. For the raw material, b batches were randomly selected from a large number of batches. Each batch was used to make k units of the chemical.

(i) Write an appropriate linear model.

(ii) Show how you can estimate the variance of experimental error and the variability among the raw materials from different batches.

[3 + (3 + 2 + 4) + 8 + (3 + 12) = 35]

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

Backpaper Examination | Date :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 error space and estimation space and obtain them in terms of the column space of X.

(c) 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) + (3 + 2 + 3) + 4 + 5 + (4 + 4) + (6 + 6) = 40]

2. (a) 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.

(b) To study the dependence of Y on $X_1, \cdots X_P$, a multiple regression model was to be fitted.

Derive the BLUE's of the coefficients and the variance of the BLUE of a regression coefficient.

[8 + 6 = 14]

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 $X_{1.2\cdots p}$

(i) Find variance of $X_{1.2\cdots p}$.

(ii) Find the covariance between X_1 and $X_{1.2\cdots p}$.

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

(c) 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

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

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

[(6 + 3 + 3) + 3 + 7 = 22]

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 is constant.

(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]