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

Semester Examination Date :1.12.10

Answer as many questions as possible. The maximum you can score is 120
The notation have their usual meaning. State clearly the results you use.

1. Consider the linear model

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

where ε follows multivariate normal distribution with mean 0 and covariance matrix $\sigma^2 I_n$.

(a) When is $l'\beta$ said to be estimable? What do you mean by a "least square (l.s.) estimator" of a parametric function ? Does it always exist ? Is it always unique ? Justify your answer.

(b) What is meant by the (unconditional) residual sum of squares (R_0^2) ? Obtain an expression for R_0^2 in terms of only X and Y.

(c) Derive the distribution of R_0^2.

(d) Show that the l.s. estimator of any estimable linear function of β is independent of R_0^2.

(e) Suppose we want to test the hypothesis $H_0 : H'\beta = \xi$. What condition H must satisfy so that it is possible to do that ? Assuming that that condition is satisfied, obtain $Cov(H'\hat{\beta})$. Show that $Cov(H'\hat{\beta})$ is nonsingular.

(f) Let R_1^2 denote the residual sum of squares under H_0 of Q(e). Obtain an expression for $R_1^2 - R_0^2$ in terms of $H, \hat{\beta}, \xi$ and $Cov(H'\hat{\beta})$.

(g) Show that $R_1^2 - R_0^2 = Y'AY$, where A is an idempotent matrix. Hence or otherwise find the distribution of $R_1^2 - R_0^2$. Show that $R_1^2 - R_0^2$ and R_0^2 are independent. Derive a test statistic for testing H_0 against its negation.

[(2+2+2+2) + (2+4) + 6 + 6 + (2+3+6) + 7 + (6+4+4+2) = 60]

2. The effects of five different catalysts (A,B,C,D,E) on the reaction time of a chemical process is being studied. The experimenter also wants to see if there is any variation in the raw material from different batches.

(a) Suppose 4 batches of raw material were used, they were divided into five parts and in each part a different catalyst was used and the reaction time noted.

Write down an appropriate linear model, assuming that the effects of the catalysts as well as the raw materials are constants. Let τ_Q denote the effect of catalyst $Q, Q = A, \cdots E$. Is $\tau_A - \tau_B$ estimable ? If so, find its BLUE.

(b) In another experiment the raw material was enough for only two trials and the following design was used.

Batch number	Catalysts	
1	A	B
2	C	D
3	C	A

See whether the following are estimable. (i) $\tau_A - \tau_B$, (ii) $\tau_A - \tau_D$.

(c) Suppose in Q(a), the batches of raw materials were selected at random from a large number of batches. Write down an appropriate model. Define error space and estimation space and describe them. Is $\tau_A - \tau_C$ estimable ?

Suppose the design of Q2(b) is used. Is $\tau_A - \tau_D$ estimable ?

[(3+2+4) + (2+3) + (3 + 2+2+4 + 2 + 3) = 30]

3. (a) Consider a random vector X following a p-variate normal distribution with mean μ and Covariance matrix Σ. Assuming Σ to be positive definite derive the density function of X.

Suppose X is partitioned as $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. State and prove the condition for independence of X_1 and X_2.

[6 + 4 = 10]

4. Consider a $p \times p$ symmetric matrix S of random variables following central Wishart distribution $W_p(n, \Sigma)$, with $n \geq p$ and Σ positive definite.

(a) Let a^{ij} denote the (i, j)th entry of A^{-1}. Show that $\sigma^{pp}/s^{pp} \sim \chi^2(n-p+1)$ and is independent of each $s_{i,j}, 1 \leq i, j \leq p-1$.

(b) Suppose S is partitioned as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

and Σ is also partitioned accordingly. Then, $S_{22} - S_{21}(S_{11})^{-1}S_{12} \sim W_t(n-r, \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12})$. [Here t and r are the orders of S_{22} and S_{11} respectively; $r + t = p$].

[8 + 8]

5. Consider a random vector $X = \begin{bmatrix} X_1 & \cdots & X_p \end{bmatrix}^T$.

(a) Suppose we want to predict X_1 from observed values of $X_2, \cdots X_p$. What is the 'best predictor' (i) among all functions and (ii) among all linear functions of $X_2, \cdots X_p$? Justify your answer.

(b) Let $R_{1.2\cdots p}$ denote the residual of X_1 after taking away the 'best linear predictor '. Show that $Cov(R_{1.2\cdots p}, X_j) = 0,\ j = 2, \cdots p$.

(c) Define multiple correlation coefficient $\rho_{1.2\cdots p}$ between X_1 and the other variables. Express $\rho_{1.2\cdots p}$ in terms of the correlation matrix (ρ) of X and that (ρ_2) of $\tilde{X_2} = \begin{bmatrix} X_2 & \cdots & X_p \end{bmatrix}^T$.

[(3+3) + 4 + (3+6) = 19]

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

Back paper Examination — Date : 21.12.10

Answer as many questions as possible. The maximum you can score is 100
The notation have their usual meaning. State clearly the results you use.

1. The effects of five different catalysts (A,B,C,D,E) on the reaction time of a chemical process is being studied. The experimenter also wants to see if there is any variation in the raw material from different batches. Suppose 4 batches of raw material were used, they were divided into five parts and in each part a different catalyst was used and the reaction time noted.

   (a) Write down an appropriate linear model, assuming that the effects of the catalysts as well as the raw materials are constants. Let τ_Q denote the effect of catalyst $Q, Q = A, \cdots E$.

   See whether τ_A and $\tau_A - \tau_B$ are estimable. Find the BLUE of the estimable one(s).

   (b) The experimenter wants to see whether (i) different catalysts and/or (ii) different batch of raw material causes difference in the reaction time. Formulate these as testing of hypothesis problems. Is it possible to test these hypotheses ? If so, derive an appropriate test procedure for each of them.

   (c) In another experiment the raw material was enough for only three trials and the following design was used.

Batch number	Catalysts		
1	A	B	C
2	D	E	
3	E	A	C

   See whether the following are estimable. (i) $\tau_A - \tau_B$, (ii) $\tau_D - \tau_C$.

   (d) Suppose the batches of raw materials were selected at random from a larger number of batches. Write down an appropriate linear model. Define error space and estimation space and describe them with justification.

   When is $l'\tau$ estimable ? Show how to find the BLUE of $l'\tau$ if it is estimable. See whether $\tau_D - \tau_C$ is estimable from the design of Q(c) under this model.

   [(2+2+5) + (3+2+10) + (2+3) + (2 + (2+2+5) + (2+6+2)) = 50]

2. (a) Consider a random vector X following a p-variate normal distribution with mean μ and Covariance matrix Σ.

   Suppose rank of Σ is $m < p$. Then show that $X = \mu + BU$, where U is a vector of m independent standard normal variates and B is a $m \times p$ matrix such that $\Sigma = BB^T$.

   (b) Suppose X is partitioned as $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$.

   (i) State and prove the condition for independence of X_1 and X_2.

   (ii) Derive the conditional distribution of X_1, given X_2.

   (c) Suppose $X_1, X_2, \cdots X_n$ are i.i.d random variables following the same distribution as X. Let $\bar{X} = \sum_{i=1}^{n} X_i$ and $S = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$.

   Assuming rank of $\Sigma = p$ show that $\bar{X}$ and S are independently distributed.

   (d) Suppose $\mu = 0$ and $\Sigma = I_n$. Let A be a real symmetric matrix and l is a vector in R^n. Prove that a necessary and sufficient condition that $l'X$ and $X'AX$ are independent is $l'A = 0$.

[6 + (3 + 5) + 6 + 10 = 30]

3. The brake horse power developed by an automobile engine is thought to be a function of the engine speed (in RPM).

   (a) Suppose a few observations are taken and a simple linear regression model is fitted to the data.

   (i) Obtain an expression for the BLUE of the regression coefficient.

   (ii) What condition the data must satisfy so that model adequacy can be checked. Assuming that the condition is satisfies, derive a test procedure for testing model adequacy.

   (b) Later, the experimenter felt that the brake horse power may also depend on the road octane number of the fuel and the engine compression. Some more observations are taken and a linear model is fitted. Show how the goodness of fit can be tested.

[(4+2+6) + 9 = 21]