Indian Statistical Institute
Bangalore Centre
B.Math (Hons.) III Year 2015-2016
First Semester
Statistics III

Semestral Examination Date : 9.11.15

Time : 4 hours

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 regression model $y_i = \alpha + \beta x_i + \epsilon_i,\ 1 \le i \le n$. Here $\epsilon_i, i = 1, 2, \cdots n$ are i.i.d normally distributed variables with mean 0 and variance σ^2. Suppose

$$
\begin{aligned}
S_x^2 &= \sum_{i=1}^{n}(x_i - \bar{x})^2, \\
S_y^2 &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ and} \\
S_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).
\end{aligned}
$$

(a) Show that $E(S_{xy}^2) = \sigma^2 S_x^2 + \beta^2 S_x^4$.

(b) Let $S^2 = S_y^2 - S_{xy}^2/S_x^2$. Obtain $E(S^2)$.

(c) Derive the distribution of S^2 without using any result of least square theory. [6 + 6 + 8 =20]

2. Suppose $X_1, X_2, \cdots X_n$ is a random sample from $N_p(\mu, \Sigma)$. Let

$$
\begin{aligned}
S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})' \\
\text{and } S_0^2 &= \sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)'.
\end{aligned}
$$

(a) Show that $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = S^2$ maximise the likelihood function over the whole parametric space.

(b) Suppose $\mu = \mu_0$. Show that $\hat{\Sigma_0} = S_0^2$ maximises the likelihood function over the restricted parametric space.

(c) Consider the testing of hypothesis problem $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

(i) Show that the likelihood ratio test statistic is a monotone function of the Hotelling's T^2 statistics.

(ii) Derive the distribution of Hotelling's T^2. [12 + 5 + 8 + 8 = 33]

3. While working with a linear model with three parameters $\beta_0, \beta_1, \beta_2$, one came across the system of normal equations $S\beta = Z$, where $Z = (5, -2, -3)'$ and

$$S = \begin{bmatrix} 10 & -5 & -5 \\ -5 & 3 & 2 \\ -5 & 2 & 3 \end{bmatrix}.$$

Find two different g-inverses G and H of S. Suppose $\hat{\beta} = GZ$ and $\tilde{\beta} = HZ$. Compute $\hat{\beta}_1 - \hat{\beta}_2$, $\tilde{\beta}_1 - \tilde{\beta}_2$, $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2$ and $\tilde{\beta}_0 + \tilde{\beta}_1 + \tilde{\beta}_2$. Explain the fact that the first two numbers are same while the last two need not be so. [4 x 2 + 4 + 5 = 17]

4. Consider the linear model

$$Y = \mu 1_n + X_T \tau + X_B \beta + \varepsilon,$$

where μ, τ and β are vectors of unknown constants of appropriate order and ε is a random vector with $E(\varepsilon) = 0$ and $Cov(\varepsilon) = \sigma^2 I_n$.

Let $C = (X_B)'(I - P_T)X_B$ and $R = (X_B)'(I - P_T)Y$, where P_T is the projection operator on $\mathcal{C}(X_T)$.

(a) Show that $E(R) = C\beta$ and $Cov(R) = \sigma^2 C$.

(c) Let $SS_B = R'C^- R$. Derive $E[SS_B]$.

(d) Now suppose β is a random vector with $E(\beta) = 0$ and $Cov(\beta) = \sigma_1^2 I_b$.

(i) Find an unbiased estimator of σ^2.

(ii) Show that the expectation of SS_B is of the form $a_1\sigma^2 + a_2\sigma_1^2$, where a_1 and a_2 are functions of X_T and X_B.

[(3 + 3) + 7 + (6 + 8) = 27]

5. Suppose a $p \times p$ random matrix S follows central Wishart distribution : $S \sim W_p(n, \Sigma)$. Assuming that both S and Σ are positive definite, prove the following results.

(a) $Z = \sigma^{pp}/s^{pp} \sim \chi^2(n - p + 1)$. [Here a^{ij} is the (i, j)-th element of A^{-1}].

(b) Z is independent of $((s_{ij}))_{1 \leq i,j \leq p-1}$.

(c) $det(S)/det(\Sigma)$ is distributed as the product of p independent χ^2 variables with degrees of freedom $n - p + 1, \cdots n - 1, n$.

[12 + 2 + 8 = 22]

6. Consider a multiple regression model with parameters $\beta_0, \beta_1, \cdots \beta_p$.

(a) Obtain the expression of variance of $\hat{\beta}_i$.

(b) What is multicollinearity ? Show how the variance of $\hat{\beta}_1$ is inflated due to the presence of multicollinearity. Define 'variance inflation factor' with justification. [5 + (2 + 7 + 4) = 18]