CSM – 68/20
Statistics
Paper – I

Time : 3 hours

Full Marks : 300

The figures in the right-hand margin indicate marks.

Candidates should attempt Q. No. 1 from Section – A and Q. No. 5 from Section – B which are compulsory and any ***three*** *of the remaining questions, selecting at least* ***one*** *from each Section.*

SECTION – A

1. Attempt any **three** of the following sub-parts : $20 \times 3 = 60$

   (a) Consider an experiment of throwing a die with 6 equally likely faces denoted by 1, 2, 3, 4, 5, 6. If you toss this die independently, until number 3 appears then (i) what is your sample space ? (ii) Check if the number of trial required to get 3 is a RV (iii) Obtain the probability distribution of X.

(Turn over)

(b) State and prove Lindberg-Levy central limit theorem.

(c) Let $Y = X\beta + \varepsilon$ be a multiple linear regression model. State the standard Ordinary Least Squares (OLS) assumptions on this model. Under thse assumptions obtain the Least Squares Estimator (LSE) of β and show that it is unbiased. Obtain the dispersion matrix of LSE.

(d) Let $X_1, X_2, \ldots\ldots, X_{109}$ be a random sample from a tri-variate normal distribution with mean vector $\mu = (0, 0, 0)'$ and dispersion matrix

$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$. If the matrix of sample

covariance is $A = \begin{pmatrix} 5 & 2 & -0.25 \\ 2 & 5 & -0.20 \\ -0.25 & -0.20 & 5 \end{pmatrix}$.

Test whether $(X_1, X_2)'$ is independent of X_3 at 5% level of significance.

2. (a) Let $A_1, A_2, \ldots, A_n$ for $n \geq 2$ be arbitrary events.

Show that $P\left(\bigcap_{j=1}^{n} A_j\right) \geq 1 - \sum_{j=1}^{n} P\left(A_j^c\right)$.

(b) Let (X, Y) be a random vector with bivariate distribution :

$$f(x, y) = \begin{cases} \frac{1}{4}\left[1 + xy\left(x^2 - y^2\right)\right] & \text{if } |x| \le 1, |y| \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Show that X and Y are uncorrelated but not independent.

(c) If $\{X_n\}$ is a sequence of random variables with $E(X_n) \to \mu < \infty$ and $\text{Var}(X_n) \to 0$ as $n \to \infty$ then show that $X_n \to \mu$ in probability.

$20 \times 3 = 60$

3. (a) Obtain the characteristics function of a random variable X with cumulative distribution function

$F(x) = \rho + (1 - \rho)(1 - e^{-\lambda x}), 0 \le x < \infty, 0 \le \rho < 1,$
$\lambda > 0.$

Hence find mean and variance of X.

(b) Show that the sum of squares of n independent and identically distributed standard normal random variables follows a Chi-square distribution with n degrees of freedom.

(c) Let X_1 and X_2 be two independent and identically distributed rvs with common probability density function (pdf)

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-x/2}, & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Obtain the pdf of $\frac{X_1}{X_1 + X_2}$. 20×3 = 60

4. (a) Let $Y_1, Y_2, \ldots\ldots, Y_n$ be a set of rvs from a simple linear regression model : $Y_i = \alpha + \beta X_i + \varepsilon_i$, $i = 1, 2, \ldots\ldots, n$, where ε_i are iid $N(0, \sigma^2)$ rvs and X_i are known values. Construct a $100(1 - \alpha)\%$ confidence interval for the regression parameter β.

(b) What is the role of principal component analysis in statistics ? Show that the first principal component is the eigen vector corresponding to the largest eigen value of the associated dispersion matrix.

(c) The following are the coded values of the amount of corn (in bushels per acre) obtained from three different varieties, say A, B, C using unequal number of experimental plots for different varieties :

Variety A : 40, 30, 50, 40, 30

Variety B : 60, 40, 55, 65

Variety C : 60, 50, 70, 65, 75, 40

Test at 5% level whether there is a significant difference between the varieties.

20×3 = 60

SECTION – B

5. Atempt any **three** of the following sub-parts :

20×3 = 60

(a) Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample from the pdf :

$$f(x) = \begin{cases} e^{-(x-\theta)}, & \text{if } x > \theta \\ 0 & \text{otherwise.} \end{cases}$$

Obtain the MLE of θ and show that it is consistent.

(b) Stating the relevant hypothesis, describe Wilcoxon signed rank test for testing the symmetry of a distribution. Obtain the critical region for the test when the sample size is large.

(c) Define Horvitz-Thompson estimator of population total under PPS sampling and show that it is unbiased.

(d) Let $X_1, X_2, \ldots\ldots, X_9$ be a random sample from a normal distribution with mean $\mu = 0$ and unknown standard deviation σ. Determine the best critical region for testing the hypothesis : $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 = 3$ at 5% level of significance.

6. (a) Based on a random sample of size n from Poisson distribution with mean θ, propose an unbiased estimator for $g(\theta) = \theta e^{-\theta}$ and determine the Cramer-Rao lower bound for its variance.

(b) Let p be the probability of getting head in a coin tossing experiment. One is interested in testing the null hypothesis $H_0 : p = \frac{1}{2}$ versus $H_1 : p = \frac{3}{4}$. After tossing the coin 5 times, H_0 is rejected if more that 3 heads are obtained. Find the probability of type I error and power of the test.

(c) Construct a Sequential Probability Ratio Test (SPRT) of strength (α, β) for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ for a Bernoulli distribution with PMF :

$$f(x) = \begin{cases} \theta^x (1-\theta)^{1-x} & \text{if } x = 0, 1, 0 < \theta < 1 \\ 0 & \text{otherwise.} \end{cases}$$

20×3 = 60

7. (a) Show that the sample mean is unbiased for the population mean under simple random sampling without replacement.

(b) Suppose that $C = C_0 + \sum_{h=1}^{L} c_h n_h$ is a cost function associated with a stratified sampling scheme, where n_h is the units to be selected from the stratum h, c_h is the sampling cost per unit and C_0 is the overhead cost. Determine the optimal n_h which minimizes the total cost by fixing the variance of the sample mean $\overline{y}_{st}$.

(c) Define Relative Efficiency (RE) of one design over the other. Obtain an expression for the RE of an RBD relative to CRD. 20×3 = 60

8. (a) Let $X_1, X_2, \ldots, X_n$ be a random sample from an arbitrary population with mean μ and variance $\sigma^2 < \infty$. Show that the moment estimator of μ is consistent and asymptotically normal for μ.

(b) Show that the regression estimator is more precise than the ratio estimator if the regression line passes through the origin and the sample size is large.

(c) Write down the statistical model for a Latin Square Design (LSD) and state all the assumptions. Stating the relevant hypothesis to be tested, present a suitable ANOVA table. State the decision rules. $20 \times 3 = 60$