STATISTICS

PAPER—I

Time Allowed : Three Hours

Maximum Marks : 200

QUESTION PAPER SPECIFIC INSTRUCTIONS

Please read each of the following instructions carefully before attempting questions

There are EIGHT questions in all, out of which FIVE are to be attempted.

Question Nos. **1** and **5** are compulsory. Out of the remaining SIX questions, THREE are to be attempted selecting at least ONE question from each of the two Sections A and B.

The number of marks carried by a question/part is indicated against it.

Unless otherwise mentioned, symbols and notations have their usual standard meanings.

Assume suitable data, if necessary, and indicate the same clearly.

Attempts of questions shall be counted in sequential order. Unless struck off, attempt of a question shall be counted even if attempted partly. Any page or portion of the page left blank in the Question-cum-Answer Booklet must be clearly struck off.

Answers must be written in ENGLISH only.

SECTION—A

1. (a) An unbiased six-sided die is thrown twice. Let X denote larger of the scores obtained. Then show that the probability mass function (p.m.f.) is given by

$$p_X(x) = \frac{2x-1}{36}, \quad x = 1, 2, \ldots, 6$$
$$= 0, \quad \text{elsewhere}$$

8

(b) If X follows the binomial (n, p), then for any $a > 0$, show that

$$P\left[\left|\frac{X}{n} - p\right| \geq a\right] \leq \frac{1}{4na^2}$$

8

(c) Let X_k, $k = 1, 2, 3, \cdots$ be i.i.d. exponential random variables with mean λ. Find the limiting distribution of first-order statistic $X_{(1)}$ of sample of size n. 8

(d) Let X belong to an exponential family of distributions of the form

$$f_X(x, \theta) = e^{x \log\theta + \log g(\theta) + w(x)}$$

Obtain maximum likelihood estimation equation based on n observations. Solve the same for Poisson distribution with variance θ. 8

(e) Find a uniformly minimum variance unbiased estimator for the parameter λ based on a random sample of size n drawn from the distribution

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

8

2. (a) Let $X_1, X_2, \cdots, X_n$ be random sample on X following $U(0, \theta)$. Show that

$$\phi_1(t) = \begin{cases} 1, & \text{if } \quad t > \theta_0 \\ \infty, & \text{if } \quad 0 < t < \theta_0 \end{cases}$$

is most powerful size α test for testing $H_0 : \theta = \theta_0$ against $\theta = \theta_1 > \theta_0$, where $T = X_{(n)}$, nth order statistic. Further, show that

$$\phi_2(t) = \begin{cases} 1, & \text{if } \quad t > \theta_0 \\ \nu(t), & \text{if } \quad 0 < t < \theta_0 \end{cases}$$

gives class of size α tests which have same power as $\phi_1(t)$ subject to the conditions that

$$\int_0^{\theta_0} \nu(t) \frac{nt^{n-1}}{\theta_0^n} dt = \infty$$

and $0 \leq \nu(t) < 1$. 15

(b) Prove that for any $n(\geq 2)$ events $A_j,\ j=1, 2, \cdots, n$

$$\sum_{j=1}^{n} P(A_j) \geq P\left(\bigcup_{j=1}^{n} A_j\right) \geq \sum_{j=1}^{n} P(A_j) - \sum_{1\leq j\leq k\leq n} P(A_j \cap A_k)$$ 15

(c) Let $\{X_n\}$ be a sequence of random variables such that

$$P[X_n = i] = \begin{cases} \frac{1}{n}, & \text{if } i=1 \\ 1-\frac{1}{n}, & \text{if } i=0 \\ 0, & \text{otherwise} \end{cases}$$

Show that $X_n \to 0$ in probability, but not almost surely. 10

3. (a) Let distribution of X belong to one-parameter exponential family of distributions with probability function

$$f_X(x, \theta) = e^{u(\theta)T(x)+v(\theta)+w(x)}$$

Then show that $T(X)$ is unbiased for $\psi(\theta) = -\frac{v'(\theta)}{u'(\theta)}$. Further, show that it is minimum variance bound estimator. 15

(b) Let X follow a geometric distribution with probability function

$$P[X=k] = \left(\frac{a-1}{a}\right)^k \left(\frac{1}{a}\right), \quad k=0, 1, \cdots$$

and a is a positive integer ≥ 2.

(i) Compute the value of $P[|X-E(X)| \geq a(a-1)]$.

(ii) What is the upper bound for the above probability? 7+8=15

(c) Let $X_1, X_2, \cdots, X_n$ be a random sample of X following $U(0, \theta)$. Show that joint likelihood has a monotone likelihood ratio in nth order statistic of the sample. 10

4. (a) For a certain industrial firm, it is found that if a person joins the firm, the probability that he is still there at times x later, where x is measured in years, is

$$\frac{1}{3}e^{-x/2} + \frac{2}{3}e^{-x/4}$$

Find the probability that a person spends between 2 and 4 years with the firm, and also find the mean and variance of length of time he works there. 15

(b) A traveller bus operator has 48-seater passenger buses and 38-seater passenger buses. With X and Y denoting number of miles travelled per day for the 48-seater passenger bus and the 38-seater passenger bus respectively, the bus operator is interested in testing the equality of the distributions $F(x)$ and $G(y)$ of X and Y respectively. That is $F(z) = G(z)$ for all z against $F(z) \neq G(z)$ for at least one value of z.

The company observed the following data on random sample $n_1 = 10$ of 48-seater and $n_2 = 11$ of 38-seater buses :

X	:	104	253	300	308	315	323	331	396	414	452	
Y	:	184	196	197	248	260	279	355	386	393	432	450

Using the normal approximation to run test, carry out run test for the problem by considering $\alpha = 0 \cdot 05$. The table value $z_{\alpha/2} = 1 \cdot 96$. 15

(c) Describe likelihood equivalence of data points. Let $X_1, X_2, \cdots, X_n$ be a random sample on X following $N(\mu, \sigma^2)$. Further, assume that $\sigma^2 = |\mu|$. Obtain minimal sufficient statistic for μ. 10

SECTION—B

5. *(a)* Explain the need for sampling with varying probabilities. Describe a method of selecting a sample with probability proportional to size. 8

(b) Why is it said that a split-plot design confounds main effects? Give an outline of analysis of a split-plot design. 8

(c) Suppose it is required to estimate the average value of output of a group of 5000 factories in a region so that the sample estimate lies within 10% of the true value with a confidence coefficient of 95%. Determine the minimum sample size required. The population coefficient of variation is known to be 60%. 8

(d) A linear model is specified by $E(Y) = X\beta$, where $\beta = (\theta_1, \theta_2, \theta_3)^T$, $Y = (Y_1, Y_2, Y_3, Y_4)^T$ and

$$X = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & -1 & -1 \end{pmatrix}$$

(i) Check whether all parametric functions are estimable.

(ii) Find an estimator for θ_2. 8

(e) Given the correlation matrix

$$P = \begin{bmatrix} 1 & 0 \cdot 5 & -0 \cdot 3 \\ 0 \cdot 5 & 1 & -0 \cdot 47 \\ -0 \cdot 3 & -0 \cdot 47 & 1 \end{bmatrix}$$

compute multiple correlation coefficient $R_{1(23)}$ and partial correlation coefficient $r_{13 \cdot 2}$. 8

6. *(a)* Under the assumption that the values of the units in population are steadily increasing by a constant amount, show that

$$\text{Var(st)} < \text{Var(sy)} < \text{Var(SRS)}$$ 15

(b) A system of partial confounding in a 2^3-experiment in block of 2^2 is

Replicate—I Control Block	Replicate—II Control Block	Replicate—III Control Block	Replicate—IV Control Block
(1)	(1)	(1)	(1)
ab	ac	bc	ab
c	b	a	ac
abc	abc	abc	bc

Identify confounded effects in different replicates. Considering r-replications of this basic model, write down the ANOVA table. 15

(c) What purpose does discriminant analysis serve? Explain how Fisher discriminant function is useful in this analysis. 10

7\. (a) Find the mean vector and dispersion matrix of the normal distribution given below :

$$f(x, y) \propto e^{-\frac{1}{2}\{2x^2 + y^2 + 2xy - 22x - 14y + 65\}}, \quad -\infty < x, y < \infty$$ 15

(b) Let $X \sim N_p(\mu, \Sigma)$ and $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$, where $X^{(1)}$ is a $q \times 1$ and $X^{(2)}$ is a $(p-q) \times 1$ vector of random variables. Obtain the marginal density of $X^{(2)}$ and the conditional expectation of $X^{(1)}$, given $X^{(2)} = x^{(2)}$. 15

(c) Prove that the number of treatments common between any two blocks in a symmetrical BIBD is λ. 10

8\. (a) (i) If $X \sim N_p(\mu, \Sigma)$, then show that $(X-\mu)^T \Sigma^{-1}(X-\mu)$ is distributed as chi-square with parameter p.

(ii) State and explain any two uses of T^2-statistic. 9+6=15

(b) Define Hotelling's T^2-statistic. Show that it is invariant under the transformation of scale. 10

(c) A sample of 30 students is to be drawn from a population consisting of 300 students belonging to two colleges *A* and *B*. The mean and SD (standard deviation) of their marks are given below :

	Total number of students (N_i)	*Mean* $(\bar{y}_{N_i})$	*SD* (σ_i)
College *A*	200	30	10
College *B*	100	60	40

Determine the sample sizes in two colleges by using proportional allocation technique. Hence, obtain variance of estimate of the population mean. 15

★ ★ ★