

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.*

*Questions no. **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.*

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.

All questions carry equal marks. 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.

*Answers must be written in **ENGLISH** only.*

SECTION A

- Q1.** (a) (i) Find α such that P is a finitely additive probability measure, where $\Omega = \{1, 2, 3\}$. \mathcal{F} consists of all subsets of Ω , and $P(\{1\}) = \frac{1}{3}$, $P(\{2\}) = \frac{1}{6}$, $P(\{3\}) = \alpha$. Compute $P(\{1, 2\})$, $P(\{1, 3\})$ and $P(\{2, 3\})$. 4
- (ii) Among $t = 60$ lottery tickets, $w = 20$ win prizes. We buy $b = 6$. What is the probability that $g = 2$ will be winning? Generalize this to arbitrary numbers t, w, b, g . 4
- (b) A fair coin is tossed independently n times. Let S_n be the number of heads obtained. Use Chebyshev's inequality to find a lower bound of the probability that $\frac{S_n}{n}$ differs from $\frac{1}{2}$ by less than 0.1 for $n = 100$. 8
- (c) If t is a consistent estimator of θ , then prove that t^2 is consistent for θ^2 . 8
- (d) Define absorbing, transient, recurrent and periodic states in a Markov chain. Also, test the periodicity of the states of a Markov chain with the following transition probability matrix : 8

$$P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0 & 0.6 & 0.4 \\ 0 & 1 & 0 \\ 0.6 & 0.4 & 0 \end{pmatrix} \end{matrix}$$

- (e) The joint probability density function of two random variables X and Y is

$$f(x, y) = \begin{cases} \frac{1}{4}(1 + xy), & |x| < 1, \quad |y| < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Show that X and Y are not independent but X^2 and Y^2 are independent. 8

- Q2.** (a) Suppose the probability generating function of a random variable X is

$$g_X(t) = e^{\lambda(t-1)}.$$

- (i) Find the probability mass function of the random variable X .
- (ii) Find the probability generating function of $Y = 3X + 2$.
- (iii) Obtain variance of Y . 3+3+4=10

(b) A coin is tossed. If it shows heads, you pay 2 Rupees. If it shows tails, you spin a wheel which gives the amount you win, distributed with uniform probability between 0 and 10 Rupees. Your gain (or loss) is a random variable X . Find the distribution function and use it to compute the probability that you will not win at least 5 Rupees. 10

(c) (i) If a random sample of size n is taken from $N(\mu, \sigma^2)$ and σ^2 is known but μ is not known, then show that

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

is not a sufficient estimator for σ^2 . Also, suggest a sufficient estimator for σ^2 . 15

(ii) Describe the role of Cramer-Rao Inequality, Rao-Blackwell and Lehmann-Scheffé theorems in the estimation of unknown parameters of the distributions. 5

Q3. (a) The observed value of mean of a random sample from $N(\theta, 1)$ distribution is 2.3. If the parameter space is $\theta = \{0, 1, 2, 3\}$, then find the maximum likelihood estimate of θ . 10

(b) Let X have a pdf

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

Find the power function of the test to test the simple hypothesis $H_0 : \theta = 1$ against the alternative simple hypothesis $H_1 : \theta = 2$ using a random sample of X_1 and X_2 of size $n = 2$ and defining the critical region to be

$$W = \left\{ (x_1, x_2) : \frac{3}{4x_1} \leq x_2 \right\}. \quad 10$$

(c) (i) Suppose that X is uniformly distributed on the interval $(-2, 3)$. Let $Y = X^2$. Find the density function of Y . 12

(ii) Let the random variables X and Y be jointly distributed. The marginal distribution of X is

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$$

and the conditional distribution of Y given $X = x$ is

$$p_{Y/X}(y/x) = \binom{x}{y} p^y (1-p)^{x-y}, \quad x > y, \quad y = 0, 1, 2, \dots, x.$$

Find the marginal distribution of Y . 8

- Q4.** (a) Verify that there exists a Minimum Variance Bound Unbiased Estimator (MVBUE) of the parameter θ of the distribution

$$f(x; \theta) = \frac{e^{-\theta} \cdot \theta^x}{x!}; x = 0, 1, 2, \dots$$

Hence, obtain variance of the MVBUE so obtained. 10

- (b) (i) Let X_1, X_2, \dots, X_n ($n > 4$) be a random sample from a population $N(\mu, \sigma^2)$. Consider the following estimators of μ :

$$U = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad V = \frac{1}{8} X_1 + \frac{3}{4(n-2)} (X_2 + \dots + X_{n-1}) + \frac{1}{8} X_n,$$

then examine whether U and V are unbiased estimators μ . Also, find which of the two estimators is more efficient. 12

- (ii) In what situation do we make use of non-parametric tests ? Test the hypothesis of no difference between the ages of male and female employees of a certain company using the Mann-Whitney U-test for the sample data given below : 8

Males (Age)	31	25	38	33	42	40	44	26	43	35
Females (Age)	44	30	34	47	35	32	35	47	48	34

Use 0.10 level of significance with $Z_{(0.10)} = 1.64$.

- (c) The joint probability density function (pdf) of two random variables X and Y is

$$f(x, y) = \frac{9(1+x+y)}{2(1+x)^4 (1+y)^4}, 0 \leq x < \infty, 0 \leq y < \infty.$$

Find the conditional distribution of Y given $X = x$. 10

SECTION B

- Q5.** (a) Explain the problem of multicollinearity. What are its consequences ? State different measures to detect the presence of multicollinearity. 8
- (b) Given below is an ANOVA table of an RBD with some missing entries denoted by (*). Find out the missing entries.

RBD ANOVA TABLE

Source	d.f.	S.S.	M.S.S.	F _{cal}
Blocks	3	*	2.040	*
Treatments	5	15.440	*	*
Error	*	7.030	*	
Total	23	28.590		

Stating clearly the hypotheses to be tested, give your conclusions. 8

Given that $F(3, 15; 5\%) = 3.29$

$F(5, 15; 5\%) = 2.90$

- (c) Define a linear model. Develop 100 (1 - α)% confidence interval for an estimable linear parametric function $\lambda'\theta$ in a linear model in a Gauss-Markov set-up ($Y; A\theta, \sigma^2 I_n$). 8
- (d) Explain Warner's randomised response technique for sensitive characteristics. 8
- (e) Let X_1, X_2, \dots, X_n be a random sample from an $N_p(\mu, \Sigma)$. Give the test statistic for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$; and test the hypothesis $H_0 : \mu = \begin{pmatrix} 7 \\ 11 \end{pmatrix}$ using the data $X = \begin{bmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{bmatrix}$. 8

Given : $F(2, 2; 5\%) = 19$

Q6. (a) The data given below is on yields of a Latin Square Design of order 6.

Total \ No.	1	2	3	4	5	6
Rows	26.35	25.85	33.45	36.75	33.25	42.85
Columns	26.75	28.90	31.45	34.40	36.40	40.60
Treatments	22.20	28.30	34.20	31.60	38.55	43.55
$\Sigma y_{ij}^2 = 1222.84$		$\Sigma y_{ij} = 198.50$				

State all hypotheses and carry out analysis. Write ANOVA, Estimates of test statistics, conclusions and interpretations. 15

Given that $F(6, 20; 5\%) = 2.60$

$F(5, 20; 5\%) = 2.71$

$F(6, 20; 1\%) = 3.87$

$F(5, 20; 1\%) = 4.10$

- (b) (i) Define a multiple linear regression model. Obtain ordinary least squares (OLS) estimator of regression coefficient β and show that it is best linear unbiased estimator (BLUE). 6
- (ii) Give the test statistic for testing linear restrictions of the type $R\beta = r$ in multiple linear regression and obtain its distribution. From this, deduce the test statistic for testing the significance of any regressor, say X_j . 4
- (iii) Define the OLS residual. Show that it is heteroscedastic and autocorrelated. How is this residual useful in detecting
- (1) heteroscedasticity,
 - (2) autocorrelation, and
 - (3) normality? 5
- (c) Define ratio estimator and regression estimator. Show that ratio estimator is biased; further obtain the bias. Derive the condition for ratio and regression estimators to be equally efficient. 10

Q7. (a) What is unequal probability sampling ? When is it to be used ? Under this scheme without replacement,

- (i) explain Lahiri's method of sample selection, and
- (ii) obtain Horvitz-Thompson estimator of variance of population total and give your comments. 6+9=15

(b) (i) Suppose X_1, X_2, \dots, X_N are independent each distributed according to $N_k(\mu, \Sigma)$ and $C = ((C_{ij}))$ be an orthogonal matrix.

Show that $Y_i = \sum_{j=1}^N C_{ij} X_j$ is multivariate normal with mean

vector $C\mu$ and variance covariance matrix Σ . 4

(ii) Let X_1, X_2, \dots, X_n be a random sample from a normal population with mean vector μ and covariance Σ . Show that

$$\hat{\mu} = \bar{X} \text{ and } \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$$

are the maximum likelihood estimators of μ and Σ . 5

(iii) Explain canonical correlation analysis. Obtain the first pair of canonical covariates and first canonical correlation given the covariance matrix of two groups of variables 6

$$X^{(1)} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \end{bmatrix} \text{ and } X^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \end{bmatrix}$$

$$\text{as cov} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 1 & 0.95 & 0 \\ 0 & 0.95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

(c) Consider the following regression output :

$$\hat{y} = 0.2033 + 0.6560 X$$

$$SE = (0.0976) \quad (0.1961)$$

Residual sum of squares = 0.0544

Regression sum of squares = 0.0358

where

Y = Labour Force Participation Rate (LFPR) of women in 1972

X = LFPR of women in 1968

The regression results were obtained from a sample of 19 cities. The figures in the parentheses give Standard Error (SE) of the estimate.

Critical t value at 1%, df = 17 is $t_{17, 1\%} = 2.567$.

- (i) How do you interpret the result ? Compute the coefficient of determination R^2 .
- (ii) Test the hypothesis that $H_0 : \beta = 0$ vs $H_1 : \beta > 0$. Which test do you use and why ? What are the underlying assumptions of the test you use ?
- (iii) Write 95% confidence interval for α and β .
- (iv) Test that the hypothesis LFPR of women in 1972 is not depending on LFPR of women in 1968. 10

Given $t(17; 5\%) = 1.740$

$F(1, 17; 5\%) = 4.45$

- Q8.**
- (a)
 - (i) What is confounding ? Explain briefly, types of confounding and compare them. 7
 - (ii) Obtain single replication of a 2^5 factorial experiment confounding interactions $X = ABC$ and $Y = ACDE$. 8
 - (b)
 - (i) Let $(Y; A\theta, \sigma^2I)$ be a Gauss-Markov set-up. Obtain the least squares estimator of θ and variance of the best estimator of estimable linear parametric function $\lambda'\theta$. 6
 - (ii) Define a linear hypothesis. Derive the maximum likelihood ratio test statistic for testing a linear hypothesis $H_0 : \xi_1 = \dots = \xi_r$ in a Gauss-Markov model $(Y; \xi, \sigma^2I_n)$, where ξ is a mean vector of Y and σ^2I_n is the dispersion matrix of Y. Obtain the distribution of the test statistic. 9
 - (c) Define : Connectedness, Orthogonality and Balancedness. State and prove a necessary and sufficient condition for an incomplete block design to be balanced. 10