

STATISTICS

(Paper I)

Time Allowed : Three Hours
Maximum Marks : 300

QUESTION PAPER SPECIFIC INSTRUCTIONS

(Please read each of the following instructions carefully before attempting questions)

There are EIGHT questions divided in two Sections.

Candidate has to attempt FIVE questions in all.

Question Nos. 1 and 5 are compulsory and out of the remaining, any THREE are to be attempted choosing at least ONE question from each Section.

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

Wherever any assumptions are made for answering a question, they must be clearly indicated.

Word limit in questions, wherever specified, should be adhered to.

SECTION 'A'

1. Answer any five parts of the following:

12×5=60

- (a) 5% residents of a locality are earning more than Rs 10 lakhs per annum. The percentage of single earner in the family among those earning more than Rs 10 lakhs per annum is 80%, whereas percentage of single earner among others (earning less than Rs 10 lakhs per annum) is 50%. A randomly selected resident of that locality was found to be single earner. What is the probability that he earns more than Rs 10 lakhs per annum?
- (b) T is a triangle in (X,Y) plane with vertices $(0,0)$, $(0,\sqrt{2})$ and $(\sqrt{2},\sqrt{2})$. Let $F(x,y)$ denote the area of intersection of T and the set $S=\{(a,b):a\leq x, b\leq y\}$. Does F define a cumulative distribution function? If yes, find the joint probability density function.
- (c) Two players A and B decide to play a series of at most 7 games. A player who wins 4 games wins the series. If both the players have equal chance of winning a game and A wins the first two games, what is the probability that A will win the series?
- (d) A city has N taxis numbered 1 to N . A person standing at a crossing noted numbers of n taxis that pass that crossing (same taxi does not pass twice). If M_n is the highest taxi number noted, find an unbiased estimate of N based on M_n .
- (e) Let $X_1, X_2, \dots, X_n$ be the lifetime of n randomly selected patients of cancer. An estimate is required for the probability that a patient survives beyond time t , i.e., $P(X>t)=S(t)$. Find the maximum likelihood estimator of $S(t)$, if lifetime X of the patients follow the following probability density function:

$$f(x)=\begin{cases} 2x/\beta \exp(-x^2/\beta), & \text{if } x>0; \\ 0, & \text{otherwise} \end{cases}$$

- (f) Let U_X be the number of Y 's those are smaller than X 's in independent random samples $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_m$. Find $E(U_X)$.

2. (a) Cumulative distribution function (c.d.f.) of a random variable X is

$$F(x)=\begin{cases} 0, & \text{if } x<0; \\ 1/4, & \text{if } 0\leq x<1; \\ x/3, & \text{if } 1\leq x<2; \\ 1, & \text{if } x\geq 2 \end{cases}$$

Find α such that we can write $F(x)=\alpha F_c(x)+(1-\alpha)F_d(x)$, where $F_c(x)$ and $F_d(x)$ are c.d.f. of continuous and discrete random variables respectively. Find $F_c(x)$ and $F_d(x)$. **20**

- (b) The joint probability density function of random variables X and Y is given as

$$f(x,y)=\begin{cases} k(y^2-x^2)e^{-y}, & \text{if } |x|<y, y>0; \\ 0, & \text{otherwise} \end{cases}$$

Comment on independence and correlation between X and Y . **20**

- (c) Explain consistency of estimators. Prove that second smallest observation in a random sample of size n from the following probability density function is consistent estimator of θ :

$$f(x|\theta) = \begin{cases} \exp(-(x-\theta)), & \text{if } x > \theta; \\ 0, & \text{otherwise} \end{cases} \quad 20$$

3. (a) The joint probability density function of X and Y is

$$f(x,y) = \begin{cases} 1/2xe^{-y}, & \text{if } 0 < x < 2, y > 0; \\ 0, & \text{elsewhere} \end{cases}$$

Derive the distribution of $X+Y$. 20

- (b) Explain convergence in r th mean and convergence almost sure. Check various modes of the convergence for $\{X_n\}$, where X_n 's are independently distributed as follows:

$$P[X_n = 0] = 1 - 1/n^r \text{ and } P[X_n = n] = 1/n^r$$

where $r \geq 2$ and $n = 1, 2, \dots$ 20

- (c) State and prove Wald's fundamental identity of sequential analysis. 20

4. (a) The following is a random sample of 10 numbers lying between (0,1): 0.22 0.90 0.96 0.78 0.66 0.18 0.73 0.58 0.11 0.97 We wish to test the hypothesis H_0 : sample comes from uniform distribution on (0,1). Explain, how you will use Kolmogorov-Smirnov test of goodness of fit for H_0 . Obtain the value of test statistic. 20

- (b) A random observation X belongs to a population having probability density function $f(x)$. Describe likelihood ratio test for testing $H_0: f(x) = f_1(x)$ against $H_1: f(x) = f_2(x)$ based on a single observation, when

$$f_1(x) = 1/\sqrt{(2\pi)}e^{-1/2x^2}, -\infty < x < \infty$$

and

$$f_2(x) = 1/2e^{-|x|}, -\infty < x < \infty \quad 20$$

- (c) $X_1, X_2, \dots, X_n$ is a random sample from Bernoulli population with parameter θ . Consider that θ has beta prior distribution with hyper-parameters (a, b) . Obtain Bayes estimator for θ under quadratic loss function and hence obtain min max estimator for θ . 20

SECTION 'B'

5. Answer any five parts of the following:

12×5=60

- (a) Consider a population of 3 units U_1, U_2 and U_3 . The size measures of U_1, U_2 and U_3 are respectively $X_1=6, X_2=4$ and $X_3=2$. A sample of two units is drawn from the population without replacement such that the first unit is selected with probability proportional to their sizes and second unit is selected with probability proportional to the sizes of the remaining units. If π_i denotes the probability of inclusion of U_i in the sample, show that

$$\pi_1 = 51/60, \pi_2 = 44/60, \pi_3 = 25/60$$

- (b) $X = [X_1, X_2, X_3]'$ has 3-variate normal distribution with mean vector μ and covariance matrix Σ , where

$$\mu = [-3; 1; 4] \text{ and } \Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Show that X_1 and X_2 are not independent but $[X_1, X_2]'$ is independent of X_3 . Further show that $1/2(X_1 + X_2)$ and X_3 are also independent.

- (c) $X = [X_1, X_2, X_3]'$ has 3-variate normal distribution with mean vector μ and dispersion matrix Σ , where

$$\mu = [2; -1; 3] \text{ and } \Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}$$

Derive the distribution of

$$Y = 4X_1 - 6X_2 + X_3 \text{ and } Z = [X_1 - X_2 + X_3; 2X_1 + X_2 - X_3]$$

- (d) Consider the mean vector for $X = [X_1; X_2]$ to be $\mu_x = [3; -2]$ and for Y to be $\mu_y = 4$. The covariance matrices for X_1, X_2 and Y are

$$\Sigma_{xx} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \sigma_{YY} = 9 \text{ and } \Sigma_{XY} = \begin{bmatrix} 3 & 1 \end{bmatrix}$$

Fit the equation $Y = b_0 + b_1 X_1 + b_2 X_2$ and calculate multiple correlation coefficient.

- (e) In a population consisting of three units U_1, U_2 and U_3 , the observations are Y_1, Y_2 and Y_3 respectively. A sample of size 2 is drawn from the population by using simple random sampling without replacement. Define an estimator T as follows:

$$T = \begin{cases} 1/2Y_1 + 1/2Y_2, & \text{if } U_1 \text{ and } U_2 \text{ selected;} \\ 1/2Y_1 + 2/3Y_3, & \text{if } U_1 \text{ and } U_3 \text{ selected;} \\ 1/2Y_2 + 1/3Y_3, & \text{if } U_2 \text{ and } U_3 \text{ selected} \end{cases}$$

Show that T is unbiased estimator of population mean and has smaller variance than sample mean if $Y_3(3Y_2 - 3Y_1 - Y_3) > 0$.

- (f) Explain in detail, the procedure of randomization (i.e., allocation of treatments randomly to plots) followed in randomized block design and Latin square design.

6. (a) Let $X = [X_1, X_2, \dots, X_p]'$ denote a column vector of variables X_i 's. Prove that Y follows a p -variate normal distribution $N_p(\mu, \Sigma)$ iff there exists a random $X = [X_1, X_2, \dots, X_p]'$ of independent standard normal variables X_i 's such that $Y = \mu + BX$ with probability one for some matrix B of full rank and $BB' = \Sigma$. 20

- (b) Distinguish between 2^3 - and 3^2 -factorial experiments. Explain, how you will get sum of squares due to main and interactions in 3^2 -factorial experiments. Give format of ANOVA table for 3^2 -factorial experiment mentioning degrees of freedom if randomized block design having r blocks containing 9 plots each is used. 20

- (c) From a population of N units, a random sample of 3 units is drawn by simple random sampling with replacement. Find the expectation of distinct units in the sample. Define T as mean of distinct units in the sample. Show that T is unbiased estimate of population mean. 20

7. (a) Let X be distributed as $N_p(\mu, \Sigma)$ (a p -variate normal distribution). Prove that $Y = CX$, where C is $r \times p$ matrix of rank r ($1 \leq r \leq p$), is distributed as $N_r(C\mu, C\Sigma C')$. 20

- (b) The variable X under study has rectangular distribution in the interval $(a, a+d)$. The interval is divided into k equal subintervals which form k strata of equal sizes. From each stratum, a simple random sample of n/k units is drawn. Let V_1 and V_2 be the variances of estimator of population mean based on stratified and unstratified samples of size n respectively. Prove that $V_1/V_2 = k^{-2}$. 20

- (c) In a randomized block design with k treatments and r replications, two observations are missing. One missing observation is related to i th treatment in j th block and the other is related to α th treatment in β th block ($i \neq \alpha$ and $j \neq \beta$). How will you estimate the missing observations? Derive the expressions for the estimate of missing observations. 20

8. (a) Define ratio estimator for estimating the population total and derive expression for the standard error of the estimator. 20

- (b) Obtain an expression for estimated value of Y for given $X = x_0$ by fitting a curve $e^{XY} = a + be^{2X}$ to n given points (x_i, y_i) , $i = 1, 2, \dots, n$ using least-square method. 20

- (c) Let $X_1, X_2, \dots, X_n$ be a random sample from p -variate normal distribution $N_p(\mu, \Sigma)$, where $X_\alpha = [X_{1\alpha}, X_{2\alpha}, \dots, X_{p\alpha}]'$, $\alpha = 1, 2, \dots, n$; $1_{p \times 1} = [1, 1, \dots, 1]'$ and $\Sigma_{p \times p}$ is known. Define

$$T_1 = 1' \bar{X} / 1' 1 \text{ and } T_2 = (1' \Sigma^{-1} \bar{X}) / (1' \Sigma^{-1} 1)$$

where $\bar{X} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p]'$ and

$$\bar{X}_i = 1/n \sum_{\alpha=1}^n X_{i\alpha}, i = 1, 2, \dots, p$$

Prove that T_1 and T_2 are unbiased estimators of μ and $V(T_2) < V(T_1)$. 20