

STATISTICS

(Paper I)

Time Allowed : Three Hours
Maximum Marks : 250

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.

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.

SECTION 'A'

1. (a) Let the probability that a family has exactly n children be αp^n ($\alpha > 0$, $0 < p < 1$). Assume that a child can be a boy or a girl with equal probability. (i) Obtain the probability that a family has exactly k (≥ 1) boys. (ii) Given that a family has at least one boy, what is the probability that there are two or more children in that family? 10
- (b) Define convergence in distribution and in mean. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random variables such that

$$X_n = \begin{cases} n^3 & \text{with probability } 1/n^2 \\ 0 & \text{with probability } 1-1/n^2 \end{cases}$$
 Show that $X_n \rightarrow 0$ in distribution. 10
- (c) Define a sufficient statistic. If X_1 and X_2 are Bernoulli (p) random variables, examine the sufficiency of T_1 and T_2 for p where $T_1 = X_1 + X_2$ and $T_2 = X_1 + 5X_2$. 10
- (d) Let $X_1, X_2, \dots, X_m$ and $Y_1, Y_2, \dots, Y_n$ be independent random samples from $N(3, \sigma_1^2)$ and $N(2, \sigma_2^2)$ respectively. Find a confidence Interval for σ_1^2/σ_2^2 at confidence level $1-\alpha$. 10
- (e) Consider n observations $X_1, X_2, \dots, X_n$ from $B(1, \theta)$. Obtain the Bayes estimator for θ under quadratic loss function when a conjugate prior is assumed for θ . 10
2. (a) How can you find out the discontinuity points of distribution of a random variable if the characteristic function $\phi(t)$ is given to you? Hence find the c.d.f. of a random variable X whose characteristic function is $1/4(1 + e^{it} + 2e^{2it})$. 20
- (b) Let X_n be a Poisson $P(n\theta)$, $n \geq 1$, $\theta > 0$. Use the Central Limit Theorem to find the smallest n such that

$$P(|X_n/n - \theta| \leq \sqrt{\theta/10}) \geq .99.$$
 15
- (c) The distribution under the null hypothesis of X is uniform over $1, 2, \dots, 10$ and under the alternative hypothesis, the distribution is given by $P\{X=k\}=0$, for $k=1, 2, 3, 4$, $P\{X=k\}=1/8$ for $k=5, 6, 7, 8, 9$ and $P\{X=10\}=3/8$. Obtain most powerful test of size 0.15 and find its power, based on a sample of size one. 15
3. (a) A trial can result in one of three possible outcomes with probabilities p_1 , p_2 and p_3 ($p_1 + p_2 + p_3 = 1$) respectively. Construct the Likelihood Ratio test of $H_0: p_1 = p_2 = 1/3$ against the alternative that these probabilities are different from $1/3$. The test needs to be constructed on the basis of n -trials. 20
- (b) How does one obtain moment estimators? If X has the pdf

$$f_X(x) = x/\theta e^{-((x-\beta))/\theta}, \quad 0 < \beta < x < \infty, \theta > 0$$
 15

and a sample of size 50 yields

$$\sum_{i=1}^{50} x_i = 8200 \text{ and } \sum_{i=1}^{50} x_i^2 = 20,00,000$$

find moment estimators of θ and β .

- (c) Find a Most Powerful test for testing

$$H_0: f(x) = 1/\sqrt{(2\pi)} e^{-1/2x^2}, -\infty < x < \infty$$

versus

$$H_1: f(x) = 1/2 e^{-|x|}, -\infty < x < \infty$$

on the basis of a random sample of size n . Here $f(x)$ is the pdf of the random variable X .

15

4. (a) Let $F(\cdot)$ be the cumulative distribution function (cdf) given by

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/2 + ((x-1))/4 & \text{if } 1 \leq x < 2 \\ 7/8 + ((x-2))/8 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

Find the set of discontinuity points of F and express F as a mixture of its discrete and continuous parts.

20

- (b) Suppose that X and Y have cdf F and G respectively. Given independent random samples $(x_1, x_2, \dots, x_m)$ and $(y_1, y_2, \dots, y_n)$ from these distributions, construct Mann-Whitney test of $H_0: F(t) = G(t)$ for all t against the alternative that $F(t) > G(t)$ for at least one t . Also indicate the test when the alternative is $F(t) \neq G(t)$ for some t . State the test you would use when m and n are large.

15

- (c) Consider observations from Poisson distribution with parameter λ . Develop a Sequential Probability Ratio Test (SPRT) to test $H_0: \lambda = \lambda_0$ against $H_1: \lambda = k\lambda_0$, $k > 1$. Obtain OC and ASN functions.

15

SECTION 'B'

5. (a) A random sample of size 4 from a bivariate normal population provides the following statistics

$$\bar{X} = ([4; 6]), S = ([6 \ -2; -2 \ 3])$$

where $\bar{X}$ is the sample mean and S is an unbiased estimator of the population dispersion matrix. Test the hypothesis $H_0: \mu = (5, 5)'$ where μ is the population mean vector. ($F_{.05, 1, 3} = 10.13$, $F_{.05, 2, 2} = 19.00$, $F_{.05, 3, 1} = 215.7$)

10

- (b) Let $X = (X_1, X_2, \dots, X_p)'$ be a p -dimensional random vector with $E(X) = 0$, $V(X) = \Sigma$. Define partial correlation coefficient $\rho_{12.(3 \dots p)}$ between X_1, X_2 . Show that
- $$\rho_{12.(3 \dots p)}^2 = (\sigma^{12})^2 / \sigma^{11} \sigma^{22}$$
- where σ^{ij} is the (i, j) th element of Σ^{-1} .

10

- (c) Consider three independent random variables Y_1, Y_2 and Y_3 having common variance σ^2 and $E(Y_1) = \beta_1 + \beta_3$, $E(Y_2) = \beta_1 + \beta_2$, $E(Y_3) = \beta_1 + \beta_3$. Determine the condition of estimability of the linear parametric function $l'\beta$. Obtain a solution of the normal equation and the S.S. due to error.

10

- (d) Show that

$$V_{\text{ran}} = V_{\text{prop}} + (N-n)/(nN(N-1)) [\sum_h N_h (\bar{Y}_h - \bar{Y})^2 - 1/N \sum_h (N - N_h) S_h^2]$$

where V_{ran} and V_{prop} are respectively the variances of the estimated means under simple random sampling and stratified random sampling with proportional allocation. All other notations have their usual interpretations.

10

- (e) Describe the layout of a 3^3 experiment in 4 replicates (with 3 blocks per replicate) using complete confounding.

10

6. (a) For an arbitrary fixed effective size sampling design with positive second order inclusion probabilities, derive the Yates-Grundy form of the variance of the Horvitz-Thompson estimator of a finite population total and hence obtain the Yates-Grundy unbiased estimator of this variance.

20

- (b) Let $X_{\sim}=(X_1, X_2, X_3)'$ have the joint moment generating function (mgf)
- $$M_{X_{\sim}}(t_{\sim})=\exp[t_1-t_2+2t_3+t_1^2+t_2^2/2+2t_3^2-t_1t_2/2-t_1t_3]$$
- where $t_{\sim}=(t_1, t_2, t_3)'$. Then: (i) Obtain the covariance matrix and the mean vector of $X_{\sim}$. (ii) Find a constant C such that $P(2X_1-3X_2+X_3>C)=0.95$. (iii) Derive the conditional distribution of X_1 given $X_2=x_2$ and $X_3=x_3$. (If necessary, you can use $P(\tau>1.645)=0.05$ and $P(\tau>1.96)=0.01$ where τ is a standard normal variate) 15
- (c) For the Gauss Markov Model $(Y_{\sim}, X\beta_{\sim}, \sigma^2 I)$, the estimator $t_{\sim}'Y_{\sim}$ is the best linear unbiased estimator (BLUE) for $E(t_{\sim}'Y_{\sim})$ iff $t_{\sim}'Y_{\sim}$ is uncorrelated with all unbiased estimators of zero. 15
7. (a) Define a balanced incomplete block design (BIBD). Carry out its intrablock analysis. 20
- (b) Suppose a random vector $X_{\sim}$ has the covariance matrix $\Sigma=([3 \ 1 \ 1; 1 \ 3 \ 1; 1 \ 1 \ 5])$. Find the principal components of $X_{\sim}$ and obtain the proportion of the total variance accounted for by the first two principal components. 15
- (c) Consider the simple linear regression model $y_i=\beta_0+\beta_1x_i+e_i, i=1,2,\dots,n$. Show that if the x_i 's are equally spaced (i.e. $x_i=u+iv$ for fixed values of u and v), then $y_i=y_0+y_1i+e_i$ is an equivalent reparametrization (in the sense that both the design matrices have the same column space). 15
8. (a) Consider a random sample $x_{\sim 1}, x_{\sim 2}, \dots, x_{\sim n}$ from a p -dimensional normal population with mean vector $\mu_{\sim}$ and dispersion matrix Σ . The purpose of the problem is to construct the likelihood ratio test (LRT) for testing $H_0:\Sigma=\sigma_p^2 I_p$. Find the maximum likelihood estimate (MLE) of μ and σ^2 . Write down the expression for $-2\log_e \Lambda$ in terms of $p, n, \bar{X}_{\sim}$ and the elements of the sample covariance matrix S . 20
- (b) What is meant by confounding in a factorial experiment? Why is confounding used even at the cost of loss of information on the confounded effects? Explain the terms 'complete confounding' and 'partial confounding'. 15
- (c) A simple random sample of n clusters, each containing M elements, is drawn from the N clusters of the population and the clusters sampled are enumerated completely. Suggest an unbiased estimator of the population mean per element and derive the expression for the variance of the proposed estimator in terms of the population intraclass correlation coefficient. 15