

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. (a) A tyre manufacturing company has three production units A, B and C which produce respectively 50%, 30% and 20% of the entire production. Unit A is known to produce 5% defectives, unit B produces 3% defectives and unit C produces 1% defectives. What percentage of items in the entire produce is defective? If an item is found to be defective, what is the probability that it was produced at unit C? 12
- (b) A bivariate random vector (X, Y) is specified by

$$P(X > x, Y > y) = e^{-(x+y+\theta xy)}; x, y > 0, 0 \leq \theta \leq 1.$$
 Obtain (i) the p.d.f. (ii) the marginal p.d.f's and (iii) the conditional p.d.f's. When will X and Y be independent? 12
- (c) Examine whether the central limit theorem holds for the sequence $\{X_n, n \geq 1\}$ of random variables with

$$P(X_n = n) = P(X_n = -n) = 1/2n^3, P(X_n = 0) = 1 - 1/n^3, n \geq 1.$$
12
- (d) Assume that X follow the Binomial, $B(n, \theta)$, and that the prior distribution of θ is the Uniform, $U(0, 1)$. Find the Bayes' estimator for θ using the squared error loss function. 12
- (e) $X_1, X_2, \dots, X_n$ is a random sample from the normal distribution with 0 mean and variance σ^2 . Examine the validity of the following statements: (i) $\sum_{i=1}^n X_i^2$ is consistent for σ^2 . (ii) $T = 1/n \sum_{i=1}^n X_i^2$ is unbiased for σ^2 . (iii) T is sufficient for σ^2 . (iv) T is the UMVUE for σ^2 . 12
2. (a) Define exponential family of distributions. Identify two distributions each which (i) belongs to the family and (ii) does not belong to the family. Substantiate your answer. 20
- (b) For a continuous non-negative random variable X with $E(X) < \infty$, show that for $t > 0$.

$$E(X - t | X > t) = 1/\bar{F}(t) \int_t^\infty \bar{F}(x) dx,$$
 where $\bar{F}(x) = P(X > x)$. Evaluate the conditional expectation for the exponential model specified by $\bar{F}(x) = e^{-x/\lambda}; x > 0$. 20
- (c) Show that for a sequence of random variables $\{X_n\}$, convergence in probability implies convergence in distribution. Is the converse true? Substantiate your answer. 20
3. (a) Based on a random sample of size n from the discrete Uniform distribution

$$p(x) = 1/N, x = 1, 2, 3, \dots, n; = 0, \text{ otherwise,}$$
 obtain the UMVUE of N. Is this estimator consistent? 20
- (b) When will the Neyman-Pearson test and likelihood ratio test be the same? Obtain the likelihood ratio test for testing $H_0: p = p_0$ against $H_1: p \neq p_0$ based on a random sample from the binomial, $b(1, p)$. 20

- (c) A random sample of size n is taken from the Poisson (λ) distribution. Let N be the number of observations which are equal to 0. Obtain an unbiased estimator for $e^{-\lambda}$ and improve the estimator using Rao-Blackwell theorem. 20
4. (a) Comment on the validity of the statement “the S.P.R.T. terminates with probability one”. For the geometric distribution with parameter p , derive the S.P.R.T. for testing $H_0: p=p_0$ against $H_1: p=p_1$. Also obtain the O.C. function. 20
- (b) Compare the Chi-square test and Kolmogorov-Smirnov test as tests for goodness of fit. Show that the Kolmogorov-Smirnov one sample test is distribution free. Also state the large sample distribution of the Kolmogorov-Smirnov test statistic. 20
- (c) Describe the relationship between confidence interval estimation and testing of hypothesis. Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$, μ known. Obtain a confidence interval for σ^2 . 20

SECTION 'B'

5. (a) X follow $N_3(0, \Sigma)$ where $\Sigma = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}$. Find ρ such that $(X_1 + X_2 + X_3)$ and $(X_1 - X_2 - X_3)$ are independent. 12
- (b) Y_1, Y_2, Y_3 and Y_4 are independent stochastic variables with common variance σ^2 and $E(Y_1) = E(Y_2) = \mu + \alpha_1$ and $E(Y_3) = E(Y_4) = \mu + \alpha_2$ (i) Examine whether α_1 is estimable. (ii) Find the best unbiased estimator of $(\alpha_1 - \alpha_2)$. (iii) Obtain the variance of the best unbiased estimator of $(\alpha_1 - \alpha_2)$. 12
- (c) If r_{12}, r_{13} and r_{23} are the simple correlation coefficients between X_1 and X_2 ; X_1 and X_3 ; X_2 and X_3 respectively then show that
- $$r_{12.3} = (r_{12} - r_{13} r_{23}) / (\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}).$$
- Also examine the validity of the statement “the range of the partial regression coefficient is $-\infty$ to $+\infty$.” 12
- (d) Distinguish between simple random sampling with and without replacement. Write down the expression for the number of samples and the probability that a specified individual is included in the sample under both the sampling schemes SRSWR and SRSWOR. 12
- (e) Derive the normal equations for estimating the parameters of a linear regression model by the method of least squares. 12
6. (a) Suppose that $X \sim$ follow the p variate normal distribution with co-variance matrix
- $$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \dots & \dots & \dots & \dots & \dots \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$
- Obtain the expression for the multiple correlation coefficient $R_{1.23\dots p}$. 20
- (b) Write down the model used for one way analysis of variance. The lifetimes (in hours) of samples taken from three different brands of batteries gave the following: $n_1=5, n_2=4, n_3=6, \bar{x}=40, \bar{y}=55, \bar{z}=60, \sum_{i=1}^5 (x_i - \bar{x})^2 = 400, \sum_{i=1}^4 (y_i - \bar{y})^2 = 350$ and $\sum_{i=1}^6 (z_i - \bar{z})^2 = 850$. Test the hypothesis that there is no difference in lifetimes for the three brands. Also write down the ANOVA ($F_{2, 12, .05} = 3.89$). 20
- (c) Suppose $X \sim (X_1, X_2, X_3)^T$ has co-variance matrix
- $$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$
- Obtain the first two principal components of X . 20
7. (a) Let $N=3, n=2$ and the possible distinct samples are $s_1=(1,2), s_2=(1,3)$ and $s_3=(2,3)$. If SRSWOR is adopted, write down $p(s_i)$ for $i=1,2,3$. If
- $$t = \{y_{1/2} + y_{2/2} \text{ if } s_1 \text{ occurs}; y_{1/2} + 2y_{3/3} \text{ if } s_2 \text{ occurs}; y_{2/2} + y_{3/3} \text{ if } s_3 \text{ occurs},\}$$
- examine whether t is unbiased for the population mean. Also find the expression for $v(t)$, if unbiased. 20

- (b) Bring out the relation between Hotelling's T^2 and Mahalanobis D^2 . Explain how T^2 can be interpreted as an extension of the student's t statistic. Describe the utility of Fisher's discriminant function in classification problems. 20
- (c) Discuss on the advantages of using a BIBD. For a BIBD with parameters (v, b, r, k, λ) , show that $\lambda(v-1)=r(k-1)$ and $b \geq v$. Also examine whether a BIBD is possible with $v=b=34$, $r=k=12$ and $\lambda=4$. 20
8. (a) What are inclusion probabilities? Use this concept to define the Horvitz-Thomson estimator. If α_k and α_{kl} denote the first and second order inclusion probabilities under an arbitrary ordered design of fixed size, show that (i) $\sum_1^N \alpha_k = v$ and (ii) $\sum \sum_{k \neq l}^N \alpha_{kl} = v(v-1)$ where $v=E(v(s))$ is the expected effective sample size. 20
- (b) If V_{ran} , V_{prop} and V_{opt} respectively denote the variance of the estimator of the population mean under SRSWOR, stratified sampling under proportional allocation and Neyman allocation, show that
- $$V_{\text{ran}} > V_{\text{prop}} > V_{\text{opt}} \quad 20$$
- (c) In a 3^3 factorial experiment, each replication is divided into 3 blocks of size 9 each and the effect ABC^2 is totally confounded. Explain the analysis of the experiment and set up the ANOVA table. 20