

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 sub-parts of the following:
 - (a) Given that A, B, C are events with A and C mutually exclusive, B and C independent and $P(A)=1/12$, $P(B)=1/3$, $P(C)=1/4$ & $P(A \cap B)=0$, calculate: (i) $P(A \cup B \cup C)$ (ii) $P(\bar{B} \cap \bar{C})$ [?] (iii) $P((A \cup B) \cap C)$ (iv) $P(A|B)$ 12
 - (b) State Inversion theorem. Find the distribution to which the characteristic function $\phi(t)=e^{-|t|}$ corresponds 12
 - (c) Independent trials consisting in rolling of two fair dice are performed. What is the probability that an outcome of 5 appears before an outcome of 7, where the outcome of a roll is the sum of the dice? 12
 - (d) A drug manufacturer wants to test the efficacy of a new medication. Denoting by θ the percentage of all the patients given the medication who recover from the disease he sets up $H_0: \theta=0.90$, $H_1: \theta=60$. His test statistic is x, the observed number of successes in $n=25$ trials and he will accept the null hypothesis if $x \geq 15$. Find the probability of type I and Type II errors denoted respectively by α and β . What happens if the acceptance region is changed to $x \geq 16$? 12
 - (e) If $E(T_1)=\theta=E(T_2)$, $\text{var}(T_1)=\sigma_1^2$, $\text{var}(T_2)=\sigma_2^2$ and T_1 and T_2 are independent, then show that $T=\lambda T_1+(1-\lambda)T_2$ is an unbiased estimator of θ for which the variance is minimized, when $\lambda=\sigma_2^2/(\sigma_1^2+\sigma_2^2)$. 12
 - (f) Explain run test. The following arrangement of defective (d) and non-defective (n) pieces produced in the given order by a certain machine was observed nnnnnn dddd nnnnnnnnnn dd nn dddd n dd nn Test for randomness at 0.01 level of significance. (Area under standard normal curve from 0 to 2.575 is 0.495) 12
2. (a) For any integer-valued random variable X, show that $\sum_{n=0}^{\infty} P(X \leq n) s^n = P(s)/(1-s)$ where $P(s)$ is the probability generating function of X. 20
 - (b) Let X, Y, Z be independent and uniformly distributed over (0, 1). Compute $P(X \geq YZ)$. 20
 - (c) Define various modes of convergence of a sequence of random variables. Prove that convergence almost surely implies convergence in probability. 20
3. (a) Giving example, define sufficient statistics and state Fisher-Neyman criterion for the existence of sufficient statistics. Let $X_1, X_2, \dots, X_n$ be a random sample from the distribution that has p.d.f.

$$f(x; \theta) = \begin{cases} \theta^x (1-\theta)^{1-x}, & x=0,1; 0 < \theta < 1; \\ 0, & \text{elsewhere} \end{cases}$$
20

Check whether $Y_1 = X_1 + X_2 + \dots + X_n$ is a sufficient statistic for θ .

(b) Describe Wilcoxon signed-rank test and derive the approximate distribution of the test statistic. 20

(c) Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\theta, 1)$ distribution, where θ is unknown. Show that there is no uniformly most powerful test of the simple hypothesis $H_0: \theta = \theta'$, where θ' is a fixed number against the alternative composite hypothesis $H_1: \theta \neq \theta'$ 20

4. (a) Discuss least squares method of estimation and the relation between least squares estimators and best linear unbiased estimators. 20

(b) Describe briefly the Wald's sequential probability ratio test. If for the choice $A = (1 - \beta)/\alpha, B = \beta/(1 - \alpha)$ the SPRT terminates with probability 1 and is of strength (α', β') , then prove that

$$\alpha' \leq \alpha/(1 - \beta), \beta' \leq \beta/(1 - \alpha) \text{ and } (\alpha' + \beta') \leq \alpha + \beta \quad 20$$

(c) Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from the uniform distribution having p.d.f. $f(x; \theta) = 1/\theta, 0 < x < \theta, 0 < \theta < \infty$ and zero elsewhere. Show that $4Y_1$ and $2Y_2$ are both unbiased statistics for θ . Find their variances. 20

SECTION 'B'

5. Answer any five sub-parts of the following:

(a) Describing briefly the concept of multiple regression, define multiple and partial correlations and obtain their formulae. 12

(b) If X is distributed as $N_5(\mu, \Sigma)$, find the distribution of $(X_2; X_4)$ 12

(c) In a simple random sample of size 200 (without replacement) from a population of size 1000, 75 persons happened to be smokers. Estimate the proportion of smokers and total number of smokers in the population and calculate 95% confidence interval for the proportion of smokers and also for the total number of smokers in the population. (Given that for 95% confidence probability, $t = 1.96$) 12

(d) If $X \sim N_p(\mu, \Sigma)$, prove that every linear function of components of X is univariate normal. 12

(e) Suppose that coefficient of variation of income of households in a region with 2000 households is 75%. How large a random sample of households is required for a margin of error $\pm 5\%$ with 95% as [?] confidence coefficient? (For 95% confidence probability, $t = 1.96$) 12

(f) What is a Latin square design? Give the analysis of an $m \times m$ Latin square design, with one missing value. 12

6. (a) Discuss the procedure of classification of an observation into one of two known multivariate normal populations. 20

(b) Describe curvilinear regression. Given that

$$\sum_1^8 X = 158, \sum_1^8 X^2 = 5530, \sum_1^8 \log y = 13.0312 \text{ and } \sum_1^8 X \log y = 212.1224,$$

fit the exponential curve $y = ab^X$ [?] applying the method of least squares. 20

(c) If $X_1, X_2, \dots, X_n$ have a joint normal distribution, show that a necessary and sufficient condition for one subset of random variables to be independent of the subset consisting of the remaining variables is that each covariance of a variable from one set and a variable from the other set be zero. 20

7. (a) Define ratio estimator. In srswor (N, n) , show that the mse of R is approximately given by

$$E(R - R)^2 = (1 - f)/n X^2 [S_y^2 + R^2 S_x^2 - 2R \rho S_y S_x]$$

(Retaining only the first term in an expansion) 20

(b) In a stratified population with two strata, the values of W_h, S_h and C_h are as follows: 20

h	W_h	S_h	C_h
1	0.35	15	4
2	0.65	25	8

If the cost function is of the form $C=c_1n_1+c_2n_2$, find the values of n_1/n and n_2/n that minimize C for a given value of $V(\bar{y}_{st})$, where the symbols have their usual meanings. Also find the sample sizes n_1 and n_2 for $V(\bar{y}_{st})=2$, ignoring f.p.c.

- (c) In double sampling for stratification, if the first sample is random (of size n'), the second sample is a stratified random subsample of the first of size $n_h=v_h n'_h$, where $0<v_h\leq 1$, and v_h are fixed, then show that

$$V(\bar{y}_{st})=S^2(1/n'-1/N)+\sum_{h=1}^L W_h S_h^2/n'(1/v_h-1)$$

Where S^2 is the population variance and other notations have their usual meanings.

20

8. (a) Discuss the two-way analysis of variance model giving its analysis and ANOVA table. 20
- (b) Explain confounding in factorial experiments. Differentiate between complete and partial confounding. Set up a statistical model for a two-factor factorial experiment in RBD and give the ANOVA table with expected mean squares. 20
- (c) What is a BIBD? Write down the parameters of a BIBD, and state and prove the relations that serve as a necessary condition for the existence of a BIBD. 20