STATISTICS

PAPER – I

Time Allowed: 3 Hours **Maximum Marks: 300**

Candidates should attempt Questions 1 and 5 which are compulsory, and any three of the remaining questions selecting at least one question from each Section.
Assume suitable data if considered necessary and indicate the same clearly.
Notations and symbols used are as usual.

SECTION A

(Probability and Statistical Inference)

1. Attempt any five sub-parts:

(a) Let X be an absolutely continuous random variable with p.d.f. f(x) and distribution function F(x). Show that

Y = F(x) has U(0, 1) distribution.

12

(b) Let X_1, X_2,……... be i.i.d. random variables with common p.d.f.

$$f(x)\begin{cases}\dfrac{1+\delta}{x^{2+\delta}}, & x \ge 1\left(\delta > 0\right)\\ 0, & otherwise\end{cases}$$

Show that the weak law of large numbers holds for the sequence.

12

(c) Let X he a random variable with p.m.f.

P(X = -2) 1/5, P(X = -1) = 1/6, P(X = 0) 1.5,

P(X = 1) 1/15, P(X = 2) = 11/30.

Obtain the distribution of $Y = X^2$.

12

(d) Let T_1 and T_2 be two unbiased estimators of θ having the same variance. Show that their correlation coefficient ρ cannot be smaller than $2e_\theta - 1$, where e_θ is the efficiency of each estimator.

12

(e) If a sufficient statistic T exists for the family

$\{f\theta : \theta \in \Theta\} = \Theta = \{\theta_0, \theta_1\}$, show that the Neyman-Pearson MP test is a function of T.

12

(f) Outline the Wilcoxon-Mann-Whitney test for the two-sample problem. Also, give the relevant large-sample test-statistic.

12

2. (a) If X is any non-negative integer valued variate and 'a' is any positive number, show that

$$P\left[P \geq a\right] \leq t^{-a} E\left[t^{x}\right], t > 1$$

and verify the inequality $P\left[X \geq 2\lambda\right] \leq \left(e/4\right)^{\lambda}$ when X follows Poisson distribution with parameter.

20

(b) Let F_n be a sequence of distribution functions defined by

$$F_n(x) = \begin{cases} 0, & x < 0 \\ 1 - \dfrac{1}{n}, & x \leq x < n \\ 1, & x \geq n. \end{cases}$$

Does X_n converge in distribution to X? If so, identify the distribution function of X. Also, examine whether

$$E\left[X_n^k\right] \rightarrow E\left[X^k\right]$$

For any k as $n \rightarrow \infty$.

20

(c) The characteristic function of a random variable X is $(q + pe^{it})^n$. Find the distribution of X.

20

3. (a) Let $X_1, X_2, \ldots\ldots\ldots X_n$ be a sample from p.d.f.

$$f(x, \theta) = \begin{cases} \dfrac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & x > 0, \theta > 0 \\ 0, & x \leq 0 \end{cases}$$

Show that $\sum_{i=1}^{n} X_i^2$ is a minimal sufficient

statistic for θ, but $\sum_{i=1}^{n} X_i$ is not sufficient.

20

(b) Let $X_1, X_2, \ldots\ldots X_n$, be a random sample from

$$f(x, \theta) = \begin{cases} \beta x^{\beta - 1}, & 0 < x < 1, \beta > 0 \\ 0, & otherwise \end{cases}$$

Find the MP test of size α for testing $H_0 : \beta = 1$ against $H_1 : \beta = 2$. Also obtain its power.

20

(c) Let the powder function of every test of

H_0 : $\theta = \Theta_0$ against H_1 : $\theta = \Theta_1$ = be continuous in θ. Show that a UMP a-similar test is UMPU provided that its size is a for testing H_0 against H_1

20

4. (a) Let $X_1, X_2, \ldots\ldots\ldots X_n$ be a random sample of size n from a distribution b(1, θ), $0 < \theta < 1$. Obtain the UMVU estimator of $\dfrac{\theta(1-\theta)}{n}$

20

(b) State and prove a result which shows that, compared to any other sequential test procedure, the SPRT has the least ASN.

20

(c) For each $\theta_0 \in \Omega$, let $A(\theta_0)$ be the acceptance region of the level-α UMP test for $H_0 : \theta = \theta_0$ against all alternatives. Show that $S_0(x)$ defined by

$$S_0(x) = \{\theta \mid x \in A_0(\theta)\}$$

is, for varying $x \in \mathcal{X}$ a family of UMA confidence sets at the level $1 - \alpha$.

20

SECTION B

(Linear Inference, Multivariate Analysis, Sampling Theory and Design of Experiments)

5. Attempt any five sub-parts

(a) If X (with p components) is distributed according to $N_p(\mu, \Sigma)$, then prove that $Z = DX$ is distributed according to $N_q(D\mu, D \Sigma D')$ where D is a $q \times p$ matrix of rank $q \le p$.

12

(b) Show that a linear function of observations belongs to the error space if and only if its coefficient vector is orthogonal to the columns of the matrix X in the usual linear model $Y = X\beta + \varepsilon$.

12

(c) A simple random sample of size $n = n_1 + n_2$ with mean $\bar{y}$ is drawn from a finite population, and a simple random sub-sample of size n_1 is drawn from it with mean $\bar{y}_1$. Show that

(i) $V(\bar{y}_1 - \bar{y}) = \left(\frac{1}{n_1} - \frac{1}{n}\right) S^2$.

(ii) $Cov(\bar{y}, \bar{y}_1 - \bar{y}) = 0$

6 + 6

(d) Bring out, giving an example, the relationship between a sampling scheme and a sampling design.

12

(e) Find a confounding arrangement of a 2^5 factorial experiment in 4 blocks such that no main effect or first order interaction is confounded.

12

(f) What is the purpose of MANOVA? Explain the one-way MANOVA model.

12

6. (a) For the model

$$y_1 = \beta_1 + \beta_2 + \varepsilon_1$$

$$y_2 = \beta_1 + \beta_3 + \varepsilon_2$$

$$y_3 = \beta_1 + \beta_2 + \varepsilon_3$$

show that $\lambda_1\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3$ is estimable if and only if $\lambda_1 = \lambda_2 + \lambda_3$.

20

(b) Explain the purpose of discriminant analysis. What is Fisher's discriminant function for two populations?

Give the allocation rule based on it.

20

(c) Outlining the purpose of factor analysis, describe the orthogonal factor model along with its covariance structure.

20

7. (a) (i) Describe a method of selecting a sample with probability proportional to size and with replacement.

10

(ii) Obtain the variance and its estimator for the Hansen-Hurwitz estimator.

10

(b) Show that the usual regression estimator is unbiased under a suitable model to be specified by you. Also, obtain the model variance of this estimator.

20

(c) Using Warner's randomized response technique, obtain the estimate of the population proportion relating to the sensitive characteristic.

20

8. (a) Give the analysis of an LSD if the yield of one of the plots is missing.

20

(b) Derive the analysis of variance table for 2-way classification with m (> I) observations per cell in the case of a random effects model.

20

(c) Construct a confounded 2^6 factorial experiment into 8 blocks of 8 treatments each by confounding the effects ABEF, ABCD and ACE. Also, identify the other effects which are confounded.

20

STATISTICS

PAPER - II

Time Allowed: 3 Hours ***Maximum Marks: 300***

SECTION A

1. Attempt any five sub-parts of the following:

(a) With reference to control charts, distinguish between (i) warning limits and action limits, and (ii) Type I and Type II errors.

12

(b) Explain the construction and use of X and R charts.

12

(c) Explain the concepts of 'availability' and 'maintainability'.

12

(d) Describe the different models used in Operations Research and state the general methods for solving them.

12

(e) What is queuing process? Obtain the probability generating function for the queue length of M / M / 1 queueing model and obtain the mean queue size.

12

(f) Consider the Markov chain $\{X_n ; n \geq 0)$ with the states (0, 1, 2) and with transition probability matrix

$$\begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/3 \end{bmatrix}$$

and the initial distribution

$P\{X_0 = i\} = 1/3; i = 0, 1, 2$

Obtain the value of:

$P\{X_0 = 1, X_2 = 2, X_1 = 1, X_0 = 2\}$

12

2. (a) Explain the statistical basis and construction of p and np charts.

25

(b) Explain with examples series and parallel systems. A series system of n components is such that each component has a failure distribution given by

$$F(x) = 1 - e^{-(x/\alpha)^{1/\sigma}}; x \geq 0.$$

Find the failure distribution of the system assuming that the components operate independently and n is large. What will be this distribution under parallel connection?

20

(c) Explain the meaning of censored and truncated samples. When are these used?

15

3. (a) Solve by Big-M method the following LP problem:

25

Maximize $Z = 3x_1 - x_2$

subject to $2x_1 + x_2 \geq 2$

$x_1 + 3x_2 \leq 3$

$x_2 \leq 4$

$x_1, x_2 \geq 0$

(b) What are replacement problems? Discuss such a problem in which items deteriorate with time and the value of money remains the same as per passage of time.

15

(c) What do you understand by 'recurrent' and 'transient' states of a discrete parameter Markov chain?

Show that a state j is recurrent if and only if

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$$

& if a state j is transient, then $p_{ij}^{(n)} \to 0$ for all i.

20

4. (a) What do you understand by sampling inspection plan for variables? Explain how it differences from the sampling inspection plan for attributes. State the use of Dodge-Roming tables.

15

(b) What is redundancy? Explain the use of redundancy in reliability improvement.

15

(c) Describe a Poisson process. Obtain the distribution of interarrival times in a Poisson process.

15

(d) What are SPSS packages? Explain their uses.

15

SECTION B

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

(a) Describe the method of moving averages for determining the trend component of a time series. How is trend eliminated?

12

(b) Define Fisher's ideal index number. Why is it called ideal?

12

(c) What are the principal publications which provide data on (i) industrial production, (ii) foreign exchange and (iii) mining activity?

12

(d) Define and interpret the following measures: 12

(i) Infant mortality rate

(ii) Expectation of life

(iii) Density of population

(e) Define a logistic curve and examine its suitability as a growth model. 12

(f) What is the 'validity' of a test? How is it measured? 12

6. (a) What are the different methods for determining the cyclical component of a time series ? Describe 'periodogram analyses'. 20

(b) Describe the essential steps for the construction of an index number of industrial production. 20

(c) Obtain the generalized least squares technique for estimating the coefficients of a linear model. Compare this with ordinary least squares estimation. 20

7. (a) Define crude and standardized death rates. Describe the direct and indirect methods for calculating standardized death rates. 20

(b) Explain the various columns of a life table giving the relationship between them. What is an abridged life table? 20

(c) Write a note on health surveys and hospital statistics 20

8. (a) What are the various official agencies responsible for data collection? Discuss their roles. 15

(b) Describe the present official statistical system in India dealing with collection of agricultural statistics. 15

(c) Why is it considered desirable to convert given raw scores into some standard scores? Define 'standardized scores' and 'normalized scores', and describe how these are derived. 15

(d) Describe any two methods commonly used for computing the 'reliability' coefficient of a test. What is the effect of lengthening of a test on its reliability? 15