STATISTICS

PAPER—IV

Time Allowed : Three Hours

Maximum Marks : 200

QUESTION PAPER SPECIFIC INSTRUCTIONS

Please read each of the following instructions carefully before attempting questions

There are FOURTEEN questions divided under SEVEN Sections.

Candidate has to choose any TWO Sections and attempt the questions therein. All the Sections carry equal marks.

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.

Diagrams/Figures, wherever required, shall be drawn in the space provided for answering the question itself.

Unless otherwise mentioned, symbols and notations have their usual standard meanings.

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 (QCA) Booklet must be clearly struck off.

Answers must be written in ENGLISH only.

SECTION—A

(Operations Research and Reliability)

1. (a) Maximize $Z = x_1 + 5x_2$

subject to

$$3x_1 + 4x_2 \le 6$$
$$x_1 + 3x_2 \ge 2$$
$$x_1,\ x_2 \ge 0$$

using Simplex procedure. 10

(b) (i) Describe dynamic programming problem and discuss its applications.

(ii) State Bellman's optimality criteria.

(iii) Minimize $Z = y_1^2 + y_2^2 + y_3^2$

subject to

$$y_1 + y_2 + y_3 \ge 18$$
$$y_1,\ y_2,\ y_3 \ge 0$$

10

(c) For two players A and B, the payoff matrix is given below :

		Player B				
		B_1	B_2	B_3	B_4	B_5
Player A	A_1	2	4	3	8	4
	A_2	5	6	3	7	8
	A_3	6	7	9	8	7
	A_4	4	2	8	4	3

Using dominance property, find the value of the game and optimal strategies of the players. Check whether strategies adopted by player A are mixed or pure. 10

(d) There are three dentists in a dental clinic. On an average, it takes 20 minutes on one patient and service times follow exponential distribution. Patients arrive according to Poisson process with mean 6/hour and queue discipline followed is FIFO (FCFS). Calculate the following :

(i) Expected number of patients in the queue

(ii) Expected amount of time spent by a patient in the clinic

(iii) Percentage of idle time of any dentist

(iv) Probability that all dentists are busy 10

(e) (i) Describe series and parallel systems made up of n components. Give their reliability functions. Consider a series system consisting of two independent components, where the lifetime of first component follows $\exp(\lambda_1)$ and the lifetime of second component follows $\exp(\lambda_2)$. Check whether IFR (Increasing Failure Rate) property holds for the system.

(ii) Show that

$$\overline{F} \text{ is IFRA} \Rightarrow \overline{F} \text{ is NBU} \Rightarrow \overline{F} \text{ is NBUE}$$

where

IFRA : Increasing Failure Rate Average

NBU : New Better than Used

NBUE : New Better than Used in Expectation

and $\overline{F}(t) = 1 - F(t)$ denotes the survival function of lifetime of a component or system. 10

2. Answer any *two* of the following :

(a) (i) The following table shows the costs incurred by assigning five jobs to five workers A, B, C, D and E :

Workers	*Jobs* 1	2	3	4	5
A	10	5	13	15	16
B	3	9	18	13	6
C	10	7	2	2	2
D	7	11	9	7	12
E	7	9	10	4	12

For the above table, find the optimal solution and hence the minimum cost by using an appropriate method of assignment. 10

(ii) Describe algorithm for solving '3 machines n jobs' sequencing algorithm. For the following table, determine the optimal sequence of performing jobs :

Jobs	*Processing time* (in hours)		
	Machine A	*Machine B*	*Machine C*
1	3	3	5
2	8	4	8
3	7	2	10
4	5	1	7
5	2	5	6

15

(b) (i) For a maximization problem, let the optimum Simplex table be as given below :

Basic variables	C_B	X_B	x_1	x_2	x_3	x_4	x_5	x_6
x_2	5	$\frac{50}{41}$	0	1	0	$\frac{15}{41}$	$\frac{8}{41}$	$-\frac{10}{41}$
x_3	4	$\frac{62}{41}$	0	0	1	$-\frac{6}{41}$	$\frac{5}{41}$	$\frac{4}{41}$
x_1	3	$\frac{89}{41}$	1	0	0	$-\frac{2}{41}$	$-\frac{12}{41}$	$\frac{15}{41}$
	$Z_j - C_j$		0	0	0	$\frac{45}{41}$	$\frac{24}{41}$	$\frac{11}{41}$

Find the ranges in which b_1, b_2 and b_3 can lie without affecting the optimality of the solution. 8

(ii) Discuss applications of Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM). A project network with six activities has the following precedence constraints :

- *A* precedes *D*
- *A* and *B* precede *C*
- *C* and *D* precede *F*
- *E* precedes *F*

Draw the network diagram. 7

(iii) A new car costs ₹ 3,50,000. The resale value of the car at the end of a year is 85% of previous year value. During the first year, operation and maintenance cost is ₹ 25,000 and it increases by 15% every year. The minimum resale value of the car can be ₹ 80,000. Determine as to when the car should be replaced so as to minimize average annual cost (ignoring value of money). 10

(c) (i) Three products I, II and III are produced by a shop in lots. The warehouse of the shop has total floor area 3000 sq. metres. The following information is available :

Items	I	II	III
Annual demand (units/year)	600	300	700
Cost per unit (in ₹)	50	30	80
Setup cost per lot (in ₹)	1,000	500	1,200
Floor area required (in sq. metres)	7	5	13

If no stockouts are allowed and warehouse capacity cannot be exceeded, find the optimal lot size for each item. 10

(ii) Suppose that n items are put on test and the test is terminated after all the items have failed. Let $(X_1, X_2, \ldots, X_n)$ denote the random failure times and suppose that failure times are distributed as gamma with parameters α and β. Find the maximum likelihood estimators of α and β. 10

(iii) A manufacturer wishes to promote a new brand of lightbulbs in the market. A random sample of 12 bulbs is taken and their failure times (in hours) are recorded as 130, 190, 200, 352, 473, 592, 612, 709, 805, 888, 915 and 989. The distribution of the failure times is assumed to be exponential with probability density function as

$$f(x \mid \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right), \quad x \geq 0, \ \sigma > 0$$

If a customer buys a bulb of new brand, estimate the probability that it will survive for at least 650 hours. 5

(d) Shipments have to go from three sources to four destinations as per the information given in the following table for the incurred costs :

		Destinations				*Availability*
		D_1	D_2	D_3	D_4	
	S_1	5	2	4	3	22
Sources	S_2	4	8	1	6	15
	S_3	4	6	7	5	8
Demand		7	12	17	9	

Using MODI method, solve the problem so that the total cost of transportation is minimum. 25

SECTION—B

(Demography and Vital Statistics)

3. (a) Write the details of National Census undertaken in India. 10

(b) List at least 10 items on which data is to be collected by census officials regarding the House-listing and Housing Census Schedule in connection with the Census of India, 2021. 10

(c) List the assumptions of life table and importance of life table. 10

(d) Given that $p_x = 0 \cdot 99$, $p_{x-1} = 0 \cdot 985$, ${}_3p_{x+1} = 0 \cdot 95$ and $q_{x+3} = 0 \cdot 02$. Calculate (i) p_{x+3}, (ii) ${}_2p_x$, (iii) ${}_2p_{x+1}$ and (iv) ${}_3p_x$. 10

(e) What is migration? Define and discuss net migration. For the following data, compute net migration in 5 years in that city :

Population of the city in 2005	1406026
Population of the city in 2010	1620140
Number of births in the city during 2005 to 2010	401240
Number of deaths in the city during 2005 to 2010	225920

10

4. Answer any *two* of the following :

(a) (i) Discuss about the data sources and data variety used by demographers. 10

(ii) The mortality function of a newborn following Gompertz law is $\mu(x) = 0 \cdot 001(1 \cdot 1)^x$. Then find—

(1) the probability that the person dies before reaching the age 30;

(2) the probability that the person lives more than 60 years;

(3) the probability of a person aged 20 dies within the age of 40;

(4) the mortality rate function of a person aged 25. 15

(b) (i) Complete the life table with usual notations :

Age x	q_x	l_x	d_x	L_x	T_x	e_x
55	—	7940	525	—	—	—
56	—	—	520	—	—	—
57	—	—	510	505	11170	—

15

(ii) A population count at time t, $p(t)$ is given by the expression

$$p(t) = \frac{\frac{1}{A}}{1 + \frac{B}{A} e^{-ut}}, \quad t \geq 0$$

where $p(0) = 46687$ thousand, $p(\infty) = 245000$ thousand and $u = 0 \cdot 023$. Determine the populations at 10, 20 and 30. 10

(c) *(i)* The following data gives the number of women, child-bearing ages and births for the respective age groups for a city. Assume that the ratio of male to female children is 16 : 15.

Age group	*Births in the age group*	*Female population*
15–19	760	15000
20–24	2300	12000
25–29	2000	9690
30–34	1300	8900
35–39	570	7600
40–44	160	6700
45–49	20	4500

(1) Calculate age-specific fertility rate for age groups 25–29, 35–39 and 45–49.

(2) Calculate total fertility rate.

(3) Compute gross reproduction rate. 15

(ii) Discuss stable population theory. What are the consequences of this to the population? How is stationary population connected to stable population? 10

(d) *(i)* Given a data set for $3n$ years, where n is a sufficient large integer. Discuss Yale's method of fitting logistic curve for N_t, population size at time t, with usual notations, where

$$N_t = \frac{K}{1+ce^{-rt}}$$ 15

(ii) Give the details of UN model for life tables. What are its drawbacks? 10

SECTION—C

(Survival Analysis and Clinical Trials)

5. *(a)* Distinguish between time-censored samples and failure-censored samples. Assume that the lifetime of a device follows the Weibull distribution with survival function

$$R(t) = \exp(-t^p/\sigma), \quad t > 0, \quad p \text{ known}$$

Obtain the likelihood function and m.l.e. of σ under failure-censored samples. Also give an estimate of the survival function. 10

(b) What is Cox proportional hazards model? Why is this model called semi-parametric? Write down the model in terms of the survival functions. Also obtain the estimate of the baseline hazard function. 10

(c) Suppose that the time to death X has an exponential distribution with hazard rate λ and that the right-censored time C is exponential with hazard rate θ. Let $T = \min(X, C)$ and

$$\delta = \begin{cases} 1 & \text{when } X \le C \\ 0 & \text{when } X > C \end{cases}$$

Assume X and C are independent. Find *(i)* $P(\delta = 1)$ and *(ii)* the distribution of T. 10

(d) Explain competing risk models. Give an illustrative example. Suppose that there are k competing risks and that the potential times are independent with survival function $S_i(t)$; $i = 1, 2, \ldots, k$. Show that the cause-specific hazard rate is same as the hazard rate of X_i. 10

(e) Explain the log-rank test for the two-sample problem, stating the assumptions involved. 10

6. Answer any *two* of the following :

(a) Define the Kaplan-Meier estimator for the survival function. Also write down the Greenwood's formula for the variance of the estimator. Also give an estimate for the hazard rate function. The times to failure of 25 units put on a life test are 17, 9, 9, 26, 24, 11, 13, 12, 12, 15, 22, 15, 9, 28, 68, 35, 30, 32, 39, 88, 50, 68, 9, 21 and 41. Obtain non-parametric estimates of reliability function and failure rate function. 25

(b) *(i)* Describe the objectives of clinical trials from the point of view of clinical practice and public health. 10

(ii) What are the different clinical trial designs? Distinguish between cross-sectional designs and longitudinal designs. Also describe the working of a factorial design. 15

(c) Distinguish between censoring and truncation. Explain right-censoring and left censoring giving two illustrations for each. X is a non-negative continuous random variable and $t, t_1, t_2 > 0$. Write down the probability distribution associated with $X \mid X > t$, $X \mid X < t$ and $X \mid t_1 < X < t_2$. 25

(d) Explain the four major phases of clinical trials detailing the objectives of each phase. 25

SECTION—D

(Quality Control)

7. *(a)* Discuss about the following statement :

Applying a control chart is equivalent to repeatedly performing a test of hypothesis.

A voltage stabilizer manufacturer checks the quality of his product daily for 15 days and estimates the fraction of non-conforming units as follows :

0·10, 0·20, 0·06, 0·04, 0·16, 0·02, 0·08, 0·06, 0·02, 0·16, 0·12, 0·14, 0·08, 0·10 and 0·06.

Obtain the three-sigma control limits for fraction defectives. 10

(b) What are *(i)* specification limits and *(ii)* control limits? 25 samples each of size 5 were drawn from a production process and measurements on a characteristic were recorded. It was seen that $\overline{\overline{X}} = 1 \cdot 5056$ and $\overline{R} = 0 \cdot 32521$ (with usual notations). Find the control limits for the $\overline{X}$ chart and R chart ($D_3 = 0$, $D_4 = 2 \cdot 114$, $A_2 = 0 \cdot 577$). If one or more points fall outside the control limits, what will be your next strategy? 10

(c) Explain how the OC curve reveals the ability of a sampling plan to distinguish between good and bad lots. How will you compare two sampling plans using the OC curve? Also interpret the OC curve for control charts with special reference to the $\overline{X}$ chart. 10

(d) What do you understand by process capability studies? Discuss the reasons for lack of process capability in the case of two-sided tolerance specification. Explain how the process capability index C_p can be used to measure process capability. 10

(e) Describe the working of a single sampling plan SSP (N, n, c). How will you design a single sampling plan by using the producer's risk (α) and consumer's risk (β)? 10

8. Answer any *two* of the following :

(a) When will you say that a production process is in statistical control? How can you decide whether a process is out of control using a control chart? Discuss the relevance of the three-sigma limits in the construction of a control chart. Distinguish between control charts for variables and control charts for attributes giving an example of each, with the control limits. 25

(b) Explain the notions of *(i)* AQL, *(ii)* LTPD, *(iii)* AOQ, *(iv)* AOQL, *(v)* ATI and *(vi)* ASN with reference to a sampling inspection plan. 25

(c) What are CUSUM charts? Explain how a decision is arrived at using CUSUM charts. How is the V-mask constructed? 25

(d) (i) Describe the assumptions involved in the computation of the OC function using (1) the hypergeometric distribution and (2) the binomial distribution in a single sampling plan. Illustrate by examples. 12

(ii) What are exponentially weighted moving average control charts for monitoring the process mean? Write down the expressions for the control limits. 13

SECTION—E

(Multivariate Analysis)

9. (a) Assuming

$$\frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2+y^2+4x-6y+13)\right]$$

to be a bivariate normal density of the form

$$\frac{\sqrt{A}}{(2\pi)^{p/2}}\exp\left[-\frac{1}{2}(x-b)'A(x-b)\right]$$

find the following :

(i) A and b

(ii) μ_x, μ_y, σ_x, σ_y and ρ_{xy} 10

(b) Let Z_1 and Z_2 follow $N(0, 1)$, defining

$$X = \mu_x + \sigma_x Z_1$$

$$Y = \mu_y + \sigma_y [\rho Z_1 + \sqrt{(1-\rho^2)}\, Z_2]$$

where μ_x, μ_y are the means; σ_x, σ_y are the standard deviations and ρ_{xy} is the correlation coefficient. Evaluate the following :

(i) $E(X|Y)$

(ii) $V(X|Y)$ 10

(c) Let X be distributed as $N(\mu, \Sigma)$, where

$$\mu' = [3 \quad -2 \quad 0]$$

$$\Sigma = \begin{bmatrix} 5 & 0 & -3 \\ 0 & 9 & 0 \\ -3 & 0 & 2 \end{bmatrix}$$

(i) Are X_2 and $2X_1 - X_3$ independent? Justify your answer.

(ii) Find the distribution of

$$\begin{pmatrix} X_1 - 3X_3 \\ 2X_1 + X_2 \end{pmatrix}$$ 10

(d) If the random vector $X = (X_1 \;\; X_2)'$, whose covariance matrix Σ is given by

$$\begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix}$$

then—

(i) show that Σ is a positive definite matrix;

(ii) determine the correlation matrix R;

(iii) determine the principal components using R;

(iv) find the proportion of total variance of X explained by the first component. 10

(e) (i) Let

$$X_1 = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} \quad X_2 = \begin{pmatrix} 4 \\ 1 \\ 6 \end{pmatrix} \quad X_3 = \begin{pmatrix} 4 \\ 0 \\ 4 \end{pmatrix}$$

be a sample of size 3 from a trivariate normal population. Examine whether it is degenerate by computing sample covariance matrix or otherwise.

(ii) Given that

$$n_{1c} = 10 = n_{2c}$$

$$n_{1m} = 2 = n_{2m}$$

are the elements of a confusion matrix, where

n_{ic} = number of items correctly classified

n_{im} = number of items misclassified

Compute the apparent error rate. 10

10. Answer any *two* of the following :

(a) Let $X_{p \times n}$ be a data matrix from $N_p(\mu, \Sigma)$. Write a graphical procedure to test whether the data follows multivariate normal or not. 25

(b) Consider the covariance matrix

$$\Sigma = \begin{bmatrix} 19 & 30 & 02 & 12 \\ 30 & 57 & 05 & 23 \\ 02 & 05 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}$$

Decompose the variances into communalities and specific variances. 25

(c) In a study of poverty, crime and deterrence, the sample correlation matrix is shown below for certain crime summary statistics :

$$R = \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline R_{21} & R_{22} \end{array}\right] = \left[\begin{array}{cc|cc} 1\cdot0 & 0\cdot4 & 0\cdot5 & 0\cdot6 \\ 0\cdot4 & 1\cdot0 & 0\cdot3 & 0\cdot4 \\ \hline 0\cdot5 & 0\cdot3 & 1\cdot0 & 0\cdot2 \\ 0\cdot6 & 0\cdot4 & 0\cdot2 & 1\cdot0 \end{array}\right]$$

The variables are—

$X_1^{(1)}$ = non-primary homicides

$X_2^{(1)}$ = primary homicides

$X_1^{(2)}$ = severity of punishment

$X_2^{(2)}$ = certainty of punishment

Find the first sample canonical variate pair (U, V) and their sample canonical correlation. 25

(d) Consider the following independent samples from three bivariate normal populations :

Sample–1 $\left[\begin{pmatrix} 9 \\ 3 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} \begin{pmatrix} 9 \\ 7 \end{pmatrix}\right]$

Sample–2 $\left[\begin{pmatrix} 0 \\ 4 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix}\right]$

Sample–3 $\left[\begin{pmatrix} 3 \\ 8 \end{pmatrix} \begin{pmatrix} 1 \\ 9 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}\right]$

Conduct MANOVA, given LOS $(\alpha) = 0\cdot01$; $F_{4,\ 2}$ (at $\alpha = 0\cdot01) = 7\cdot01$. 25

SECTION—F

(Design and Analysis of Experiments)

11. (a) What are randomization and replication in experimental design? How are these principles useful? 10

(b) Suppose $y_1, y_2, \dots, y_n$ are independently distributed following a normal distribution with mean

$$E(y_i) = \alpha + \sum_{j=1}^{r} \beta_j x_{ij}$$

and variance $V(y_i) = \sigma^2$, $i = 1, 2, \dots, n$. Then prove that maximum likelihood estimator is same as the least square estimator. 10

(c) What are the two important hypotheses tested in RBD? Discuss the procedure to carry out this for a given data. 10

(d) What is meant by 'analysis of covariance'? Write down the model for analysis of covariance in case of CRD. State the hypothesis that will be tested. Also write the ANCOVA table. 10

(e) Write a model for split-plot experiment with whole plots in an RBD. Also write the ANOVA table. 10

12. Answer any *two* of the following :

(a) Rats were given one of four different diets at random, and the response measure was liver weight as a percentage of body weight. The responses were—

Treatments

T_1	T_2	T_3	T_4
3·52	3·47	3·54	3·74
3·36	3·73	3·52	3·83
3·57	3·38	3·61	3·87
4·19	3·87	3·76	4·08
3·88	3·69	3·65	4·31
3·76	3·51	3·51	3·98
3·94	3·35	3·86	
3·64	3·71		

(i) Compute the treatment effects.

(ii) Compute the analysis of variance table for these data. What would you conclude about the four diets?

Certain F critical values at 5% are—

$F\,(0{\cdot}05,\ 4,\ 28) = 4{\cdot}675465$

$F\,(0{\cdot}05,\ 4,\ 24) = 2{\cdot}776289$

$F\,(0{\cdot}05,\ 3,\ 25) = 2{\cdot}991241$

25

(b) (i) With reference to RBD, split the total sum of squares into three components as carried out in the analysis of variance. Write the three mean sum of squares and its null distributions and expected value. 15

(ii) Suppose two observations are missing in an RBD. Derive the solution process to obtain the estimates of missing values. 10

(c) *(i)* Write down the model for Latin Square Design (LSD). Also write its ANOVA table. There are four treatments A, B, C and D to be used in an LSD. Construct an appropriate layout. 15

(ii) What is non-orthogonal data? What is the impact of this in regular analysis of data? How could such analysis be misleading? 10

(d) Taking an example of 2^3-factorial experiment, with A, B, C treatments, describe how average effect of A at various levels can be obtained. Also write how interaction effect is calculated. Present a form of ANOVA for this. 25

SECTION—G

(Computing with C and R)

13. Write C programs for questions *(a)* to *(c)* and R-code for questions *(d)* and *(e)* :

(a) To compute binomial (5, 0·3) probabilities and print the following table :

X	*Probabilities* (to be computed)
0	p_0
1	p_1
2	p_2
3	p_3
4	p_4
5	p_5

10

(b) To evaluate Hotelling's T^2 statistic to test

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad \mu \neq \mu_0$$

and to print T^2 value 10

(c) To calculate quartile deviation (QD) of n observations and print Q_1, Q_3 and QD 10

(d) To compute multiple and partial correlation coefficients for a trivariate sample and print the results 10

(e) To calculate geometric and harmonic means 10

14. Answer any *two* of the following [For questions *(a)* to *(c)*, write C programs and to print the results] :

(a) For decomposing the given symmetric matrix A into $L'L$, i.e., $A = L'L$, where L is a lower triangular matrix 25

(b) *(i)* To isolate the root and *(ii)* to find a positive root of

$$f(x) = x^3 - 0\cdot 2x^2 - 0\cdot 2x - 1\cdot 2 = 0$$

using method of chords and print the results 25

(c) To compute two types of generalized variances based on the sample covariance matrix 25

(d) Write R-code to find the value of an $m \times n$ two-person zero-sum strictly determined game $G_{m\times n}$. 25

★ ★ ★