IES/ISS Exam,2021

SDT-Q-STT

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. Graph paper can be found at the end of the booklet.

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 Booklet must be clearly struck off.

*Answers must be written in **ENGLISH** only.*

SECTION A

(Operations Research and Reliability)

Q1. (a) Find the sequence that minimises the total elapsed time required to complete the following tasks : *10*

Tasks	A	B	C	D	E	F	G
Time on Machine I	3	8	7	4	9	8	7
Time on Machine II	4	3	2	5	1	4	3
Time on Machine III	6	7	5	11	5	6	12

(b) For the following pay-off table, transform the zero-sum game into an equivalent Linear Programming Problem and solve the game by Simplex Method : *10*

Player B

$$\text{Player A}\quad \begin{array}{c|ccc} & B_1 & B_2 & B_3 \\ \hline A_1 & 1 & -1 & 3 \\ A_2 & 3 & 5 & -3 \\ A_3 & 6 & 2 & -2 \end{array}$$

(c) A supermarket has two girls ringing up sales at the counters. The service time for each counter is exponential with mean 4 minutes, and people arrive in a Poisson fashion at the rate of 10 an hour.

Calculate :

(i) Probability of having to wait for service.

(ii) Expected percentage of idle time for each girl.

(iii) If the customer has to wait, what is the expected length of his waiting time ? *10*

(d) A manufacturing company purchases 9000 parts of a machine for its annual requirements, ordering one month's requirement at a time. Each part cost ₹ 20. The ordering cost per order is ₹ 15 and the carrying charges are 15 percent of the average inventory per year. You have been assigned to suggest a more economical purchasing policy for the company. What advice would you offer and how much would it save the company per year ? *10*

(e) Describe Project Evaluation and Review Technique (PERT) and its procedure stepwise.

A small project is composed of activities whose time estimates are listed in the following table. Activities are identified by their beginning (i) and ending (j) node numbers.

Activity	Estimated Duration (in weeks)		
i – j	Optimistic	Most Likely	Pessimistic
1 – 2	1	1	7
1 – 3	1	4	7
1 – 4	2	2	8
2 – 5	1	1	1
3 – 5	2	5	14
4 – 6	2	5	8
5 – 6	3	6	15

(i) Draw the Project Network.

(ii) Find the expected duration and variance of each activity.

(iii) What is the expected projected length ?

(iv) Calculate the Variance and Standard deviation of the project length. *10*

Q2. Answer any ***two*** questions of the following :

(a) (i) A marketing manager has 5 salesmen and 5 sales districts. Considering the capabilities of the salesmen and the nature of the districts, the marketing manager estimates that sales per month (in hundred rupees) for each salesman in each district would be as follows :

Salesmen

		A	B	C	D	E
	1	32	38	40	28	40
	2	40	24	28	21	36
Districts	3	41	27	33	30	37
	4	22	38	41	36	36
	5	29	41	40	35	39

Find the assignment of salesmen to districts that will result in maximum sales. *10*

(ii) Describe Algorithm for solving n-jobs through 2-machines. The following table gives machine processing times for the five jobs on the two machines A and B :

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

Determine the optimal sequence of jobs that minimizes the total elapsed time. Also find the idle time of Machine A and Machine B. *15*

(b) Find the Basic Feasible Solution of the following Transportation Problem by Vogel's Approximation Method :

		D_1	D_2	D_3	D_4	D_5	Available ↓
Origins	O_1	4	3	1	2	6	80
	O_2	5	2	3	4	5	60
	O_3	3	5	6	3	2	40
	O_4	2	4	4	5	3	20
Demand →		60	60	30	40	10	

Using MODI Method, also find the optimum transportation plan. *25*

(c) (i) A commodity is to be supplied at a constant rate of 200 units per day. Supplies of any amounts can be had at any required time, but each ordering costs ₹ 50. Cost of holding in inventory is ₹ 2 per unit per day while delay in the supply of the items induces a penalty of ₹ 10 per unit per delay of 1 day. Find the optimal policy (q, t) where 't' is the reorder cycle period and 'q' is the inventory level after reorder. What would be the best policy if the penalty cost becomes infinite ? *10*

(ii) For $(M|M|1) : (\infty|FIFO)$ queue system, derive the steady state distribution of the queue length. Also obtain the probability density function of waiting time excluding service time distribution. *15*

(d) (i) A recorded lifetime (in hours) for identical items are as follows :

10·2, 89·6, 54·0, 96·0, 23·3, 30·4, 41·2, 0·8, 3·6, 28·0, 73·2, 31·6

Assuming lifetime has Gamma Distribution with parameters λ and ν, find the estimate of the failure rate λ. *10*

(ii) Consider an item with failure rate function given by

$$h(t) = \alpha\lambda\,(\lambda t)^{\alpha-1};\ \lambda > 0.$$

Obtain the Reliability Function, Conditional Survival Function $P[T > x + t \mid T > t]$, Mean Time To Failure (MTTF) and Mean Residual Lifetime. *15*

SECTION B

(Demography and Vital Statistics)

Q3. (a) Discuss on hospital records and ad hoc surveys in relation to vital statistics information. *10*

(b) Discuss internal migrations, rural-urban migrations and international migrations. *10*

(c) Distinguish between the terms fecundity and fertility. Mention the various measures of fertility. Also discuss their relative merits and demerits. *10*

(d) Given that the complete expectations of life at ages 25 and 26 for a specific group are 22·08 and 21·93 years respectively. The number of people living at age 25 is 45324. Find the number that attains age 26. *10*

(e) Describe intercensal and post-censal estimates of population growth assuming linear growth and exponential growth in mathematical method. *10*

Q4. Answer any ***two*** of the following :

(a) Write a critical note on the salient features of Indian Censuses 1991 and 2001. *25*

(b) Explain Pearl and Reed method of fitting logistic curve for population projection. *25*

(c) Calculate

(i) General Fertility Rate (GFR),

(ii) Specific Fertility Rate (SFR),

(iii) Total Fertility Rate (TFR), and

(iv) Gross Reproduction Rate (GRR)

from the data given below : *25*

Age group of child bearing females	Number of Women ('000)	Total Births
15 – 19	16·0	260
20 – 24	16·4	2244
25 – 29	15·8	1894
30 – 34	15·2	1320
35 – 39	14·8	916
40 – 44	15·0	280
45 – 49	14·5	145

Assume that the proportion of female births is 46·2 percent.

(d) Fill in the blanks of the following table which are marked with question marks (?) : 25

Age x	l_x	d_x	q_x	p_x	L_x	e_x^0
20	693435	?	?	?	?	35081126
21	690673	–	–	–	–	?

SECTION C
(Survival Analysis and Clinical Trial)

Q5. (a) Define the survival function and the hazard function. Derive these functions for the Weibull distribution. *10*

(b) Define Kaplan-Meier estimator of the survival function using standard notation. *10*

(c) Define the proportional hazards model in terms of the survival function. Explain why it is called a semi-parametric model. Derive the estimator of its baseline hazard function. *10*

(d) Make a critical comparison between cross-sectional and longitudinal designs. Describe their relative advantages and disadvantages. *10*

(e) Discuss the need of clinical trials and explain the ethical considerations involved in clinical trials. *10*

Q6. Answer any ***two*** of the following :

(a) Consider a survival distribution with constant hazard $\lambda = 0{\cdot}07$ on $(0, 5]$ and $\lambda = 0{\cdot}14$ on $(5, \infty)$. Plot this hazard function and the corresponding survival function on $(0, 10)$.

What is the median survival time ? *25*

(b) Define time and order censoring. If the lifetime distribution is exponential with parameter λ, derive the maximum likelihood estimator of λ under these two censoring schemes.

Also, derive Fisher information in the two cases. *25*

(c) Describe the four phases of clinical trials explaining their respective objectives in detail. *25*

(d) Explain the difference between censoring and truncation. Also, explain right-censoring and left-censoring with the help of two examples each. Suppose X is a non-negative random variable and t, t_1, t_2 are positive real constants, identify the probability distributions of X conditional on

(i) $X > t$,

(ii) $X < t$, and

(iii) $t_1 < X < t_2$

respectively. *25*

SECTION D
(Quality Control)

Q7. (a) (i) What do you mean by Statistical Quality Control (SQC) ? Discuss its need and utility.

(ii) Distinguish between process control and product control. Discuss the situations where they are used. *5+5=10*

(b) Discuss various sources of assignable causes and random causes of variations. Also, state how they are detected in manufacturing process. *10*

(c) What is control chart ? Explain the basic principles underlying the control charts. Discuss the role of control charts in manufacturing process. *10*

(d) Explain the construction of control chart for $\bar{X}$ and R. Also, mention the criterion for detecting 'Lack of Control' in $\bar{X}$ and R charts. *10*

(e) What is an acceptance sampling plan ? Discuss a single sampling plan (n, c), where the sampling is carried out using a binomial model. Find P_a if $n = 10$, $c = 3$ and $p = 0{\cdot}05$. *10*

Q8. Answer any ***two*** of the following :

(a) Evaluate the ARL to detect a shift in the process average (no change in process variation) by one unit of standard deviation in the higher side. Assume the use of $\bar{X}$-chart with subgroup size 5 and probability of not detecting the shift is at most 0·05. *25*

(b) What is Average Sample Number (ASN) ? Explain the method of calculation for single sampling and double sampling plan and also discuss the relative merits and demerits. Why is ASN calculated ? *25*

(c) An improvement in the process has resulted in increase of Cpk from 0·06 to 0·09. Estimate the reduction in % of non-conforming products. Assume the following : Process average has not changed, Cp = Cpk and process is in statistical control. *25*

(d) (i) Describe the construction of p-chart (control chart for fraction defective). Discuss the cases when standards are specified and standards are not specified. *10*

(ii) The following are the figures of defectives in 22 lots each containing 2000 rubber belts :

425, 430, 216, 341, 225, 322, 280, 306, 337, 305, 356,
402, 216, 264, 126, 409, 193, 326, 280, 389, 451, 420

Find control limits. *15*

SECTION E

(Multivariate Analysis)

Q9. (a) Define bivariate normal distribution and obtain its

(i) Characteristic function, and

(ii) Marginal distribution of X_1. *10*

(b) Let $\mathbf{X} \sim N_3(\mu, \Sigma)$ with $\mu' = (2, 3, -1)$ and $\sum = \begin{bmatrix} 7 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 2 \end{bmatrix}$.

(i) Obtain the conditional distribution of X_1 and X_2 given $X_3 = 2$.

(ii) Find the partial correlation coefficient between X_1 and X_2 given X_3. *10*

(c) Define Hotelling's T^2 and mention its applications. Show that

$$T^2(p, m) = \frac{mp}{m-p+1} \cdot F_{p,\, m-p+1}$$ *10*

(d) Let $\mathbf{X} \sim N_3(\mu, \Sigma)$ with $\mu' = (-3, 1, 4)$ and $\sum = \begin{bmatrix} 1 & -2 & 0 \\ 2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

Which of the following random variables are independent ? Explain. *10*

(i) X_1 and X_2

(ii) X_2 and X_3

(iii) (X_1, X_2) and X_3

(e) Define Wishart distribution and obtain its characteristic function. Also establish its additive property. *10*

Q10. Answer any ***two*** of the following :

(a) Discuss the procedure for testing the null hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ regarding the specified mean vector of the multivariate normal distribution $N_P(\mu, \Sigma)$.

Let the data matrix for a random sample of size n = 3 from a bivariate normal distribution be

$$X = \begin{bmatrix} 6 & 10 & 8 \\ 9 & 6 & 3 \end{bmatrix}.$$

Evaluate the observed value of T^2 for $\mu_0' = (9, 5)$. What is the sampling distribution of T^2 in this case ? *25*

(b) Define canonical correlation coefficient and canonical variates. In the usual notations, show that the canonical correlations are the solution of the determinantal equation

$$\begin{vmatrix} -\lambda\,\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\,\Sigma_{22} \end{vmatrix} = 0.$$

Hence or otherwise, show that multiple correlation and simple correlation are special cases of canonical correlation. 25

(c) Define Wishart matrix D and its distribution. Also, point out the applications of this distribution. If $D \sim W_P(D;\, n;\, \Sigma)$, then show that $\frac{|D|}{|\Sigma|}$ is distributed as the product of p independent chi-square variates with d.f. n, n – 1, n – 2, ..., n – p + 1, respectively. Also, obtain $E(|D|^h)$ for h = 1, 2, 3, ... If $\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3$ are independently and identically distributed as $N_2(\mu, \Sigma)$, with $\mu = 0$ and $\sum = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, obtain $E[\mathbf{X}_1\mathbf{X}_1' + \mathbf{X}_2\mathbf{X}_2' + \mathbf{X}_3\mathbf{X}_3']$. 25

(d) A researcher wants to determine a procedure for discriminating between two multivariate populations. The researcher has enough data available to estimate the density functions $f_1(\mathbf{X})$ and $f_2(\mathbf{X})$ associated with populations π_1 and π_2 respectively, where $C(2\,|\,1) = 50$ and $C(1\,|\,2) = 100$. In addition, it is known that about 20% of all possible items (for which the measurements $\mathbf{X}$ can be recorded) belong to π_2.

(i) Give the minimum ECM rule (in general form) for assigning a new item to one of the two populations.

(ii) Measurements recorded on a new item yield the density values $f_1(\mathbf{X}) = 0{\cdot}3$ and $f_2(\mathbf{X}) = 0{\cdot}5$. Given the preceding information, assign this item to population π_1 or population π_2. 25

SECTION F

(Design and Analysis of Experiments)

Q11. (a) For the one-way classified fixed effect model :

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \,;\, (i = 1, 2 \ldots, k, j = 1, 2, \ldots, n_i),$$

where $\varepsilon_{ij} \sim N(0, \sigma_e^2)$.

(i) Obtain the Least Square Estimates of μ & α_i's.

(ii) Obtain E(SST) and E(SSE),

where

SST = Sum of Squares due to treatment.

SSE = Sum of Squares due to error.

(iii) Also give ANOVA table. *10*

(b) Distinguish between 'total confounding' and 'partial confounding' in a 2^3-factorial experiment.

The following are three replicates each consisting of a complete block of a 2^3-factorial experiments in blocks of 4 plots involving three fertilizers N, P and K each at two levels.

Replicate I

Block 1	np	npk	(1)	k
Block 2	p	n	pk	nk

Replicate II

Block 3	(1)	npk	nk	p
Block 4	n	pk	np	k

Replicate III

Block 5	pk	nk	(1)	np
Block 6	np	npk	p	k

Identify the treatment effects confounded in each of the replicates and outline its ANOVA. *10*

(c) Explain what you understand by 'Missing plot technique'. Suppose one observation is missing in an m^2-Latin Square design. Let the missing observation be x. Obtain the estimate of missing observation x in a Latin Square design and analyse the data. *10*

(d) (i) In a 2^5-factorial experiment, the key block is given by

(1), ad, bc, abcd, abe, ace, cde and bde.

Identify the Confounding Effects. *5*

(ii) Given the principal block of 2^4-design as

(1), ab, cd, abcd.

Identify the Confounding Effect. *5*

(e) Explain a Split-plot Experiment. How does it differ from Randomised Block Design (RBD) ? Discuss the advantages and disadvantages relative to RBD. Also, give comparative study of the Split-plot design and Factorial design. *10*

Q12. Answer any ***two*** of the following :

(a) Let Y_1, Y_2 and Y_3 be independent random variables with common variance σ^2 and Expectation given by :

$$E(Y_1) = \theta_1 + \theta_3$$

$$E(Y_2) = \theta_2 + \theta_3$$

$$E(Y_3) = \theta_1 + \theta_3$$

(i) Prove that $b_1 + b_2\theta_2 + b_3\theta_3$ is estimable if and only if $b_3 = b_1 + b_2$.

(ii) If $b_1\theta_1 + b_2\theta_2 + b_3\theta_3$ is estimable, find its Best Linear Unbiased Estimator (BLUE) and its variance.

(iii) Find the unbiased estimators of σ^2. *25*

(b) In a Randomised Block Design (RBD) with r-blocks and v-treatments, the yield of treatment 1 and 2 in the first block are mixed up and only their total yield is known. Estimate the mixed up yields and show that the bias in the estimated treatment sum of squares is given by

$$\text{bias} = \frac{(T_1 - T_2)^2}{2\,(r-1)^2}$$

and that the loss of efficiency due to mixing of the yield is

$$[1 + (v - 1)\,(r - 1)]^{-1}.$$ *25*

(c) Let Y_{ij} (i = 1, 2 ..., s, j = 1, 2, ..., r) be independent Normal observations with common variance σ^2 and expectation given by

$$E(Y_{ij}) = \mu + \alpha_i + t_j + \beta x_{ij}\ ;\ (i = 1, 2 \ldots, r,\ j = 1, 2, \ldots, s),$$

where μ, α_i's, t_j's and β are unknown parameters.

Derive the statistics for testing (i) $\beta = 0$, and (ii) $t_1 = t_2 = \ldots = t_r = 0$. *25*

(d) (i) Explain basic principles of design of experiments pointing out the role each one plays in the valid and accurate interpretation of the data. Explain how these principles are used in Randomised Block Design (RBD). Explain the situations in which RBD is considered as an improvement over CRD. *10*

(ii) Explain the Analysis of Variance of two-way classification data with one observation per cell. Show that Mean Square Error is an unbiased estimate of population variance. *15*

SECTION G

(Computing with C and R)

Q13. (a) Express $A * \frac{B + C}{D}$ in prefix and postfix forms. *10*

(b) Write a C program to check if the given matrix is idempotent. *10*

(c) Write a C program to find the product of the digits of a 7-digit positive integer, where none of the digits is zero. *10*

(d) Given n positive integers, write R code to arrange these numbers in descending order. *10*

(e) Given bivariate data (x_i, y_i), i = 1, 2, ..., n, write R code to compute the rank correlation coefficient between X and Y. *10*

Q14. (a) Suppose X follows the gamma distribution with parameters a and b, where a, b > 1. Write a C program to compute

$$P[X > t] = \int_t^\infty \text{(density function of X)}\, dx.$$

25

(b) State the relative advantages and disadvantages of call by value and call by reference in C programming. Write a C program to evaluate the series

$$\exp\{x\} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots,$$

without using any built-in function. *25*