IES/ISS EXAM, 2018

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.

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

Assume suitable data, if necessary, and indicate the same clearly.

Normal Distribution Table is given at the end.

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 QCA Booklet must be clearly struck off.

Answers must be written in ENGLISH only.

SECTION-A

(Operations Research and Reliability)

1. (a) The costs per year (in ₹) for running a truck and its resale value are given below:

Year : 1 2 3 4 5 6 7

Running cost : 1,000 1,200 1,400 1,800 2,300 2,800 3,400

Resale value : 3,000 1,500 750 375 200 200 200

Determine at what year the replacement is due if the purchase price is ₹6,000.

(b) Five jobs have to be processed by machines A and B in the order 'first A, then B'. The processing times of the jobs in each machine are given below:

Jobs : 1 2 3 4 5

Machine A : 6 2 10 4 11

Machine B : 3 7 8 9 5

Determine the sequence of jobs that will minimize the total elapsed time. What are the idle times for the machines?

- (c) Explain the criteria—(i) optimistic, (ii) pessimistic and (iii) minimax regret in decision making under conditions of uncertainty.
- (d) The cumulative distribution function of times to failure (measured in months) for a system is

$$F(t) = 1 - \frac{100}{(t+10)^2}, \ t \ge 0$$

- (i) What is the reliability function?
- (ii) What is the failure rate as a function of time?
- (iii) Does the failure rate increase or decrease?
- (iv) What is MTTF?

(e) Reduce the game (whose payoff matrix is given below) by principle of dominance and solve it:

		Player B						
		I	$I\!I$	Ш	IV			
	I II III 1 3 2 4 2 3 4 2	4	0					
Diagram A	2	3	4	2	4			
Player A	3	4	2	4	0			
	4	0	4	0	8			

- 2. Answer any two of the following:
 - (a) (i) The following table gives the profit earned by assigning jobs 1, 2, 3 and 4 to contractors A, B, C and D:

Find the assignment rule that will yield the maximum profit.

(ii) Given the simplex tableau for a maximization-type linear programming problem :

X_B	X_1	X_2	X_3	X_4	X_5	X_6	RHS
X_2	. 0	1	-1	а	0	-1	b
X_5	0	0	c	-3	1	3	18
X_1	1	0	2	-1	0	2	16
$Z_j - C_j$	0	0	d	e	0	1	

Find the set of values for a, b, c, d and e, so that each of the following cases is true:

The current tableau

- (1) represents an optimal solution
- (2) represents an infeasible solution
- (3) is feasible but the problem has no finite optimum

15

(b) (i) Consider a system of n components connected in series. The failure time, T_i , of ith component has Weibull distribution $W(\alpha, \lambda_i)$, i = 1, 2, ..., n. Assuming T_i are independent, find the MTTF of the system.

10

(ii) The recorded lifetimes (in hours) for identical items are as follows:

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

Assuming lifetime has exponential distribution with parameter λ , find the failure rate λ and estimate the reliability at t = 10 hours.

15

(c) There are totally nine different routes from plants P, Q and R (each with respective capacity 100, 80 and 120) to markets A, B and C (each with respective demand 150, 50 and 100). The costs associated with the transportation of one unit of commodity from any plant to any market are given below:

			Markets	3
		Α	В	C
	P	4	8	10
Plants	Q	2	12	4
	R	8	14	2

Find the optimal set of routes and the number of items to be transported in the optimal route.

25

(d) (i) For a single-period inventory model with probabilistic demand and no setup cost, let c denote unit cost, h holding cost per item and unit time and p shortage cost per item and unit time.

Derive the optimal order quantity by assuming continuous density function for the demand.

10

(ii) A servicing facility has three channels to receive customers. On the average, 48 customers arrive for service in an 8-hour day. Each channel spends 15 minutes on an average to serve.

Find the average number of customers in the system, average number of customers waiting to be served and average time a customer spends in the system and in the queue.

SECTION-B

(Demography and Vital Statistics)

- 3. (a) Discuss the direct and indirect method of standardizing death rates.
 - (b) Fill in the blanks of the following table which are marked with question marks: 10

Age
$$x$$
 l_x
 d_x
 q_x
 p_x
 L_x
 T_x
 e_x^2

 20
 6,93,435
 ?
 ?
 ?
 35,081,126
 ?

 21
 6,90,673
 —
 —
 —
 ?
 ?

(c) Compute (i) GFR, (ii) SFR and (iii) TFR for the following data:

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

- (d) The number of persons dying at age 75 is 476 and the complete expectation of life at 75 and 76 years are 3.92 and 3.66 years respectively. Find the number of persons living at ages 75 and 76.
- (e) Explain E. C. Rhodes method of fitting logistic curve for population forecasting. 10

- 4. Answer any two of the following:
 - (a) What is Vital Statistics? State the uses of vital statistics. Explain registration and census method of obtaining vital statistics in brief. What are the shortcomings of these methods?

(b) Write the different measures of mortality. Discuss the merits and demerits of each.

25

(c) For the following values of l(x) of a life table, fit a Makeham curve:

25

Age x :35 30

45

50 55

l(x): 89685 86137 82277 77918 72795 66566

(d) How would you decide that the population has a tendency to increase, decrease or remain stable? Hence, define appropriate measures for measuring growth in population. Establish the relationship between gross reproduction rate and net reproduction rate.

25

SECTION-C

(Survival Analysis and Clinical Trials)

(a) Define cumulative hazard function, survival function and mean residual life function. Obtain the same for a Pareto distribution and describe its IFR/DFR property.

10

(b) In the usual notations, show that

$${}_{n} q_{x} = \frac{n({}_{n} m_{x})}{\left[1 + \frac{n}{2}({}_{n} m_{x})\right]}$$
 10

(c) Explain Gehan test. The survival times from the time of randomization of Treatment A patients and Treatment B patients are as follows:

Treatment A: 2, 3, 5, 5+, 5+, 7, 8+, 8+, 10

Treatment B: 2+, 4, 4, 7, 7+, 8, 9, 9, 10+, 11

where + denotes censored observation. Compute Gehan large sample statistic.

[Normal distribution table has been provided at the end]

10

(d) Explain the steps for planning and conduct of multicentre trials.

10

(e) Explain the aspects of data management in clinical trials.

- 6. Answer any two of the following:
 - (a) (i) Explain Cox proportional hazards model with several covariates. Discuss the maximum likelihood estimation of the parameter vector involved in the model.

(ii) 43 female breast cancer patients with negative auxiliary lymph nodes and a minimum 10-year follow-up were selected from Ohio State University Hospital Cancer Registry. Of the 43 patients, 9 were immuno- peroxidase positive and the remaining 34 remained negative. Survival times (in months) for both groups of patients are as follows:

Immunoperoxidase Negative:

19, 25, 30, 34, 37, 46, 47, 51, 56, 57, 61, 66, 67, 74, 78, 86, 122+, 123+, 130+, 134+, 136+, 141+, 143+, 148+, 151+, 152+, 153+, 154+, 156+, 162+, 164+, 165+, 182+, 189+

Immunoperoxidase Positive:

22, 23, 38, 42, 73, 77, 89, 115, 144+

where + denotes censored observation.

Perform a proportional hazards regression (two iterations only) with immunoperoxidase status as the single covariate in the model. Start with an initial guess equal to zero.

[Normal distribution table has been provided at the end]

15

(b) (i) Derive the proper survival distribution for cause C_{α} , $\alpha = 1, 2, 3, ..., k$ in competing risks. Prove the properties of various survival distributions when crude hazard rates are proportional.

15

10

(ii) Suppose 20 patients are followed for a period of 1 year and to the nearest tenth of a month, deaths were observed at the following times:

0.5, 1.5, 1.5, 3.0, 4.8, 6.2, 10.5 months

In addition, losses to follow up were recorded at:

0.6, 2.0, 3.5, 4.0, 8.5, 9.0 months

Construct Kaplan-Meier life table for 20 patients and variances.

- (c) Discuss the types of endpoints in clinical trials and their advantages and limitations.
- (d) Explain crossover trials and their limitations. What are their advantages over parallel trials? Discuss various aspects in designing and monitoring phase III sequential clinical trials.
 10+15

SECTION-D

(Quality Control)

- 7. (a) (i) Explain the use of p-charts and c-charts. What is the key difference between these two?
 - (ii) What is meant by process capability? How can it be measured? Why is it important?
 - (b) An \overline{X} -chart with 3-sigma control limits has the following parameters :

Suppose that the measured characteristic is normally distributed and has true mean of 98 and standard deviation of 8. What is the probability that the control chart will raise an out-of-control alarm at the fourth sample point? [Normal distribution table has been provided at the end]

(c) Compute measures of capability C_p and C_{pk} for the data given below and comment on it:

USL = 80, LSL = 50, Process
$$\mu$$
 = 60, Process σ = 5

(d) The control charts for \overline{X} and R are to be set up. The sample size is 5, and \overline{X} and R values are computed for each of 35 samples. The summary data are

$$\sum_{i=1}^{35} \overline{X}_i = 7805, \qquad \sum_{i=1}^{35} R_i = 1200$$

- (i) Find the trial control limits for \overline{X} and R.
- (ii) Assuming that the process is in control, estimate the process mean and standard deviation.

Use the following values:

n	4	5	6	7
A_2	0.729	0.577	0.483	0.4419
d_2	2.059	2.326	2.534	2.704
D_3	0	0	0	0.076
D_4	2.782	2.115	2.004	0.1924

10

(e)	Expl	ain the following terms :	10
	(i)	AQL	
	(ii)	LTPD	
	(iii)	Consumer's risk	
	(iv)	Producer's risk	
	(v)	OC curve	
Ans	swer a	any two of the following:	
(a)	(i)	Describe a CUSUM control chart. Compare this with a Shewhart chart with respect to performance. How is V-mask useful in CUSUM charts?	15
	(ii)	What are Average Sample Number (ASN) and Average Total Inspection (ATI)?	10
(b)	(i)	Three bagging machines of a company have the following characteristics :	
		Machines : A B C	
		Standard deviation : 0.2 0.3 0.05	
		If specifications are set as LSL = 12.35 and USL = 12.65 , determine which of the machines are capable of producing within specifications.	5
	(ii)	The number of defective bulbs in 10 random samples with 30 observations each are found to be	
		1, 3, 3, 1, 0, 5, 1, 1, 1	
		Construct a suitable 3-sigma control chart and offer your comments.	20
(c)	(i)	Discuss the key differences between chance causes and assignable causes.	5
	(ii)	Describe single and double sampling plans used in acceptance sampling. Define the OC function of a sampling plan.	10
	(iii)	Explain the moving average charts and exponentially weighted moving average charts.	10

8.

9560402898

SUDHIR SIR

[P.T.O.

(d) A quality control inspector of a company has taken 10 samples with four observations each. The data are given below:

1	2	3	4	5	6	7	8	9	10
12.5	12.8	12.1	12.2	12.4	12.3	12.6	12.4	12.6	12.1
12.3	12.4	12.6	12.6	12.5	12.4	12.7	12.3	12.5	12.7
12.6	12.4	12.5	12.5	12.5	12.6	12.5	12.6	12.3	12.5
12.7	12.8	12.4	12.3	12.5	12.6	12.8	12.5	12.6	12.8

Draw the control charts-

- (i) if the process standard deviation is given to be 0.2;
- (ii) by using sample ranges to find an estimate of process variability.

25

SECTION-E

(Multivariate Analysis)

- 9. (a) Let X_1 , X_2 , X_3 and X_4 be independent $N_p(\mu, \Sigma)$ random vectors.
 - (i) Find the marginal distribution for each of the random vectors

$$\mathbf{V}_1 = \frac{1}{4} \mathbf{X}_1 - \frac{1}{4} \mathbf{X}_2 + \frac{1}{4} \mathbf{X}_3 - \frac{1}{4} \mathbf{X}_4$$

and $\mathbf{V}_2 = \frac{1}{4}\mathbf{X}_1 + \frac{1}{4}\mathbf{X}_2 - \frac{1}{4}\mathbf{X}_3 - \frac{1}{4}\mathbf{X}_4$

- (ii) Find the joint density of the random vectors \mathbf{V}_1 and \mathbf{V}_2 defined in (i).
- (b) Let $\mathbf{X} \sim N_3(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu}' = (2, 3, -1)$ and

$$\sum = \begin{pmatrix} 7 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 2 \end{pmatrix}$$

- (i) Obtain the conditional distribution of x_1 and x_2 , given $x_3 = 2$.
- (ii) Obtain the partial correlation coefficient between x_1 and x_2 , given x_3 , namely $\rho_{12\cdot 3}$.

- (c) Explain the technique of principal component for handling a problem in multivariate analysis with too many variables. If $\sum = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho > 0$, then find the principal components associated with matrix Σ and find the percentage of total variance explained by the first principal component.
 - 10
- (d) Let X_1 , X_2 , ..., X_n be n independent observations from N_p (μ , Σ) (n > p) population. Obtain maximum likelihood estimators of the parameters. Are they independently distributed? Write the distribution of the estimators you have obtained.
- 10
- (e) Define sample correlation coefficient matrix $R = (r_{ij})$ and obtain its null distribution.

10. 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 population $N_p(\mu, \Sigma)$.

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

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

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) 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. Let 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 $f_1(\mathbf{X}) = 0.3$ and $f_2(\mathbf{X}) = 0.5$. Given the preceding information, assign this item to population Π_1 or population Π_2 .

Define canonical correlation coefficients and canonical variates. In the usual notations, show that the canonical correlations are solution of the determinant equation

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

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

Define Wishart matrix and the distribution of the Wishart matrix D. Write the applications of Wishart distribution. If $D \sim W_p(D \mid n \mid \Sigma)$, then show that $\frac{|D|}{|\Sigma|}$ is distributed as the product of p independent Chi-square variates with $n, (n-1), (n-2), \dots, (n-p+1)$ respectively. Also obtain $E(|D|^n)$, $h=1,\ 2,\ 3,\ \dots$. If $\mathbf{X}_1,\ \mathbf{X}_2$ and \mathbf{X}_3 are identically independently distributed as $N_2(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = \mathbf{0}$ and $\sum = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, obtain $E[\mathbf{X}_1\mathbf{X}_1' + \mathbf{X}_2\mathbf{X}_2' + \mathbf{X}_3\mathbf{X}_3']$. 25

SECTION-F

(Design and Analysis of Experiments)

In the usual notations, considering the following linear model, obtain the best estimates of the parameters : 10

$$y_1 = 2\beta_1 + 3\beta_2 + \epsilon_1$$

 $y_2 = 3\beta_1 + 4\beta_2 + \epsilon_2$
 $y_3 = 4\beta_1 + 5\beta_2 + \epsilon_3$

- (b) If S_L and S_M are orthogonal sum of squares with n_1 and n_2 degrees of freedom, also E (error) = 0, then establish that $\frac{S_L/n_1}{S_M/n_2}$ follows non-central F-distribution. 10
- (c) In the usual notations, assuming that the linear model of CRD is $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ ($i = 1, 2, \dots, m$ and $j = i, i + 1, \dots, n_i$) and if $a\mu + \sum t_i \tau_i$ is estimable, then prove that the best estimate is $\sum t_i Y_i$. 10

- (d) Mentioning the need for RBD and its advantages over CRD and specifying its statistical model, write the detailed statistical analysis with one missing observation.
- 10
- (e) Derive the efficiency of LSD in comparison with the corresponding RBD in which either rows or columns are omitted.
 - 10

12. Answer any two of the following:

- (a) Write a detailed note on Kruskal-Wallis test to carry out ANOVA in one-way classification.
- (b) Discussing the layout of an RBD to conduct 3³ factorial experiment in 3 blocks, explain its ANOVA along with detailed computational procedure. 25
- (c) Considering a 2³ design, describe the layout of the design by confounding second-order interaction in each replicate. Explain ANOVA with 3 replications. 25
- (d) Distinguish between ANOVA and ANCOVA. Explain detailed analysis of covariance for RBD.

SECTION—G

(Computing with C and R)

13. (a) Write a C-program to evaluate and print

$$Q = \sum_{j=1}^{m} \left[\sum_{i=1}^{j} (3j + ji)^{i+1} \right]^{2}$$

- (b) Write a C-program to verify and output whether a given square matrix is orthogonal or not.
- (c) Write a C-program to verify whether a given string is a PALINDROME.
- (d) Write an R-program to find 'SADDLE POINT' in a two-person zero-sum game and print its position in the matrix and also its value.
- (e) Given $X_{n \times k}$ and $Y_{n \times l}$, write an R-program to print the vector $\hat{\beta} = (X'X)^{-1}X'Y$.

14. Answer any two of the following:

(a) Write a suitable function in C and main program to invoke the same to print the values of the following:

$$P = \frac{x + \sin(x)}{\sqrt{1 + \cos\left(\frac{x}{x^2 + y^2}\right)}} \text{ and } Q = \frac{\tan(x)}{x^2 + 5\sqrt{1 + \cos\left(\frac{x^3}{x^4 + 1}\right)}}$$

(b) Write a C-function to compute

$$\sin^{-1}(x) = x + \frac{1}{2} \frac{x^3}{3} + \frac{1}{2} \frac{3}{4} \frac{x^5}{5} + \frac{1}{2} \frac{3}{4} \frac{5}{6} \frac{x^7}{7} + \dots \infty$$

and main program to print the value of $\cos^{-1}(x)$.

15+10

- (c) Given the 3 vectors X, Y and Z, write a program to compute—
 - (i) multiple correlation coefficient;
 - (ii) partial correlation coefficient.

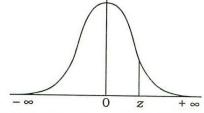
12+13

- (d) Write a program to-
 - (i) create stack on n elements;
 - (ii) insert an element after rth element;
 - (iii) delete rth element.

9+8+8

* * *

NORMAL DISTRIBUTION TABLE



							W. W	Ü	Z	1 00
	.00	·01	·02	.03	·04	.05	.06	·07	.08	.09
.0	.5000	.5040	.5080	.5120	·5160	·5199	.5239	.5279	.5319	·5359
.1	·5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
·2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	·6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
_	6015	6050	500							
.5	·6915	·6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	·7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	·7580	.7611	.7642	.7673	.7704	.7734	·7764	.7794	.7823	.7852
.8	·7881	.7910	.7939	.7967	.7995	·8023	·8051	.8078	·8106	.8133
.9	·8159	·8186	.8212	.8238	·8264	.8289	·8315	·8340	.8365	.8389
1.0	·8413	·8438	·8461	.0405	0500	0501	0== 4			
1.1	.8643	.8665	.8686	·8485	·8508	·8531	·8554	·8577	.8599	.8621
1.2	·8849	.8869		·8708	·8729	·8749	·8770	·8790	·8810	.8830
1.3	.9032	.9049	.8888	·8907	·8925	·8944	·8962	.8980	.8997	.9015
1.4	·9032		·9066	.9082	.9099	·9115	.9131	·9147	.9162	.9177
1 4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	·9535	.9545
1.7	.9554	.9564	.9573	.9582	9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
								3.00	3701	3101
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
0 =			200000000000000000000000000000000000000							
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.0000	0000	0000	0000		
3.1	.9990	.9991	.9991		.9988	.9989	.9989	.9989	.9990	.9990
3.2	.9993			.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.3		.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
	.9995	.9995	.9995	.9996	·9996	.9996	.9996	.9996	.9996	.9997
3.4	·9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998