ZLX–F–STS

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.

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.

Normal Distribution and Chi-square Distribution Tables are given at the end.

Answers must be written in ENGLISH only.

SECTION—A

(Operations Research and Reliability)

1. *(a)* A newspaper boy has the following probabilities of selling a magazine :

No. of copies sold	:	10	11	12	13	14
Probabilities	:	0·10	0·15	0·20	0·25	0·30

Cost of a copy is 30 paise and sale price is 50 paise. He cannot return unsold copies. How many copies should he order? Solve by EOL criterion. 10

(b) Find an optimum solution to the following transportation problem : 10

		Store				*Supply*
		1	2	3	4	
Factory	*A*	4	6	8	13	50
	B	13	11	10	8	70
	C	14	4	10	13	30
	D	9	11	13	8	50
	Demand	25	35	105	20	

(c) A supermarket has two salesgirls at the sales counters. If the service time for each customer is exponential with a mean of 4 minutes, and if the people arrive in a Poisson fashion at the rate of 10 an hour, then calculate *(i)* the probability that a customer has to wait for being served and *(ii)* the expected percentage of idle time for each salesgirl. *(iii)* If a customer has to wait, what is the expected length of his waiting time? 10

(d) Let h be the reliability function of a coherent system. Then prove that *(i)* $h(p \amalg p') \geq h(p) \amalg h(p')$ and *(ii)* $h(p \cdot p') \leq h(p)h(p')$ for all $0 \leq p \leq 1$, $0 \leq p' \leq 1$. 10

(e) Obtain the reliability function and the hazard function for the Weibull distribution. 10

2. Answer any *two* of the following :

(a) An electronic device consists of four components, each of which must function for the system to function. The system reliability can be improved by installing parallel units in one or more of the components. The reliability (R) of a component with one, two or three parallel units and the corresponding cost (C) are as given below. The maximum amount available for this device is 100. The problem is to determine the number of parallel units in each component. Use dynamic programming to solve the problem : 25

No. of units	*Components*							
	1		2		3		4	
	R	C	R	C	R	C	R	C
1	0·7	10	0·5	20	0·7	10	0·6	20
2	0·8	20	0·7	40	0·9	30	0·7	30
3	0·9	30	0·8	50	0·95	40	0·9	40

(b) The following network diagram represents activities associated with a project :

Activity	:	1–2	1–3	1–4	2–5	2–6	3–6	4–7	5–7	6–7
Optimistic time	:	5	18	26	16	15	6	7	7	3
Pessimistic time	:	10	22	40	20	25	12	12	9	5
Most likely time	:	8	20	33	18	20	9	10	8	4

Determine the *(i)* critical path, *(ii)* total and free floats and *(iii)* duration of the project that will have 95 percent chance of being completed. 25

(c) 25 items were placed on test and the test was terminated after a pre-assigned time $t_0 = 20$ hours from the starting time. 10 items failed before time t_0 at 3·8, 5·9, 6·3, 6·5, 7·5, 7·7, 10·5, 13·2, 13·5 and 14·0 hours. If the underlying distribution is Weibull with shape parameter equal to 2, then compute m.l.e. of the scale parameter and estimated standard error of it. 25

(d) *(i)* Develop an exact test statistic for testing the shape parameter of the gamma distribution p when the scale parameter σ is known, based on moment estimator of the shape parameter and against right-sided alternative.

(ii) For a prototype test vehicle, electrical failures for the engine operation occurred at the following kilometers : 28820, 36707, 46128 and 68345. The total test schedule was 72000 kilometers. Estimate the reliability function assuming exponential electrical failures. Also estimate the kilometers at which 10% of the vehicles will have failed. 13+12

SECTION—B

(Demography and Vital Statistics)

3. (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) Define crude and specific death rates. Explain clearly the purpose of standardizing death rates. 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 postcensal estimates of population growth assuming linear growth and exponential growth in mathematical method. 10

4. 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) Find the missing values (——) in the following table : 25

x	l_x	d_x	p_x	L_x	T_x	e_x^0	m_x
30	762227	—	—	—	27296732	—	—
31	758580						

(d) What is meant by fertility? How is it measured? Describe the various fertility rates commonly used. Discuss their relative merits. 25

SECTION—C

(Survival Analysis and Clinical Trials)

5. (a) A random variable $T(\geq 0)$ has hazard rate at time t which is given by

$$h(t) = \alpha \lambda t^{\alpha - 1}; \quad t \geq 0$$

Derive the distribution function $F(t)$ and survival function $S(t)$. Calculate $S(t)$ at $t = 5$, if $\alpha = 2$, $\lambda = 1$. 10

(b) Define gamma distribution as a failure time model. Discuss the monotonicity property of its hazard rate. 10

(c) Explain the log-rank test in a two-sample problem. 10

(d) Which quality control processes are to be used by the data analysis centre to assess the data quality and to handle problems that are observed relating to data and data analysis? 10

(e) Describe the general principles to be observed in preparation of Case-Report Form and the organization of the form. 10

6. Answer any *two* of the following :

(a) The survival times (in months) of 12 patients who have undergone a treatment are given below :

$$10 \quad 25 \quad 20 \quad 15^{+} \quad 60 \quad 18 \quad 16 \quad 17 \quad 28^{+} \quad 2 \quad 30 \quad 35^{+}$$

where + denotes right censored observations.

Assuming that the survival times follow exponential distribution with mean $1/\lambda$, obtain the maximum likelihood estimate for the mean.

Obtain the maximum likelihood estimate for $\hat{S}(t)$.

Compute $\hat{S}(t)$ at $t = 10$ months. 15+5+5

(b) The following data represents the survival times of extremely ill AIDS patients observed in days :

$$2 \quad 72 \quad 51^{+} \quad 60 \quad 33^{+} \quad 27 \quad 14 \quad 24 \quad 4 \quad 21^{+}$$

Compute Kaplan-Meier estimate of the survival function $S(t)$.

Sketch the graph of $S(t)$ against t.

Compute $\hat{S}(10)$ and $\hat{S}(25)$, and compare. 15+5+5

(c) Explain briefly the key categories of clinical trials data that are to be collected. 25

(d) Explain Phase I and Phase II clinical trials. What are MTD and DLT related to Phase I trials? Explain the traditional design approach used in Phase I trials. 25

SECTION—D

(Quality Control)

7. (a) Explain the basic differences among the chance and special causes of variation that affect process results. 10

(b) State your agreement or disagreement to the statement given below, with brief justification : 10

"A statistically controlled process will always produce 100% results within tolerance limits."

(c) Derive the control limits for the control chart for defects in a constant and varying number of sample units. 10

(d) Explain the chart which describes the sensitivity of the sampling plan to detect the lot quality level for acceptance. 10

(e) What is Cusum chart? Explain the methodology of using it for process control. 10

8. 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 use of $\overline{X}$-chart with subgroup size 5 and probability of not detecting the shift is at most 0·05. 25

(b) Explain the exponentially weighted moving average chart with at least one example of its potential application. 25

(c) Describe the variable sampling scheme and compare it with the attribute sampling scheme. 25

(d) An improvement in the process has resulted in increase of C_{PK} from 0·6 to 0·95. Estimate the reduction in % of nonconforming products. Assume the following : Process average has not changed, $C_p = C_{PK}$ and process is in statistical control. 25

SECTION—E

(Multivariate Analysis)

9. *(a)* In the ISS examination of a certain year, the distribution of scores in Statistics Papers I, II, III and IV is found to follow a 4-variate normal distribution with known parameters. Indicate how you would estimate the percentage of candidates in that examination faring better *(i)* in Papers I and II compared to Papers III and IV, both in the aggregate, and *(ii)* in Paper I compared to Paper IV. 5+5

(b) In multivariate analysis, often the issue of handling too many variables poses a big problem. How would you propose to solve the problem? Indicate the procedure. 10

(c) Assuming the form of the conditional distribution of a subset $(X_1, X_2, \cdots, X_q)$ of random variables $(X_1, X_2, \cdots, X_p)$, $q < p$, following multivariate normal law for fixed values of $(X_{q+1}, \cdots, X_p)$, derive the expression for the conditional mean of X_1 when $X_2 = x_2, \cdots, X_p = x_p$. 10

(d) Define Fisher's linear discriminant function. Give a classifier rule for classifying an individual based on his or her measurements on p identified variables, into one of two mutually exclusive and exhaustive classes, clearly stating the underlying assumptions and explaining the notation you have used. 10

(e) Show that in random sampling from a p-variate normal distribution, the sample variances and covariances follow Wishart distribution. Identify the parameters. 10

10. Answer any *two* of the following :

(a) *(i)* Show that if $\underset{\sim}{X}_{p\times 1}$ follows $N_p(\underset{\sim}{\mu}_{p\times 1}, \Sigma_{p\times p})$, then $Q = (\underset{\sim}{X} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$ follows central chi-square distribution with p degrees of freedom.

(ii) Suppose $\underset{\sim}{X}_{p\times 1}$ follows $N_p(\underset{\sim}{\mu}_{p\times 1}, \Sigma_{p\times p})$ with $E(X_i) = \mu \ \forall\, i$, $V(X_i) = \sigma^2 \ \forall\, i$ and $\text{Cov}(X_i, X_j) = \rho\sigma^2 \ \forall\, i \neq j$. Show that

$$T^2 = \frac{1}{\sigma^2}\left[\frac{\sum_{i=1}^{p}(X_i - \overline{X})^2}{1-\rho} + \frac{p(\overline{X} - \mu)^2}{1 + (p-1)\rho}\right]$$

follows chi-square distribution with p degrees of freedom, where $\overline{X} = \sum_{i=1}^{p} X_i \Big/ p$. 10+15

(b) *(i)* Define multiple correlation coefficient $r_{1\cdot 23\cdots p}$ of X_1 on $X_2, X_3, \cdots, X_p$ taken together, based on the observed set of data x_{ij}, $i = 1(1)p$, $j = 1(1)n$, and write down its expression.

(ii) Derive the simplified form of $r_{1\cdot 23\cdots p}$ for the special case $r_{ij} = r \ \forall\, i \neq j = 1(1)p$. 10+15

(c) Derive the maximum likelihood estimators for the parameters of a multivariate normal distribution. 25

(d) Discuss the procedure for testing the null hypothesis

$$H_0 : \underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)}$$

regarding the mean vectors of two multivariate normal distributions having identical dispersion matrix. 25

SECTION—F

(Design and Analysis of Experiments)

11. *(a)* Discuss the usefulness of the principle of randomization in the context of design of experiments. Use the following set of random numbers to obtain the layout of a randomized block design with 5 blocks of size 4 each for testing the treatments A, B, C and D : 10

518	275	950	305	490	333	068	088	389	933	548	275
950	305	490	334	069	736	545	774	864	062	951	005

(b) A 2^4 experiment has been laid out in 4 replicates, in each replicate confounding one of the 3-factor interactions. Give the compositions of the 4 key blocks, and write down the ANOVA table. 10

(c) What is a Latin Square Design? What are the main disadvantages of using such a design in field experiments? Indicate the procedure for obtaining the layout of a 5×5 Latin Square Design. 10

(d) A 3^2 experiment with factors A and B each at levels 0, 1 and 2 is to be designed in $2r$ replicates of 3 incomplete blocks each, such that in r replicates AB is confounded, while in the remaining r, AB^2 is confounded, the other factorial effects remaining unconfounded in the entire design.

Give the layout of the 3 incomplete blocks of each of the two types of replicates, and write down the ANOVA table. 10

(e) The factor A has p levels which require relatively large experimental units. Another factor B has q levels requiring relatively small experimental units. Suggest a suitable design for accommodating both the factors in a single experiment. Discuss the procedure for testing the main effects A and B, as also the interaction effect AB. 10

12. Answer any *two* of the following :

(a) It is desired to use a randomized block design with 4 blocks of size 6 each for testing the effects of 5 treatments A, B, C, D and E. In each block, treatments B, C, D and E are replicated once each, while treatment A is replicated twice to ensure more precise estimation and testing for A.

Using a suitable model, give the expressions for different sums of squares, and write down the ANOVA table.

Discuss the procedures for testing equality of all treatment effects, as also for equality of effects of A and B, if the hypothesis of equality of all treatment effects is rejected. 13+12

(b) A 3^3 factorial experiment with factors A, B and C each at levels 0, 1 and 2 is to be laid out in 9 incomplete blocks of size 3 each, totally sacrificing information on the factorial effects ABC^2, AC^2 and AB^2C^2.

Give the layout of the design.

If this basic pattern is replicated r times, discuss the procedure for analyzing the experiment. 10+15

(c) Discuss, in detail, the procedure for analysis of covariance in a completely randomized design for testing the effects of t treatments with $r_1, r_2, \cdots, r_t$ replications respectively in the presence of just one concomitant variable. 25

(d) A 2^3 experiment is proposed to be carried out in 2 incomplete blocks of size 4 each per replicate, retaining full information on all main effects, and sacrificing equal amount of information on the remaining four factorial effects each.

Suggest a scheme of partial confounding ensuring the above, and give the intrablock subgroup in each replicate.

Discuss the method of analysis for the data. 10+15

SECTION—G

(Computing with C and R)

13. (a) Write a C-program that will generate every nth integer, beginning with n start (i.e., $i = n$ start, n start $+n$, n start $+2n$, etc.), ending with n stop. Calculate the sum of those integers that are evenly divisible by some positive integer k. 10

(b) Write a C-program to find whether a given matrix is a symmetric matrix and print the comment. 10

(c) Write a C-program to calculate

$$y = 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{24} - \cdots + (-1)^n \frac{x^n}{n!}$$

defining a recursive function and print it. 10

(d) Write an R-program to fit a linear relationship of y on x, to get the summary of the relationship and to predict y for a given value of x. 10

(e) Write an R-program to calculate the median of a set of observations. 10

14. Answer any *two* of the following :

(a) Write a C-program that will examine each character in a line of text and determine how many of the characters are letters, how many are digits, how many are white space characters, and how many are other kinds of characters (e.g., punctuation characters) and print them. 25

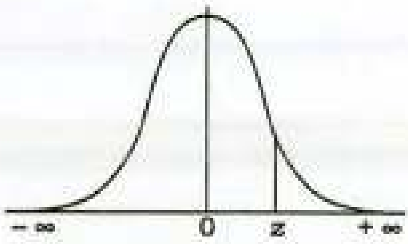
(b) Write a C-program to create a linear linked list, delete an existing component from the list defining suitable functions. 25

(c) Write a C-program to evaluate $y = x^n$, where x and y are floating-point values and n can be an integer or floating point and print it. 25

(d) Write a C-program to fit a normal distribution to a given observed frequency distribution and test for its goodness of fit and print the result. 25

★ ★ ★

NORMAL DISTRIBUTION TABLE



	·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
·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	·8485	·8508	·8531	·8554	·8577	·8599	·8621
1·1	·8643	·8665	·8686	·8708	·8729	·8749	·8770	·8790	·8810	·8830
1·2	·8849	·8869	·8888	·8907	·8925	·8944	·8962	·8980	·8997	·9015
1·3	·9032	·9049	·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
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
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	·9988	·9989	·9989	·9989	·9990	·9990
3·1	·9990	·9991	·9991	·9991	·9992	·9992	·9992	·9992	·9993	·9993
3·2	·9993	·9993	·9994	·9994	·9994	·9994	·9994	·9995	·9995	·9995
3·3	·9995	·9995	·9995	·9996	·9996	·9996	·9996	·9996	·9996	·9997
3·4	·9997	·9997	·9997	·9997	·9997	·9997	·9997	·9997	·9997	·9998

χ^2-DISTRIBUTION : VALUES OF $\chi^2_{\alpha \cdot \nu}$

ν \ α	0·995	0·99	0·975	0·95	0·05	0·025	0·01	0·005
1	0·000	0·000	0·001	0·004	3·841	5·024	6·635	7·879
2	0·010	0·020	0·051	0·103	5·991	7·378	9·210	10·597
3	0·072	0·115	0·216	0·352	7·815	9·348	11·345	12·838
4	0·207	0·297	0·484	0·711	9·488	11·143	13·277	14·860
5	0·412	0·554	0·831	1·145	11·070	12·832	15·086	16·750
6	0·676	0·872	1·237	1·635	12·592	14·449	16·812	18·548
7	0·989	1·239	1·690	2·167	14·067	16·013	18·475	20·278
8	1·344	1·646	2·180	2·733	15·507	17·535	20·090	21·955
9	1·735	2·088	2·700	3·325	16·919	19·023	21·666	23·589
10	2·156	2·558	3·247	3·940	18·307	20·483	23·209	25·188
11	2·603	3·053	3·816	4·575	19·675	21·920	24·725	26·757
12	3·074	3·571	4·404	5·226	21·026	23·337	26·217	28·300
13	3·565	4·107	5·009	5·892	22·362	24·736	27·688	29·819
14	4·075	4·660	5·629	6·571	23·685	26·119	29·141	31·319
15	4·601	5·229	6·262	7·261	24·996	27·488	30·578	32·801
16	5·142	5·812	6·908	7·962	26·296	28·845	32·000	34·267
17	5·697	6·408	7·564	8·672	27·587	30·191	33·409	35·718
18	6·265	7·015	8·231	9·390	28·869	31·526	34·805	37·156
19	6·844	7·633	8·907	10·117	30·144	32·852	36·191	38·582
20	7·434	8·260	9·591	10·851	31·410	34·170	37·566	39·997
21	8·034	8·897	10·283	11·591	32·671	35·479	38·932	41·401
22	8·643	9·542	10·982	12·338	33·924	36·781	40·289	42·796
23	9·260	10·196	11·688	13·091	35·172	38·076	41·638	44·181
24	9·886	10·856	12·401	13·848	36·415	39·364	42·980	45·558
25	10·520	11·524	13·120	14·611	37·652	40·646	44·314	46·928
26	11·160	12·198	13·844	15·379	38·885	41·923	45·642	48·290
27	11·808	12·879	14·573	16·151	40·113	43·194	46·963	49·645
28	12·461	13·565	15·308	16·928	41·337	44·461	48·278	50·993
29	13·121	14·256	16·047	17·708	42·557	45·722	49·588	52·336
30	13·787	14·953	16·791	18·493	43·773	46·979	50·892	53·672
40	20·706	22·164	24·433	26·509	55·759	59·342	63·691	66·766
50	27·991	29·707	32·357	34·764	67·505	71·420	76·154	79·490
60	35·535	37·485	40·482	43·188	79·082	83·298	88·379	91·952
70	43·275	45·442	48·758	51·739	90·531	95·023	100·425	104·215
80	51·172	53·540	57·153	60·391	101·879	106·629	112·329	116·321
90	59·196	61·754	65·647	69·126	113·145	118·136	124·116	128·299
100	67·328	70·065	74·222	77·929	124·342	129·561	135·807	140·169

For larger values of v, the variable $\sqrt{2\chi^2} - \sqrt{2v-1}$ may be used as a standard normal variable.