INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-2018

Course Name: M. Tech. (QR & OR) 2nd YEAR

Subject Name: Advanced Multivariate Analysis

Date of Examination: 04.09.2017 Maximum Marks: 70 Duration: $2\frac{1}{2}$ hours

Note:
1. This paper carries 75 marks. Answer all questions but the maximum you can score is 70.
2. All notations have their usual meanings

---

1. Distinguish between *dependence* and *interdependence* techniques. Give at least two examples each. [5]

2. Let Y = be $N_3(\underline{\mu}, \sum)$ where

$$\underline{\mu} = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \quad \textit{and} \quad \Sigma = \begin{bmatrix} 9 & 0 & -5 \\ 0 & 12 & 0 \\ -5 & 0 & 10 \end{bmatrix}$$

a) Which of the following random variables are independent? State the necessary results.

i) y_1 and y_2 ii) y_1 and y_3 iii) y_2 and y_3

iv) (y_1, y_2) and y_3 v) (y_1, y_3) and y_2 vi) (y_2, y_3) and y_1

a) Explain how will you obtain a random variable Z from Y such that Z follows $N_3(\underline{0}, \underline{I})$. State the necessary results that you use

[4+4 = 8]

3. Assume y and x are sub vectors, each 2×1, where $\begin{pmatrix} x \\ y \end{pmatrix}$ is $N_4(\mu, \Sigma)$ with

$$\mu = \begin{bmatrix} 2 \\ -1 \\ 3 \\ 1 \end{bmatrix} \quad \text{and} \quad \Sigma = \left[\begin{array}{cc|cc} 7 & 3 & -3 & 2 \\ 3 & 6 & 0 & 4 \\ \hline -3 & 0 & 5 & -2 \\ 2 & 4 & -2 & 4 \end{array}\right]$$

i) Find $E(\mathbf{y}|\mathbf{x})$

ii) Find $\text{Cov}(\mathbf{y}|\mathbf{x})$ [3 + 3 = 6]

4. Four psychological tests were given to randomly selected 32 men and 32 women respectively.

The variables measured were, y_1 = Pictorial Inconsistencies, y_2 = Paper form board, y_3 = Tool Recognition and y_4 = Vocabulary.

Assume that the four performance characteristics for men and women populations jointly follow multivariate normal distributions with the same covariance matrix.

The researcher wants to test if the mean vectors of the two populations (men and women) are equal. He / She also wants to know if the four performance characteristics contributes additionally towards group separation. The researcher computed Hotelling's T^2 values as described below.

Write down the hypotheses to be tested. Test the Hypotheses and draw conclusions.

		Variables used
$T_4^2 =$	96.603	y_1, y_2, y_3, y_4
$T_3^2 =$	78.8733	y_2, y_3, y_4
$T_3^2 =$	90.8348	y_1, y_3, y_4
$T_3^2 =$	32.6253	y_1, y_2, y_4
$T_3^2 =$	74.5926	y_1, y_2, y_3

[2 + 3 x 4 + 2 = 16]

5. Municipal wastewater treatment plants are required by law to monitor their discharge into rivers and streams on a regular basis. Concern about the reliability of data from one of these self-monitoring programs led to a study in which samples of effluent were divided and sent to two Laboratories for testing. One-half of each sample was sent to the state Laboratory of Hygiene and one-half was sent to a private laboratory routinely used in the monitoring program. Measurements of Biochemical Oxygen Demand (BOD) and Suspended Solids (SS) were obtained for n=11 sample splits, from the two laboratories. The data are shown in the following table

Sample	Private Laboratory		State Laboratory of Hygiene	
	(BOD)	SS	(BOD)	SS
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31
6	34	75	44	64
7	28	26	42	30
8	71	26	54	64
9	43	54	34	56
10	33	30	29	20
11	20	14	39	21

a) Write down the underlying model explaining the notations used and also state the associated assumptions.

b) Write down the hypothesis to be tested.

c) Test the hypotheses.

d) Draw conclusions.

[3 + 1 + 12 + 2 = 18]

6. In a study to assess the effects of Solder-Bath Temperature (SBT) and Conveyor Speed (CS) on generation of defects in PCBs a 3^2 design was run. Three PCBs were soldered at each factor combination. Two defect types D1 and D2 were observed. Some intermediate computations are furnished below. Notations have their usual meaning

e) Write down the underlying model explaining the notations used and also state the associated assumptions

f) Write down the hypotheses that are to be tested.

g) Test the hypotheses.

h) Draw conclusions.

$$E = \begin{bmatrix} 526.67 & -151.33 \\ -151.33 & 4636.67 \end{bmatrix}, \quad H_{SBT} = \begin{bmatrix} 96.52 & 295.37 \\ 295.37 & 2716.96 \end{bmatrix},$$

$$H_{CS} = \begin{bmatrix} 6.741 & -343.52 \\ -343.52 & 26787.19 \end{bmatrix}, \quad H_{SBT \times CS} = \begin{bmatrix} 117.48 & 34.85 \\ 34.85 & 3251.70 \end{bmatrix},$$

[4 + 3 + 12 + 3 = 22]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2017-18

Course name	:	M. Tech. (QR & OR)-II	
Subject Name	:	Industrial Experimentation	
Date: 05/09/2017		Maximum Marks: 70	Duration: 2 hours

NOTE: (i) This paper carries 80 marks. Answer as much as you can but the maximum you can score is 70. The marks are indicated in [ ] on the right margin.

(ii) The symbols and notations have the usual meaning as introduced in your class.

1. *Montgomery, Peck, and Vining* in *Introduction to Linear Regression Analysis*, describe an acetone-butyl alcohol distillation process for which *concentration of acetone* in the output product stream is important. Factors that may affect the concentration are *reboil temperature*, *condensate temperature* and *reflux rate*. The following records from the process are obtain and archive routinely:

   (i) concentration of acetone in an hourly test sample of output product, (ii) reboil temperature log, which is a plot of the reboil temperature over time, (iii) condenser temperature controller log, and (iv) nominal reflux rate each hour.

   The reflux rate is held constant for this process. Consequently, it didn't change much over the historical period. Thus, a study to discover the relationships among the two temperatures on the acetone concentration in the output product stream is undertaken.

   It is decided to use a sample of the historical process data, archived over a long period of time, for this study. An appropriate *full factorial design* with *factors at three levels each* are used to sample the historical data for observations on reflux rate, reboil temperature and condenser temperature. These data are then suitably analysed.

   a) Do you consider the preceding to be an experiment in light of the definition given in your class? Give justification.

   b) "In an experiment, the results and conclusions that can be drawn depend to a large extent on the manner in which the data are collected". Illustrate this point with a hypothetical experiment.

   c) What is blocking?

   [5 + 5 + 3 = 13]

2. a) What is a *components of variance* model? Illustrate with a simple example.

   b) Compute $E(SS_{Treatments})$ for a CRD with *'a'* random treatments and *'n'* replicates per treatment. (Describe the notations used.)

   [4 + 9 = 13]

3. Consider a RCBD to compare four treatments in four blocks under the usual effects model $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, i = 1,2,3,4 \,\&\, j = 1,2,3,4$ introduced in your class where ϵ_{ij} is the usual $NID(0,\sigma^2)$ random error term.

   a) When will a parametric function be estimable?

b) In light of your preceding answer, which of the following functions are estimable? Also provide an estimator for the estimable function(s) and justification for non-estimable one(s). [Simply stating 'estimable' / 'not estimable' without answering part (a) will not fetch any mark.]

(i) $3\mu + 2\tau_1 + \tau_3 + 3\beta_3$, (ii) $4\mu + 2\tau_1 + \tau_2 + \tau_4$, and (iii) $2\tau_1 - \tau_2 - 2\tau_3 + \tau_4$

[4 + 2 × 3 = 10]

4. What is a Latin square? When can we use such a square to design an experiment? Write down the model describing the notations and stating appropriate assumptions. Obtain the least squares estimators of your model parameters. How will you obtain the residuals (not the residual sum of squares) for your design?

[2 + 3 + 3 + 9 + 2 = 19]

5. Analyse the data from the following BIB design where the response 'grain size' is a 'smaller the better' characteristic.

Stirring Rate (rpm)	Furnace (Block)			
	1	2	3	4
5	13	–	14	17
10	4	5	6	–
15	–	6	9	3
20	6	9	–	6

a) Is there any evidence that stirring rate affects grain size?

b) What should be your recommendation concerning the choice of stirring rate?

[10 + 5 = 15]

6. Discuss the decomposition of the total corrected sum of squares into various sum of squares components in a two factor factorial design with interaction. Provide some justification for attributing different components to different sources, e.g., why sum of squares due to factor A is so named etc.

[10]

F distribution (5%) Table

$F_{0.05,\, v1,\, v2}$

	Degree of freedom for the Numerator (v_1)										
Degree of freedom for the Denominator (v_2)	1	2	3	4	5	6	7	8	10	12	24
2	18.5	19.0	19.2	19.2	9.3	19.3	19.4	19.4	19.4	19.4	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.61
12	4.75	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.24
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.19
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.41	2.34	2.15
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08

PERCENTAGE POINTS OF THE T DISTRIBUTION

		Tail Probabilities							
One Tail		0.10	0.05	0.025	0.01	0.005	0.001	0.0005	
Two Tails		0.20	0.10	0.05	0.02	0.01	0.002	0.001	
D	1	3.078	6.314	12.71	31.82	63.66	318.3	637	1
E	2	1.886	2.920	4.303	6.965	9.925	22.330	31.6	2
G	3	1.638	2.353	3.182	4.541	5.841	10.210	12.92	3
R	4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	4
E	5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	5
E	6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	6
S	7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7
	8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	8
O	9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	9
F	10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	10
	11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	11
F	12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	12
R	13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	13
E	14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	14
E	15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	15
D	16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	16
O	17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	17
M	18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	18
	19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	19
	20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	20
	21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	21
	22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	22
	23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	23
	24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	24
	25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	25
	26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	26
	27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	27
	28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	28
	29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	29
	30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	30
	32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	32
	34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	34
	36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	36
	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566	38
	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551	40
	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538	42
	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526	44
	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515	46
	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505	48
	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496	50
	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476	55
	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460	60
	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447	65
	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435	70
	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416	80
	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390	100
	150	1.287	1.655	1.976	2.351	2.609	3.145	3.357	150
	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340	200
Two Tails		0.20	0.10	0.05	0.02	0.01	0.002	0.001	
One Tail		0.10	0.05	0.025	0.01	0.005	0.001	0.0005	
		Tail Probabilities							

Critical values $r_{0.05}(p, df)$ for Duncan's multiple range tests

df *p*->	2	3	4	5	6	7	8	9	10	11	12	13	14
1	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.501	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516
4	3.926	4.013	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033
5	3.635	3.749	3.796	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814
6	3.460	3.586	3.649	3.680	3.694	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697
7	3.344	3.477	3.548	3.588	3.611	3.622	3.625	3.625	3.625	3.625	3.625	3.625	3.625
8	3.261	3.398	3.475	3.521	3.549	3.566	3.575	3.579	3.579	3.579	3.579	3.579	3.579
9	3.199	3.339	3.420	3.470	3.502	3.523	3.536	3.544	3.547	3.547	3.547	3.547	3.547
10	3.151	3.293	3.376	3.430	3.465	3.489	3.505	3.516	3.522	3.525	3.525	3.525	3.525
11	3.113	3.256	3.341	3.397	3.435	3.462	3.480	3.493	3.501	3.506	3.509	3.510	3.510
12	3.081	3.225	3.312	3.370	3.410	3.439	3.459	3.474	3.484	3.491	3.495	3.498	3.498
13	3.055	3.200	3.288	3.348	3.389	3.419	3.441	3.458	3.470	3.478	3.484	3.488	3.490
14	3.033	3.178	3.268	3.328	3.371	3.403	3.426	3.444	3.457	3.467	3.474	3.479	3.482
15	3.014	3.160	3.250	3.312	3.356	3.389	3.413	3.432	3.446	3.457	3.465	3.471	3.476
16	2.998	3.144	3.235	3.297	3.343	3.376	3.402	3.422	3.437	3.449	3.458	3.465	3.470
17	2.984	3.130	3.222	3.285	3.331	3.365	3.392	3.412	3.429	3.441	3.451	3.459	3.465
18	2.971	3.117	3.210	3.274	3.320	3.356	3.383	3.404	3.421	3.435	3.445	3.454	3.460
19	2.960	3.106	3.199	3.264	3.311	3.347	3.375	3.397	3.415	3.429	3.440	3.449	3.456
20	2.950	3.097	3.190	3.255	3.303	3.339	3.368	3.390	3.409	3.423	3.435	3.445	3.452
21	2.941	3.088	3.181	3.247	3.295	3.332	3.361	3.385	3.403	3.418	3.431	3.441	3.449
22	2.933	3.080	3.173	3.239	3.288	3.326	3.355	3.379	3.398	3.414	3.427	3.437	3.446
23	2.926	3.072	3.166	3.233	3.282	3.320	3.350	3.374	3.394	3.410	3.423	3.434	3.443
24	2.919	3.066	3.160	3.226	3.276	3.315	3.345	3.370	3.390	3.406	3.420	3.431	3.441
25	2.913	3.059	3.154	3.221	3.271	3.310	3.341	3.366	3.386	3.403	3.417	3.429	3.439
26	2.907	3.054	3.149	3.216	3.266	3.305	3.336	3.362	3.382	3.400	3.414	3.426	3.436
27	2.902	3.049	3.144	3.211	3.262	3.301	3.332	3.358	3.379	3.397	3.412	3.424	3.434
28	2.897	3.044	3.139	3.206	3.257	3.297	3.329	3.355	3.376	3.394	3.409	3.422	3.433
29	2.892	3.039	3.135	3.202	3.253	3.293	3.326	3.352	3.373	3.392	3.407	3.420	3.431
30	2.888	3.035	3.131	3.199	3.250	3.290	3.322	3.349	3.371	3.389	3.405	3.418	3.429

Indian Statistical Institute
Mid-semester Question paper on Reliability II
M.Tech (QR & OR).

Please answer all questions. Marks allotted to each question is given in [ ].

Maximum Marks = 75 Time = 2 1/2 hrs. Date : 6 September 2017.

1. Describe the steps involved in planning for accelerated testing. Describe also the broad categories of models used in accelerated life testing. Explain the basic assumptions required for modelling accelerated life test data.

[10+5+5=20]

2. a) Write clearly the assumptions of the Jelinski-Moranda software reliability model.

   b) What changes in the assumptions were made in the Moranda's model?

   c) Show how to estimate the parameters of the Moranda's de-eutrophication model based on the assumptions.

[5+2+8=15]

3. Let the prior distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ of the undetected number of faults be given as {0.01, 0.02, 0.03, 0.04, 0.2, 0.3, 0.2, 0.1, 0.05, 0.05, 0, 0, 0, ...}. Consider that the conditional probability of success given that there were i bugs in the software is denoted by $p_i = \frac{\pi i}{20}$. Find out the two-step-posterior probability π_2 ($\delta\delta$, 2).

[10]

4. Let the prior distribution of the number of undetected bugs be denoted by π (h,t) with the history 'h' is known and 't' denotes the number of test cases tested. Let $g_k(\pi(h,t))$, $k \geq 1$, denotes the maximum expected gain after k test cases are tested and a reward of getting one bug and debugging it immediately be 1 unit, whereas the cost of testing each test case is 'c' unit, where $c<1$. Assume that $g_o(\pi(h,t))=0$. Show that $g_k(\pi(h,t))$ is bounded above by $\mu_{\pi(h,t)}$, the finite mean of the prior distribution.

[20]

5. Explain the differences between hardware and software reliability.

[10]

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination: 2017 – 18

Course Name: M Tech (QROR), 2nd Year
Subject Name: Business Analytics

Date: 7 September 2017 Maximum Marks: 50 Duration: 2 Hours

Notes: Answer questions 3 and 4 and any one from 1 and 2

1. Define the following terms with examples
    a. Supervised learning
    b. Over fitting
    c. Predictive analytics
    d. K fold cross validation
    e. Area under ROC curve [5 X 3 = 15]
2. Consider the following problems of analytics. For each problem, identify the response variable of interest. Explain briefly how the variable may be measured. Also explain with reasons whether the problem is explanatory or predictive.
    a. A company wishes to forecast the demand of its products based on past demands as well as past and current economic conditions
    b. An e-commerce company wants to know the impact of making changes in the portal or sales policy on the volume of sales
    c. A credit card company wishes to identify fraudulent transactions so that timely actions, including stoppage of transaction may be initiated
    d. A manufacturing company wants to know which of the causes are more likely to lead to product failure
    e. The railways wish to understand whether the spate of complaint regarding quality of food is indeed a reflection of deterioration of quality. [3 X 5 = 15]
3. Consider the following data on the frequency of awarding death penalty in murder cases in the United States on the basis of the race of the defendant and victim. It may be noted that the race of the defendant and victim may be white or black. Out of 151 cases of white defendant and victim, 19 were awarded death penalty. Out of 9 cases of white defendant and black victim, none were awarded death penalty. Out of 63 cases of black defendant and white victim, 11 were awarded death penalty. Finally, out of 103 cases of black defendant and victim, 6 were awraded death penaty. Note that defendant is the person who has committed the crime, in this case murder, and the sentence is awarded to the defendant.
    a. Present this data in the form of a multi-way table indicating the race of the defendant and the victim, the number of cases when death penalty was awarded and the number of cases when death penalty was not awarded.
    b. Suppose it is proposed to carry out explanatory analytics to understand the impact of different explanatory variables on the verdict of awarding death penalty.
        i. What are the explanatory and response variables?
        ii. Would you recommend fitting a binary logistic regression model? Explain briefly.
    c. In case logistic regression is unsuitable, would it be reasonable to recommend a discriminant analysis model? Explain.

d. Consider a binary classification problem with one explanatory variable. Suppose you are aware that the conditional distribution of the explanatory varible given the response follows normal distribution for both values of the response. Work out the linear discriminant function stating the assumptions clearly. Explain why is it called the linear discriminant function. [4 + 2 + 2 + 2 + 4 + 1 = 15]

4. A telecom service provider finds that about one in 1000 international calls using their network is farudulent leading to revenue loss. They decide to implement a classification algorithm that can detect fraudulent / non-fraudulent calls with high accuracy so that the fraudulent calls may be stopped automatically.
   a. Would a confusion matrix be a good method to assess the quality of the classifier? In particular what would be your reaction if you find that the overall predictive accuracy of the model is over 99%? Explain briefly.
   b. What are the meanings of true and false positives in this case? Explain by stating the events of interest clearly.
   c. Do you think a naïve Bayes' classifier can be used in this situation? If yes, how? Explain briefly the events that you may have to define to fit the naïve Bayes' classifier.
   d. Suppose a naïve Bayes' model has been fitted. Explain how you would verify whether the developed model over fits the data?
   e. Suppose the company has collected data on the characteristics of fraudulent as well as non-fraudulent calls from its repository. Data on very large number of calls (both types exceeding several thousands) were collected and it was decided to fit a logistic regression model. Do you think it is a wise decision? If not, why not? [4 + 6 + 4 + 3 + 3 = 20]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2017-18

Course Name: M. Tech. (QROR)

Subject Name: Advanced SQC

Date: 08.09.17 **Maximum Marks:** 50 **Duration:** 2 hours

Note: Answer any three questions

1. (a) What is a multiple stream process (MSP)? Give one example of an MSP.

   (b) State the assumptions that are necessary for developing a Group control chart for monitoring an MSP.

   (c) Suppose a process has 5 streams. What will be the one-sided in-control ARL for the event that consecutive four times sample mean of a particular stream is the largest or smallest.

   (d) Use the following data to set up appropriate short-run $\bar{x}$ and R charts, assuming that the standard deviations of the measured characteristic for each part type are the same. The nominal dimensions for each part are $T_A = 100$, $T_B = 200$ and $T_C = 2000$.

Sample no.	Part type	Measurements			
		M1	M2	M3	M4
1	A	120	95	100	110
2	A	115	123	99	102
3	A	116	105	114	108
4	A	120	116	100	96
5	A	112	100	98	107
6	A	98	110	116	105
7	B	230	210	190	216
8	B	225	198	236	190
9	B	218	230	199	195
10	B	210	225	200	215
11	B	190	218	212	225
12	C	2150	2230	1900	1925
13	C	2200	2116	2000	1950
14	C	1900	2000	2115	1990
15	C	1968	2250	2160	2100
16	C	2500	2225	2475	2390

[3 + 2 + 3 + 12 = 20]

2. (a) What do you mean by 'Economic Design of $\bar{x}$ Chart'?

   b) State the assumptions about the process characteristics made by Duncan for the 'Single Assignable Cause Model for Economic Design of $\bar{x}$ Chart' and then, derive the expected length of a production cycle under these assumptions.

   c) Can a run of 50000 toothpicks produced in a toothpick manufacturing organization be called a short run? Explain your answer.

[2 + 15 + 3 = 20]

3. The tensile strength (x_1) and diameter (x_2) of a textile fibre are two important quality characteristics that are to be jointly controlled. The quality engineer has decided to use $n = 10$ fibre specimens in each sample. He has taken 20 preliminary samples. The nominal values of these quality characteristics and their sample covariance matrix have been estimated from the analysis of these preliminary samples as follows:

$$\bar{\bar{\mathbf{x}}} = \begin{bmatrix} 115.59 \\ 0.0106 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 1.23 & 0.79 \\ 0.79 & 0.83 \end{bmatrix}$$

   (a) Determine the Phase-I and Phase-II control limits of the T^2 control chart that may be used for joint monitoring of the means of the two quality characteristics. Given that $F_{0.001,2,179} = 7.18$.

   (b) Determine the control limits of the control chart for generalized variance that may be used to monitor the multivariate dispersion.

   (c) Suppose, at a time point t, the values of $\bar{x}_1$ and $\bar{x}_2$ are found to be 115.25 and 0.0104 respectively. On the other hand, at the time point t, the values of S_1^2, S_2^2 and S_{12} are found to be 1.25, 0.87 and 0.80 respectively.

      (i) Compute the T^2 value at time point t

      (ii) Compute the Generalized variance at time point t

   (d) Give the expressions of the following two models that may be used for modeling autocorrelated process data, x_t:

      (i) first-order mixed model

      (ii) first-order integrated moving average model

[4 + 6 + (5 + 2) + 3 = 20]

4. (a) State a graphical approach for detection of autocorrelation in process data.

(b) The viscosity of a chemical product is read every 2 minutes. Some data from this process are shown in the table given below (read down, then across from left to right). Set up an unweighted batch means (UBM) control chart with batch = 10, and use it to assess the current state of statistical control of the chemical process.

29.33	30.80	32.43	33.61	28.17
19.98	30.45	32.44	36.54	28.58
25.76	36.61	29.39	35.70	30.76
29.00	31.40	23.45	33.68	30.62
31.03	30.83	23.62	29.29	20.84
32.68	33.22	28.12	25.12	16.56
33.56	30.15	29.94	27.23	25.23
27.50	27.08	30.56	30.61	31.79
26.75	33.66	32.30	29.06	32.52
30.55	36.58	31.58	28.48	30.28
28.94	29.04	27.99	32.01	26.14
28.50	28.08	24.13	31.89	19.03
28.19	30.28	29.20	31.72	24.34
26.13	29.35	34.30	29.09	31.53
27.79	33.60	26.41	31.92	31.95
27.63	30.29	28.78	24.28	31.68
29.89	20.11	21.28	22.690	29.10
28.18	17.51	21.71	26.60	23.15
26.65	23.71	31.47	28.86	26.74
30.01	24.22	24.71	28.27	32.44

[2 + 18 = 20]

INDIAN STATISTICAL INSTITUTE

M. Tech. (QR & OR) 2nd YEAR

Year: 2017

MID SEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 08.09.2017 Max. Marks: 100 Time: 3 hours

Answer any five.

1. Write a general model of mathematical programming problem. State under what condition this general can be considered as LP and QP. State the duality theory. Explain the complementary slackness property.

[4+4+6+6=20]

2. a) Suppose that $f: R^n \to R$ is differentiable at $\bar{x}$. Prove that if there is a vector d such that $\nabla f(\bar{x})^t d < 0$, then there exists a $\delta > 0$ such that $f(\bar{x} + \lambda d) < f(\bar{x})$ for each $\lambda \in (0, \delta)$.

   b) Suppose that $f: R^n \to R$ is differentiable at $\bar{x}$. Prove that then $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive semidefinite if $\bar{x}$ is a local minimum.

   c) Let S be a nonempty open convex set in R^n and $f: S \to R$ be differentiable on S. Prove that f is convex if and only if

$$[\nabla f(x_2) - \nabla f(x_1)]^t (x_2 - x_1) \geq 0 \text{ for each } x_1, x_2 \in S.$$

[4+8+8=20]

3. a) Define positive definite matrix, positive semidefinite matrix, copositive matrix, and copositive plus matrix.

   b) Let $f(x_1, x_2) = 2x_2 + 2x_1^2 - 4x_1x_2 + 3x_2^2$. Find the Hessian matrix $H(x)$ and show that $H(x) \in$ PSD.

   c) Suppose A is an $m \times n$ matrix and c is an n vector. Then, show that exactly one of the following two systems has a solution:

   System 1 $Ax \leq 0$ and $c^t x > 0$ for some $x \in R^n$

   System 2 $A^t y = c$ and $y \geq 0$ for some $y \in R^m$.

[8+4+8 = 20]

P.T.O

4. a) Define epigraph and sub-gradient of a function.
   b) Let S be a nonempty convex set in R^n and let $f : S \to R$. Then show that f is convex if and only if *epi* f is a convex set.
   c) Let S be a nonempty closed convex set in R^n and $y \notin S$. Then show that there exists a nonzero vector p and a scalar α such that $p^t y > \alpha$ and $p^t x \leq \alpha$ for each $x \in S$.

[4+8+8 = 20]

5. a) Define convex, pseudo-convex and quasi-convex function.
   b) State von Neumann's minimax theorem for two-person zero-sum game.
   c) Formulate two-person zero-sum game as linear programming problem.

[6+6 + 8 = 20]

6. a) Define pure strategy and mixed strategy of a game.
   b) State Nash equilibrium point of a Bi-matrix game.
   c) Formulate Nash equilibrium point as linear complementarity problem.

[4+8 + 8 = 20]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination : 2017-18
M. TECH. (QR & OR) II YEAR
Applied Stochastic Processes and Time Series Modelling

Date: 11 September 2017 Maximum Marks: 60 Duration: 2-1/2 hours

This paper carries 70 marks. However, maximum you can score is 60.
Notation have usual meaning.

---

1 Eight light bulbs are turned on at time $t = 0$. The lifetime of any particular bulb is independent of lifetimes of all other bulbs and is described by exponential distribution having mean $1/\lambda$. Let Y be the time of third failure. Find $E[Y]$ and $V[Y]$.

[5+5 = 10]

2 Consider a Poisson process with rate λ. Compute

(a) $P\{$the 10th event occurs 2 or more time units after 9th event$\}$,

(b) $P\{$the 10th event occurs later than time 20$\}$, and

(c) $P\{$ 2 events in [1,4] and 3 events in [3,5]$\}$.

[3+6+6 = 15]

3 Prove that sum of two independent Poisson processes is a Poisson process.

[10]

4 Let $\{N(t), t \geq 0\}$ be a non-homogeneous Poisson process with intensity function $\lambda(t)$. Show that for $0 < s < t$,

$$N(s) \mid N(t) = n \quad \sim \quad \text{Bin}\left(n, \frac{\Lambda(s)}{\Lambda(t)}\right),$$

where $\Lambda(p) = \int_0^p \lambda(x)dx$.

[10]

5 A store opens at 8 A.M. and remains open till 5 in the evening. From 8 until 10, customers arrive at a Poisson rate of four an hour. Between 10 and 12, they arrive at a Poisson rate of eight an hour. From 12 to 2, the arrival rate increases steadily from eight per hour at 12 Noon to 10 per hour at 2 P.M.; and from 2 to 5, the arrival rate drops steadily from ten per hour to four per hour. Determine the probability distribution of the number of customers that arrive the store during a given day. [Hint: Give the associated intensity function and mean function explicitly.]

[12]

[P.T.O.]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2017-18

M. Tech. (QR & OR)-II

Industrial Experimentation

Date: 20/11/2017 Maximum Marks: 100 Duration 3 hours

NOTE: (i) This paper carries 120 marks. Answer as much as you can but the maximum you can score is 100. The marks are indicated in [ ] on the right margin.

(ii) The symbols and notations have the usual meaning as introduced in your class.

1. (i) What is a factorial experiment?

(ii) What is a symmetrical factorial experiment?

(iii) An engineer is interested in the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the tool life (y) in hours of a machine tool. Two levels of each factor are chosen, and three replicates of a 2^3 factorial design is run. The coded data (i.e., $y - 40$) on tool life are given in Table 1.

Table 1: Data on Tool Life (Coded)

Factor Code			Replicate		
A	B	C	I	II	III
-	-	-	-18	-9	-15
+	-	-	-8	3	-11
-	+	-	-5	-6	10
+	+	-	15	7	6
-	-	+	4	5	-2
+	-	+	0	-3	-4
-	+	+	20	10	14
+	+	+	-1	1	7

(a) Estimate the factorial effects.

(b) Use ANOVA to identify the significant effects.

(c) Write down a regression model for predicting tool life (in hours) based on the results of this experiment and your analysis.

(d) Find the standard error of the factorial effects.

[3+2+(5+4+3+3) = 20]

2. (a) What is confounding?

(b) Describe the method of confounding a 2^k factorial design in 2^p blocks of size 2^{k-p} each, and illustrate the process for $k = 6$ and $p = 3$. Use $ABEF$, $ABCD$, and ACE as the independent effects for the illustration. [Do not try to generate all the blocks.]

What other effects are confounded with the blocks? How will you compute the sum of squares due to blocks if there are two (2) replications?

(c) Introduce the concept of partial confounding with a small example? What is meant by relative information for the confounded effects?

[2+(10+4+3)+(5+2) = 26]

3. A process engineer is testing the yield of a product manufactured on three machines. Each machine can be operated at two power settings. Furthermore, a machine has three stations on which the product is formed. An experiment is conducted in which each machine is tested at both power settings, and three observations on yield are taken from each station. The runs are made in random order, and the coded results are given in Table 2. Analyze this experiment, assuming that all three factors are fixed.

Table 2: Date on Coded Yield of the Process

	Machine 1			Machine 2			Machine 3		
Station	1	2	3	1	2	3	1	2	3
Power setting 1	5.0	4.6	7.1	1.1	3.0	2.8	2.9	3.8	3.6
	1.2	5.8	7.7	3.4	4.6	5.1	3.0	3.4	2.8
	2.5	5.9	8.0	3.9	3.8	4.3	3.1	2.8	1.7
Power setting 2	-4.8	-1.0	-3.4	-6.0	-6.0	-4.0	-5.8	-6.8	-5.3
	-2.8	0.2	-3.0	-5.1	-5.0	-2.9	-3.9	-2.6	-8.0
	-2.0	-0.5	-4.2	-3.8	-2.2	-6.1	-4.7	-2.0	-5.2

[12]

4. (a) What is meant by Response Surface Methodology (RSM)?

(b) What is a response surface?

(c) Write down the second-order response surface model where the response y is influenced by k controllable process variables?

(d) What are the designs used for fitting response surfaces called?

(e) Why is it said that RSM is a sequential procedure?

(f) Describe the method of steepest ascent.

(g) In a maximization problem, if the fitted first-order response surface model is $\hat{y} = 40.44 + 0.886x_1 + 0.443x_2$ and if the basic step size is determined by the x variable whose subscript is 1 if your roll number is odd and 2 if it is even then obtain three (3) points along the path of steepest ascent, in terms of the coded variables. [State your roll number to justify your choice of subscript.]

(h) What is a simplex design? Are such designs orthogonal? Starting from the following matrix A obtain a simplex design in three factors.

$$A = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & -1 & 1 \end{bmatrix}$$

[3+3+2+1+5+7+6+(3+1+3) = 34]

5. (a) What is robust design or robust parameter design (RPD)?

(b) Who introduced the concepts of such designs?

(c) What was the result of comprehensive peer review of the pioneer's approach in the late 1980s?

(d) Briefly indicate the steps in the pioneer's approach to RPD.

(e) What information, crucial to the solution of a RPD problem, is provided by his approach? Discuss the role of this information in RPD with a suitable pictorial illustration.

(f) Describe the response surface approach to robust design problems involving two control variables and two noise variables with an appropriate first-order response model of your choice.

[3+1+2+5+(2+5)+10 = 28]

Table 3: F distribution (5%) Table

$F_{0.05, v1, v2}$

	Degree of freedom for the Numerator (v_1)										
Degree of freedom for the Denominator (v_2)	1	2	3	4	5	6	7	8	10	12	24
2	18.5	19.0	19.2	19.2	9.3	19.3	19.4	19.4	19.4	19.4	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.61
12	4.75	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.24
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.19
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.41	2.34	2.15
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

M.Tech (QR&OR) II year

Reliability II

Date: 22/11/2017 Maximum Marks: 100 Duration: 3 hours

Please answer all questions. Marks allotted to each question is given in [ ].

1) a) Define IFR and IFRA classes of life distributions.
   b) Prove that if F is IFR, then F is IFRA. Construct a counter example to show that the Converse is not true. [5+10 = 15]

2) Which one of the classes between IFR and IFRA has a closure property? Give a detailed proof of the closure property of that particular class of life distribution. [2+18=20]

3) Describe the steps involved in proper sequence while planning for an Accelerated life testing experiment. [7]

4) Differentiate between HPP, NHPP and renewal processes with examples. Also give an example where bivariate exponential distribution may be used. [9+4 = 13]

5) A group of six lives was observed over a period of time as part of a mortality Investigation. Each of the lives was under observation at all ages from age 55 until they died or were censored. The table below shows the sex, age at exit and reason for exit from the investigation.

Life	Sex	Age at exit	Reason for exit
1	M	56	Death
2	F	62	Censored
3	F	63	Death
4	M	66	Death
5	M	67	Censored
6	M	67	Censored

The following model has been suggested for the force of mortality:

$\mu (x \mid Z = z) = \mu_o (x)\exp(\beta z)$

P.T.O

Where:

- x denotes age
- $\mu_o(x)$ is the baseline hazard
- $z = 0$ for males and $z = 1$ for females.

Write down the partial likelihood for these observations using the model above. [15]

6) Let the prior distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ of the undetected number of faults in a software be given as {0.01, 0.02, 0.03, 0.04, 0.2, 0.3, 0.2, 0.1, 0.05, 0.05, 0, 0, 0, ...}. Consider that the conditional probability of success given that there were i bugs in the software is denoted by $p_i = \frac{\pi_i}{20}$. Find out the two-step-posterior probability $\pi_2(ss, 2)$.

[10]

7) Let the prior distribution of the number of undetected bugs be denoted by π (h,t) with the history 'h' is known and 't' denotes the number of test cases tested. Let $g_k(\pi(h,t))$, $k \geq 1$, denotes the maximum expected gain after k test cases are tested and a reward of getting one bug and debugging it immediately be 1 unit, whereas the cost of testing each test case is 'c' unit, where $c<1$. Assume that $g_o(\pi(h,t))=0$. Show that $g_k(\pi(h,t))$ is bounded above by $\mu_{\pi(h,t)}$, the finite mean of the prior distribution.

[20]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017 – 18

Course Name: M Tech (QROR), 2nd Year
Subject Name: Business Analytics

Date: 24.11.2017 Maximum Marks: 100 Duration: 3 Hours

Notes: There are 5 questions. Answer any four.

1. Answer the following
   a. Explain the concepts of bagging and random forests. When is random forest expected to perform better than bagging? [4 + 4 + 4 = 12]
   b. What are the parameters of boosting? [3]
   c. Write the boosting algorithm. [7]
   d. Which techniques out of bagging, boosting and random forests may suffer from over fitting? Explain briefly. [3]
2. Answer the following
   a. In cluster analysis, how would you measure the distance between two entities when the cluster variate has binary variables? How is distance measured when the cluster variate has categorical variables with more than 2 levels? [4 + 3 = 7]
   b. Explain the concepts of single linkage, complete linkage and average linkage methods used for hierarchical clustering. [3 X 5 = 15]
   c. Why is it important to standardize the cluster variates? [3]
3. Answer the following
   a. In the case of a binary classification, when will you prefer discriminant analysis over logistic regression? Why? [5]
   b. Can you use the concept of area under ROC curve to measure the classification effectiveness in case of a 2 class discriminant analysis? How? [3]
   c. Suppose someone says that the overall classification accuracy of a model is 92%. However, an expert comments that it does not tell much about the actual capability of the model? Do you agree? Explain briefly. [4]
   d. Explain how Hosmer Lemeshow test is used to check the accuracy of a logistic regression model? [8]
   e. Suppose a retailer has collected point of sales data to see which customers having a particular loyalty card has bought a particular product. [Note: Point of sale data refers to the data on the customers who have bought something. These data are collected when the customer pays at the payment counter of a retail outlet]. They want to fit a logistic regression to predict sales for different customers. Do you see any difficulty in this modelling approach? Explain briefly. [5]
4. Answer the following
   a. What are the formulations for ridge regression and lasso? [5 + 5 = 10]
   b. Ridge regression cannot be used for feature selection but lasso can be used for this purpose. Explain. [6]
   c. When is ridge regression preferred over usual multiple regression? [3]
   d. Explain basis function representation in the context of regression. [3]
   e. How many degrees of freedom are needed to fit a cubic spline with K knots? [3]

5. Answer the following
    a. What is the difference between problems of segmentation and grouping? Explain with examples. [8]
    b. Explain the PPR formulation. Why is it called universal approximator? [6 + 2 = 8]
    c. Explain the concepts of
        i. Over fitting
        ii. K fold cross validation
        iii. Supervised learning [3 X 3 = 9]

INDIAN STATISTICAL INSTITUTE

First Semester EXAMINATION: 2017

Course Name: M. TECH (QROR) II

Subject Name: OR II

Date: 27/11/2017 **Maximum Marks: 100** **Duration: 3 hours**

Answer 8. and any six from 1. To 7.

1. a) Define linear complementarity problem LCP (q, M).
   b) Formulate linear programming problem and quadratic programming problem as linear complementarity problem. State the Lemke's algorithm.

[4+3+3+5=15]

2. a) Define primal feasibility, dual feasibility and complementary slackness condition of a nonlinear programming problem.

   b) State the KKT sufficient conditions of optimality.

[6+9 = 15]

3. a) Let S be a nonempty open set in R^n and $f: R^n \to R$, $g_i: R^n \to R$ for $i=1, \ldots m$. Consider the problem to minimize $f(x)$ subject to $x \in X$ and $g_i(x) \le 0$ for $i = 1, \ldots, m$. Let $\bar{x}$ be a feasible solution and suppose that f and g_i are differentiable at $\bar{x}$. If $\bar{x}$ locally solves the problem, then show that there exist scalars u_0 and u_i such that

$$u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) = 0$$
$$u_i g_i(\bar{x}) = 0 \quad \text{for } i = 1, \ldots, m$$
$$u_0, u_i \ge 0 \quad \text{for } i = 1, \ldots, m$$
$$(u_0, u) \ne (0, 0)$$

   b) State under what condition the above statement can be written as

$$\nabla f(\bar{x}) + \sum_{i=1}^{n} \bar{u}_i \nabla g_i(\bar{x}) = 0$$
$$\bar{u}_i g_i(\bar{x}) = 0$$
$$\bar{u}_i \ge 0 \quad \text{with at least one } \bar{u}_i > 0.$$

[10+ 5 = 15]

P.T.O

4. a) State a method to solve a linear fractional programming problem as linear programming problem.
   b) How do you solve a separable nonlinear programming problem as linear programming problem?

[6 + 9 = 15]

5. Find the optimum integer solution using Gomory's cutting plane algorithm to the following problem:

$$\text{Max } x_1 + 2x_2$$
$$\text{Subject to } 2x_2 \leq 7$$
$$x_1 + x_2 \leq 7$$
$$2x_1 \leq 11$$
$$x_1, x_2 \geq 0.$$

[15]

6. State max flow, spanning tree and shortest path network optimization problems with examples. Discuss the method of finding solution for each problem.

[9+6=15]

7. State *k* out of *n* problem. Under what condition will we get integer solution directly by using simplex algorithm? Formulate max flow problem as linear programming problem. Why do we consider minimum ratio test for selecting departing variable?

[4+3+5+3 = 15]

8. Give an example of order 2 matrix for each of the five cases:

   a) copositive star matrix which is not copositive plus,
   b) non-symmetric P matrix which is not positive definite,
   c) skew symmetric matrix,
   d) copositive matrix which is not positive semidefinite
   e) copositive plus matrix

[10]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

M. Tech. (QROR): Second Year

Advanced SQC

Date: 27.11.2017 **Maximum Marks:** 100 **Duration:** 3 hours

Note : Symbols have usual meaning

Answer any five questions

1. (a) Let m be the ideal value for length of a component. The specification for the length is $m \pm \Delta$ and the loss incurred at the terminal points of the specification is 'A'. Determine the average loss per product due to variation.

   (b) At present, diagnosis on production process of one part is being done at its assembly process. That is, the production process of this part is being diagnosed by using the information from 100% inspection at its assembly process. Since any defective can be found readily at the assembly process, its diagnosis cost is nil. The loss due to a defective part is ₹50, adjustment cost is ₹5000 and the time lag is 800 parts. The production during the last 3 months was 300000 parts and finally the number of troubles during the last 3 months was 20. Now, the management has considered diagnoses immediately after production of parts. In that case, diagnosis cost increased to ₹100 and time lag decreased to 5 parts. Moreover, loss due to defective decreased to ₹40 for reduction of supplemental value. The adjustment remained unchanged at ₹5000.

   (i) Discuss the loss or gain for this proposal in which diagnoses are done immediately after production.

   (ii) As preventive maintenance, a periodical tool change with $\bar{u}' = 10000$ parts is introduced where the risk of trouble before a tool change is 0.02. Discuss the loss or gain for both the above diagnosis systems when this preventive maintenance is introduced. Assume that the cost of preventive maintenance is equal to the adjustment cost.

   [7+(6+7) = 20]

2. (a) Consider a quality control system where the process is diagnosed after every 'n' production units by checking a product. If the process is found abnormal, then it is

recovered to the normal condition. On the other hand, if the process is found normal, the production is continued without any recovering operation. Suppose, loss due to a defective product is A, diagnosis cost is B, adjustment cost is C, average trouble occurrence interval is $\bar{u}$ and the time lag of diagnosis, i.e. number of production units produced before the process is stopped after a product is found abnormal on diagnosis is l. Determine the optimal diagnosis interval for this quality control system.

(b) At a production process of ring hollowing of automobile part, two types of defectives occur where one type is reworkable defective and is 20% of all defectives and the other type is non-reworkable defective and is 80% of all defectives. The loss due to producing one reworkable defective item is ₹60 and the loss due to producing one non-reworkable defective item is ₹500. Diagnosis cost is ₹15 and adjustment cost is ₹9800. During the last one month, 2 machine-troubles occurred while the total production was 8500 items. Time lag is 4 items. The tool cost within the adjustment cost is ₹2500. Now, tool with longer life than the present one has come into the market with unit cost of ₹12000. Find how many times the expected life of this new tool has to be of the present one such that it is profitable to use this new tool instead of the present one. Assume when machine trouble occurs, the adjustment is done by changing the tool.

[10+10 = 20]

3. (a) Suppose, the target value of a quality characteristic of a chemical process is y_0. The quality characteristic is measureable and adjustable. After an interval of Δ_t, the value of the quality characteristic is measured (predicted) and instead of adjusting the whole gap between the measured and the target values, β times the gap is adjusted (β is nonnegative). Determine the optimal value of β.

(b) Describe an approach of combining SPC and EPC for controlling a production process.

(c) Describe the procedure of continuous sampling plan CSP-1 using a flow diagram.

[10+5+5 = 20]

4. (a) A product has two quality characteristics, X_1 and X_2. The nominal values of these quality characteristics and their sample covariance matrix have been determined from the analysis of 30 preliminary samples of size 10 as follows:

$$\bar{\bar{\mathbf{x}}} = \begin{bmatrix} 3.0 \\ 3.5 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 1.40 & 1.02 \\ 1.02 & 1.35 \end{bmatrix}$$

At a time point t, the values of $\bar{x}_1$ and $\bar{x}_2$ are found to be 3.1 and 3.7 respectively.

(i) Compute the T^2 value at time point t.

(ii) State Phase-II control limits of the T^2 control chart that may be used for joint monitoring of the means of the two quality characteristics.

(b) The quality of a product is defined by two characteristics: Brinell hardness number (BHN) and circular diameter. The specifications of these characteristics are:

Hardness in BHN: 250 ± 5

Diameter : 1.0000 ± 0.002 inch

The following BHN measurements were taken:

248	250	249	252	253
249	247	249	250	251
250	249	248	250	251
249	245	246	249	254

The following measurements of the diameter were also taken:

1.0010	1.0020	1.0015	1.0009	1.0019
0.9998	0.9999	1.0020	1.0011	0.9997
0.9980	1.0010	1.0009	0.9996	0.9990
1.0000	1.0013	1.0009	1.0009	1.0009

The loss caused by unacceptable BHN is ₹20 and the loss caused by unacceptable diameter is ₹30. What is the total expected loss caused by deviations from target values?

$[(5 + 2) + 13 = 20]$

5. (a) An $\bar{x}$ chart is used for monitoring a quality characteristic. Only a single assignable cause can occur at random in the process according to a Poisson process with an intensity of λ occurrences per hour. Samples are taken at intervals of h hours. The probability of getting a false alarm is α. Derive the expected number of false alarms within a production cycle.

(b) An $\bar{x}$ chart is used to maintain current control of a process. A single assignable cause of magnitude 2σ occurs and the time that the process remains in control is an exponential random variable with mean 100 hours. Suppose that sampling costs are ₹0.50 per sample and ₹0.10 per unit, it costs ₹5.00 to investigate a false alarm, ₹2.50 to find the assignable

cause, and ₹100 is the penalty cost per hour to operate in the out-of-control state. The time required to collect and evaluate a sample is 0.05 hour, and it takes 2 hours to identify and eliminate the assignable cause. Assume that the process is allowed to continue operating during searches for the assignable cause. What is the cost associated with the arbitrary control chart design $n = 5$, $k = 3$ and $h = 1$?

[8 + 12 = 20]

6. (a) Describe the procedure for Skip-lot sampling plan SkSP-2.

   (b) Compute the probability of acceptance of a lot in case of a Chain sampling plan ChSP-1 plan with $n = 5$, $c = 0$ and $i = 3$ applied to the lots with process quality $p = 0.10$.

   (c) Consider a simple situation involving a process in which feedback adjustment is appropriate and highly effective. Suppose, the process output characteristic of interest at time period $t = y_t$ and the target value for the output variable = T. We wish to keep y_t as close as possible to the target T. This process has a manipulatable variable x and a change in x will produce all of its effect on y within one period, i.e. $y_{t+1} - T = gx_t$, where g is a constant usually called the process gain. Show that the actual adjustment to the manipulatable variable made at time t, i.e. $x_t - x_{t-1} = -\frac{\lambda}{g}(y_t - T)$. Assume that the disturbance can be predicted adequately using an EWMA and $0 < \lambda \leq 1$ is the weighting factor for the EWMA.

[5 + 5 + 10 = 20]

7. (a) Define generalized C_p, C_{pl}, C_{pu} and C_{pk}.

   (b) Name four non-normal distributions that are most commonly fitted to non-normal process data for estimation of process capability indices. State two advantages and two disadvantages of this approach for estimation of process capability indices from non-normal data.

   (c) Describe the procedure for calculation of process capability indices from non-normal process data using Pearson's Tables.

[4+ (2+2+2) + 10 = 20]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2017-18

Programme: M. Tech. (QR & OR) 2nd YEAR

Course: Advanced Multivariate Analysis

Date of Examination: 29.11.2017 Maximum Marks: 100 Duration: 3½ hours

Note:
1. This paper carries 121 marks. Answer all questions but the maximum you can score is 100.
2. All notations have their usual meanings as used in the class.

---

1. Fill in the gaps:

   i) The factors in *factor analysis* are _____ _____ and not the directly observed variables.
   ii) _____ _____ are defined as linear combinations of the original variables, whereas in _____ _____ the original variables are expressed as linear combinations of the (background) factors.
   iii) _____ _____ is an exploratory data analysis technique that attempts to discover hitherto unknown groups of objects in a data set.
   iv) Objects in a *cluster* will have _____ _____ similarity and those belonging to different *clusters* will have _____ _____ similarity.

   [1+ 2 + 1 + 2 = 6]

2. i) What do you understand by multicollinearity?

   ii) How does multicollinearity affect the least square estimates of regression coefficients? Explain considering two regressor variables.

   iii) In a multiple linear regression problem with six regressors, the correlation matrix of the regressors is -

$$R = \begin{bmatrix} 1.000 & 0.052 & -0.343 & -0.498 & 0.417 & -0.192 \\ & 1.000 & -0.432 & -0.371 & 0.485 & -0.317 \\ & & 1.000 & -0.355 & -0.505 & 0.494 \\ & & & 1.000 & -0.215 & -0.087 \\ & & & & 1.000 & -0.123 \\ & & & & & 1.000 \end{bmatrix}$$

   Is there an indication of multicollinearity in the above correlation matrix? If pair-wise correlation matrix does not give any indication of multicollinearity, does it mean that there is no multicollinearity? Justify your answer. You may utilize the above example.

   iv) Discuss, briefly, two formal methods of detecting the multicollinearity.

   v) What are the methods of dealing with multicollinearity?

   [3 + 5 + 4 + (4+4) + 5= 25]

3. i) What are the purposes of principal component analysis?

ii) Is it a dependence or an interdependence technique ? Give a brief justification.

iii) Let X be a random vector of p variables with mean vector μ and dispersion matrix Σ. Let the eigenvalue and eigenvector pairs of Σ be $(\lambda_i, e_i), i = 1,2,...p$. $\lambda_1 \geq \lambda_2 \geq .....\geq \lambda_p > 0$. Show that the proportion of total variability explained by the i'th principal component is

$$\lambda_i \Big/ \sum_{j=1}^{p} \lambda_i \quad , \quad j = 1,2,....p$$

iv) If some of the eigenvalues are very close to zero, what conclusion can be made regarding collinearity of the variables?

v) If the dispersion matrix Σ is a diagonal matrix will there be any gain in extracting the principal components?

vi) If some the λ_i 's are equal, what can we say about the uniqueness of the corresponding principal component.

[2+2+ 6+2+2+3 =17]

4. i) Explain the differences and similarities between principal component analysis and factor analysis?

ii) Whether Factor Analysis is a dependence or an interdependence technique? Give a brief justification.

iii) Write down the factor model and associated assumptions.

iv) What is the significance of factor loading in factor model?

v) Suppose there are five variables X_1, X_2, X_3, X_5 and X_5. The eigenvalues and eigenvectors corresponding to sample correlation matrix are given below:

$\hat{\lambda}_1$ =	2.857	$\hat{e}_1$ =	[0.464, 0.457, 0.470, 0.421, 0.417]
$\hat{\lambda}_2$ =	0.809	$\hat{e}_2$ =	[0.240, 0.509, 0.260, − 0.526, − 0.582]
$\hat{\lambda}_3$ =	0.540	$\hat{e}_3$ =	[−0.612, 0.178, 0.335, 0.541, − 0.435]
$\hat{\lambda}_4$ =	0.452	$\hat{e}_4$ =	[0.387, 0.206, −0.662, 0.472, − 0.383]
$\hat{\lambda}_5$ =	0.343	$\hat{e}_5$ =	[−0.451, 0.676, −0.400, −0.176, 0.333]

a) Determine the number of factors that may be extracted?

b) Determine the loading matrix by principal component method.

c) Estimate the communalities and specific variance.

[4 + 2 + 4 + 1+ (4+4+3) = 22]

5. (i) Samples of size three are drawn from two bivariate normal populations with a common dispersion matrix and possibly different mean vectors μ_1 and μ_2

$$X_1 = \left[\begin{pmatrix}3\\7\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix}, \begin{pmatrix}4\\7\end{pmatrix}\right] \quad and \quad X_2 = \left[\begin{pmatrix}4\\8\end{pmatrix}, \begin{pmatrix}5\\7\end{pmatrix}, \begin{pmatrix}6\\9\end{pmatrix}\right]$$

(ii) Test for the difference between the population mean vectors using an appropriate test procedure.

(iii) Construct sample linear discriminant function to separate the groups.

(iv) Whether Discriminant Function Analysis is a dependence or an interdependence technique? Give a brief justification.

(v) Using the linear discriminant function obtained in (iii) can you say which variable is responsible for the difference?

(vi) Classify the observation $x_o' = [2, 7]$ as coming from population-1 or population-2 with equal prior probability and equal misclassification costs.

(vii) What is optimum error rate of a classification rule? Derive the expression for optimum error rate for a classification rule for two normal populations with equal covariance matrices.

(viii) Write down the expression for expected cost of misclassification for classification to several populations.

[(8 + 5 + 4 +2 + 2) + (2 + 8) + 5 = 36]

6. (i) What is the difference between Agglomerative and Divisive Method of clustering?

(ii) The sample correlation matrix of five variables is given below:

	X_1	X_2	X_3	X_4	X_5
X_1	1				
X_2	0.58	1			
X_3	0.51	0.60	1		
X_4	0.39	0.39	0.44	1	
X_5	0.46	0.32	0.43	0.52	1

Treating the sample correlation as similarity measure, cluster the variables by complete linkage method. Draw the dendrogram and decide on the number of cluster.

[3 + 12 = 15]

INDIAN STATISTICAL INSTITUTE

Semestral Examination : 2017-18

M. TECH. (QR & OR) II YEAR

Applied Stochastic Processes and Time Series Modelling

Date: 01-12-2017 Maximum Marks: 100 Duration: 3 hours

This paper carries 110 marks. However, maximum you can score is 100.
Notation have usual meaning.

---

1. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate $\lambda = 2$ per minute. What is the probability that the time elapsed between at least two of the first three events of the process is smaller than or equal to one minute?

[10]

2. We define

$$Y(t) = \begin{cases} 1 & \text{if } N(t) = 0, 2, 4, \ldots \\ 0 & \text{if } N(t) = 1, 3, 5, \ldots, \end{cases}$$

where $\{N(t), t \geq 0\}$ is a Poisson process with rate λ.

(a) Show that $P[Y(t) = 1] = [1 + \exp(-2\lambda t)]/2$ for $t \geq 0$.

(c) Find $P[Y(s) = 1 \mid N(t) = 1]$, where $0 < s < t$.

[5+5 = 10]

3. Customer arrivals to a shop is governed by Poisson process with time dependent rate:

$$\lambda(t) = \begin{cases} 4 + 2t & \text{if } 0 \leq t \leq 4 \\ 24 - 3t & \text{if } 4 < t \leq 8. \end{cases}$$

(a) Sketch the graphs of $\lambda(t)$ and $\Lambda(t)$ for $0 \leq t \leq 8$.

(b) Find the probability of no arrival in the interval (3,5].

(c) Determine the expected number of arrivals in the last 5 hours (i.e., in the interval (3,8]), given that 15 customers have arrived in the last 3 hours (i.e., in the interval (5,8]).

[8+4+8 = 20]

4. State and prove the following (explaining all the notation used) for CTMC:

(a) Chapman-Kolmogorov Equation,

(b) Kolmogorov Forward Differential Equation.

[8+7 = 15]

[P.T.O.]

5. The time between successive renewals, for the renewal process $\{N(t), t \geq 0\}$, has U(0, 1) distribution.

   (a) Show that $m(t) = e^t - 1$ for $0 < t < 1$.

   (b) Use (a) to compute $m(t)$ for $1 \leq t < 2$.

[5+8 = 13]

6. Let $\{N(t), t \geq 0\}$ be a renewal process induced by $\{X_n, n \geq 1\}$ with $\mu = \mathrm{E}[X_1] < \infty$ and $\sigma^2 = \mathrm{V}[X_1] < \infty$. Prove that

$$\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} \sim \mathrm{N}(0,1) \text{ as } t \to \infty.$$

[12]

7. (a) Define a weakly stationary process.

   (b) Find the autocorrelation function $\rho(k)$ (in general form), for $k = 0, 1, 2, \ldots$, of the stationary second-order AR process

$$X_t = \frac{1}{12}X_{t-1} + \frac{1}{12}X_{t-2} + Z_t, \text{ where } Z_t \sim \mathrm{WN}(0, \sigma^2)$$

   (c) Explain that the partial autocorrelation of a staionary AR(p) process is zero for lag $k > p$.

[2 + 8 +3 = 13]

8. Consider a time series model $X_t = a_0 + a_1 t + a_1 t^2 + Z_t$, where $\{Z_t\} \sim WN(0, \sigma^2)$ and, a_0 and a_1 are constants. Show that the differenced process with appropriate order becomes a stationary process. Identify the differenced process. [5+2 =7]

9. (a) Suppose a time series of monthly sales for 3 years contains both a linear trend and a seasonality effect. Suggest a suitable model for incorporating both linear trend and monthly seasonal effects.

   (b) Find the forecasts for h steps ahead for an ARIMA(1, 1, 2) process.

[3+7=10]

———*** xXx ***———

INDIAN STATISTICAL INSTITUTE

First Semestral Examination (Back Paper): 2017-18

Programme: M. Tech. (QR OR); II Year

Course: Advanced Multivariate Analysis

Date: 01/01/2018 Maximum Marks: 100 Duration: 3 hours

Note: Answer all questions

1) Each of 11 students wrote an informal and a formal essay. The variables recorded were the number of words and the number of verbs:

Y_1 = number of words in the informal essay
Y_2 = number of verbs in the informal essay
X_1 = number of words in the formal essay
X_2 = number of verbs in the formal essay.

Table: Number of words and Number of Verbs:

Student	Informal		Formal			
	y_1	y_2	x_1	x_2	$y_1 - x_1$	$y_2 - x_2$
1.	148	20	137	15	+11	+5
2.	159	24	164	25	-5	-1
3.	144	19	224	27	-80	-8
4.	103	18	208	33	-105	-15
5.	121	17	178	24	-57	-7
6.	89	11	128	20	-39	-9
7.	119	17	154	18	-35	-1
8.	123	13	158	16	-35	-3
9.	76	16	102	21	-26	-5
10.	217	29	214	25	+3	+4
11.	148	22	209	24	-61	-2

The data are given in the preceding table. It is required to test whether informal and formal essays are different in respect of usage of words and verbs.

(a) State and test the appropriate hypothesis.

(b) Carry out univariate tests.

(c) What are your conclusions?

[10+ 6+ 2 = 18]

2) a) What are the purposes of principal component analysis?

b) Six haematological variables x_1, x_2, x_3, x_4, x_5, and x_6 were measured on 51 workers. The summary of the data is given below:

Let S and R be the covariance matrix and correlation matrix respectively. The diagonal elements of S are 0.69, 5.4, 2006682.4, 90.3, 56.4, and 18.1. Eigenvalue of S and R are as follows:

Eigenvalues

2006760	2.42
65	1.4
18	1.03
7	0.92
3	0.2
0	0.03

First 3 eigenvectors of S and R are :

S			**R**		
e_1	**e_2**	**e_3**	**e_1**	**e_2**	**e_3**
0.00016	0.005	-0.0136	0.424	-0.561	-0.150
0.00051	0.017	0.0787	0.426	-0.528	0.087
0.99998	- 0.001	-0.0002	0.563	0.387	-0.051
0.00529	0.698	0.0174	0.454	0.267	0.166
0.00322	-0.716	0.0195	0.303	0.425	-0.296
0.00020	0.025	0.9965	0.073	0.069	0.293

(i) How many principal components you should retain, separately for S and R ?

(ii) Does the large variance of X_3 affect the pattern of the components of S?

(iii)Should we carry out the PCA with S? Justify your answer.

(iv)Interpret the components of either S or R.

[3+(3+2+3+3) = 14]

3) Consider the following observations from two bivariate normal populations with mean vectors μ_1 and μ_2 and common dispersion matrix Σ .

Population1 : $\begin{pmatrix}4\\7\end{pmatrix}, \begin{pmatrix}5\\7\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix}, \begin{pmatrix}3\\8\end{pmatrix}$

Population2 : $\begin{pmatrix}6\\8\end{pmatrix}, \begin{pmatrix}5\\7\end{pmatrix}, \begin{pmatrix}4\\8\end{pmatrix}$

a) Obtain an unbiased estimate of Σ.

b) Construct linear discriminant function for discriminating the two populations.

[4+8=12]

4) Observations on two responses y_1 and y_2 were collected for each of three treatments. The observation vectors $(y_1, y_2)'$ are as follows:

Treatment 1: $\begin{bmatrix} 4 \\ 5 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \end{bmatrix} \begin{bmatrix} 7 \\ 6 \end{bmatrix}$

Treatment 2: $\begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

Treatment 3: $\begin{bmatrix} 2 \\ 3 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix}$

a) Write down the one way MANOVA model and associated assumptions.

b) Construct the one way MANOVA table for the equality of treatment effects.

c) If the null hypothesis is rejected, test for each variable separately.

d) Draw Conclusions.

[4 + 12 + 4 +2 = 22]

5) a) What are the similarities and differences between Principal Component Analysis and factor analysis?

b) Write down the factor model and the associated assumptions.

c) Show that the factor model and communalities remain unchanged under orthogonal transformation.

d) Explain how orthogonal transformation help in factor extraction?

e) If $\Lambda = ((\lambda_{ij}))_{p \times p}$ be the loading matrix then show that the contribution of the j^{th} factor to $V(Y_i)$ is λ_{ij}

[5+ 3 + 5 + 4 + 3 = 20]

6) a) What is the difference between hierarchical and partitioning method of clustering?

b) Discuss suitable graphical techniques to detect grouping in a set of observations?

c) Consider the following distance matrix for 4 items. Apply complete-linkage method to obtain the clustering.

	1	2	3	4
1	0			
2	516	0		
3	590	833	0	
4	693	881	464	0

[3+5+6 = 14]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

M.Tech (QR&OR) II year

Reliability II

Date: 01/01/2018 Maximum Marks: 100 Duration: 3 hours

Please answer all questions. Marks allotted to each question is given in [ ].

Back Paper

1. a) Define the following classes of life distributions.

   i) DFR
   ii) DFRA
   iii) NWU
   iv) NWUE

   b) Show and prove the relationships between the above four classes of life distributions.

   [2x4+ 3x3=17]

2. a) Define likelihood ratio ordering of probability distributions.
   b) Suppose $\pi(h,t)$ denote the prior probability distribution of undetected number of bugs in a software with history h and t test cases having tested already. Suppose the likelihood ratio ordering of two probability distributions be denoted by π^* (h, t) $>^{LR}$ $\pi(h,t)$ indicating π^* (h, t) as a preferable prior. Show that if $\pi^*(h_1,t_1) >^{LR} \pi(h_2,t_2)$ then

   $\sum_{n\geq k} \pi_n^*(h_1,t_1) \geq \sum_{n\geq k} \pi_n(h_2,t_2)$ for all k≥1.

   [5+15=20]

3. Let X and Y be two random variables with hazard rate functions $h_X(x)$ and $h_Y(x)$ respectively. Suppose that X and Yare related as

   $$h_Y(x|Z=z) = h_X(x)\alpha(x),$$

   where $\alpha(x) = e^{bz(x)}$, z(x) is the time dependent covariate, b is the regression parameter. Then prove that

   (i) If $\alpha(x)$ is increasing in x and X is IFR then Y is also IFR;
   (ii) If $\alpha(x)$ is increasing in x and X is IFRA then Y is also IFRA.

P.T.O

IFR : Increasing failure rate, IFRA : Increasing failure rate average. [15]

4. a) Describe clearly the steps involved in planning for accelerated life testing for finding out reliability of a device in short term.

   b) Describe briefly the cox proportional hazard model. How and when such a model may be applied in the context of accelerated life testing?

   [7+3+3 = 13]

5. a) Discuss the differences between software and hardware reliability. Describe at least two sources of uncertainty in using software by an average user.
   b) State clearly the assumptions of Jelinski - Moranda model. Develop the model based on the assumptions and derive the expressions for the unknown parameters when 'n' time between failures are noted as $t_1, t_2, \ldots, t_n$.

   [(5+2) +8=15]

6. a) Give a schematic view of statistical analysis of successive inter-arrival times of a repairable system.
   b) Discuss any two methods to test whether a trend exists in the data of inter-arrival times of a repairable system. [10+(5+5) =20]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : 2017-18

Course name : M. Tech. (QR & OR)-II

Subject Name : Industrial Experimentation

Date: 2 / 1 / 2018 Maximum Marks: 100 Duration 3 hours

NOTE: (i) This paper carries 100 marks. Answer all the questions. The marks are indicated in [ ] on the right margin.

(ii) The symbols and notations have the usual meaning as introduced in your class.

1. What is a randomized complete block design? Write down the effects model for the design stating all your assumptions. How do you propose to estimate a missing observation for an approximate analysis of variance? Obtain the estimator of the missing value for the ith treatment in jth block. What will be the error degrees of freedom for an approximate analysis of variance with this estimated value?

[3+3+1+3+2 = 12]

2. The yield of a chemical process was measured using five batches of raw material, five acid concentrations, five standing times (A, B, C, D, E), and five catalyst concentrations ($\alpha, \beta, \gamma, \delta, \epsilon$). The Graeco-Latin square that follows, in Table 1, was used. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

Table 1: Date on yield of Chemical Process

	Acid Concentration				
Batch	1	2	3	4	5
1	$A\alpha = 26$	$B\beta = 16$	$C\gamma = 19$	$D\delta = 16$	$E\epsilon = 13$
2	$B\gamma = 18$	$C\delta = 21$	$D\epsilon = 18$	$E\alpha = 11$	$A\beta = 21$
3	$C\epsilon = 20$	$D\alpha = 12$	$E\beta = 16$	$A\gamma = 25$	$B\delta = 13$
4	$D\beta = 15$	$E\gamma = 15$	$A\delta = 22$	$B\epsilon = 14$	$C\alpha = 17$
5	$E\delta = 10$	$A\epsilon = 24$	$B\alpha = 17$	$C\beta = 17$	$D\gamma = 14$

[10]

3. (a) What is a balance incomplete block design, BIBD (a, b, r, k, λ), where the number of treatments, blocks, replications and block size are a, b, r, and k respectively and any given pair of treatments appear together in λ blocks of the design. Show that $b \geq a + r - k$. Write the associated model. Obtain the least squares estimators of the treatment effects in the BIBD.

(b) Does a BIBD (22, 22, 7, 7, 2) exist? Justify your answer. (No derivation is necessary but no credit will be given without justification)

[(3+7+3+7)+3 = 23]

4. For a 2^4 factorial experiment write the contrast corresponding to the interaction ABD using the labels (1), a, b, ab etc. for treatment combinations. If the value of the contrast is -16 and if there are 3 replications of this experiment then find out the value of the sum of squares of ABD.

[2+2 = 4]

5. Obtain a one-ninth fraction of a 3^4 design with AB^2C and ABD^2 as independent generators. Obtain the complete defining relation for this design. Write the alias structure for the main effect of A.

[8+4+6 = 18]

6. (a) How to find the location of the stationary point on a second-order response surface and how to characterize it?
(b) What is a stationary ridge and a rising ridge?
(c) What is a central composite design (CCD)? Why are they used? What is rotatability of a design? How a CCD is made rotatable? What is a spherical CCD?

[(3+5)+(3+3)+(3+1+2+2+2) = 24]

7. List and discuss, in brief, three different focus areas of robust parameter design.

[9]

F distribution (5%) Table

$F_{0.05, v1, v2}$

	Degree of freedom for the Numerator (v_1)										
Degree of freedom for the Denominator (v_2)	1	2	3	4	5	6	7	8	10	12	24
2	18.5	19.0	19.2	19.2	9.3	19.3	19.4	19.4	19.4	19.4	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.61
12	4.75	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.24
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.19
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.41	2.34	2.15
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08