INDIAN STATISTICAL INSTITUTE
Mid-semester Examination: 2015-2016 (First Semester)

M. Tech (QR & OR) II Year

Applied Stochastic Processes

Date: September 14, 2015 | Full Marks: 75 | Duration: 3 hours.

Note: Answer all questions

1. Consider a Poisson process $\{N(t), t \geq 0\}$ with rate λ.
   (a) For $0 \leq t_1 \leq t_2$, find the distribution of $N(t_1)$ given $N(t_2)$.
   (b) Consider a process $\{X(t), t \geq 0\}$ defined by $X(t) = (-1)^{N(t)}$. Find $E[X(t)]$ and Cov$[X(t), X(t+s)]$ for $s, t \geq 0$.

[5+(4+6)=15]

2. Let $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ be two independent Poisson processes with rates λ_1 and λ_2, respectively. Consider the process $\{Z(t), t \geq 0\}$ defined by $Z(t) = N_1(t) - N_2(t)$. Give an example of this process. Find the probability generating function of $Z(t)$.

[2+5=7]

3. The independent visitors of a certain Web site may be divided into two groups: those who arrived on this site voluntarily (Type 1) and those who arrived there by chance or by error (Type 2). Let $N(t)$ be the total number of visitors in the interval $[0, t]$. Suppose that $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda = 10$ per hour, and that 80 % of the visitors are Type 1. Let $N_i(t)$ be the number of Type i visitors in $(0, t]$, for $i = 1, 2$. Derive the distributions of $N_1(t)$ and $N_2(t)$.

[10]

4. Suppose that people arrive at a bus stop in accordance with a Poisson process having rate λ. The bus departs at time t. Suppose S denote the total amount of waiting time of all those that get on the bus at time t. Let $N(t)$ denote the total number arrivals by time t. Find $E(S)$.

[8]

5. Suppose $T_1 < T_2 <, \ldots, < T_n$ are the occurrence times of events of a Poisson process $\{N(t), t \geq 0\}$ with intensity function $\lambda(t)$.
   (a) Derive the joint probability density function of $T_1, T_2, \ldots, T_n$.
   (b) Consider a repairable system, which is observed until ten failures and the waiting time until the 10th failure is 230 days. Assume that the failure process can be modelled by a homogeneous Poisson process with parameter λ. Find the maximum likelihood estimate of λ. Derive a 95% confidence interval for λ.

[7 + (3 + 5) = 15]

6. Suppose that the intensity function $\lambda(t)$ of a nonhomogeneous Poisson process $\{N(t), t \geq 0\}$ is given by

$$\lambda(t) = \frac{t}{t^2 + 1}\lambda; \quad t \geq 0, \lambda > 0.$$

Let T_1 be the arrival time of the first event of the process. Calculate $P[T_1 \leq s | N(1) = 1]$, for $s \in (0, 1)$.

[5]

[P.T.O.]

7. (a) Define a covariance stationary time series.

(b) What is the difference between seasonal and cyclical components?

(c) Consider the following time series model

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, \quad t = 0, 1, 2, \ldots,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$.
Show that the autocorrelation function cuts off after lag 2.

(d) Consider a time series model:

$$X_t = a_0 + a_1 t + a_2 t^2 + Z_t,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and, a_0, a_1 and a_2 are constants.
Verify whether the process is stationary. Show that the differenced process with an appropriate order becomes an ARMA (p, q) process. Mention the values of p and q.

[2+2+5+(2+4)=15]

INDIAN STATISTICAL INSTITUTE
M. Tech. (QR & OR) 2nd YEAR
Year: 2015
MID SEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 15.09.2015 **Max. Marks: 100** **Time: 3 hours**

Answer any five from 1. to 7.

1. State Johnson's rule for an n jobs, 2 machines sequencing problem. How to find optimum solution for n jobs, 3 machines sequencing problem? Find the optimum solution of the following 2 machines, 6 jobs problem.

	Job					
Machine	A	B	C	D	E	F
1	30	120	50	20	90	110
2	80	100	90	60	30	10

[3+5+6=14]

2. Write the difference between CPM and PERT in network optimization. Use graphical method to find the minimum time needed to process the following jobs on the machines as shown. Here two jobs are to be processed on four machines A, B, C and D. The technological order for these two jobs is: Job 1 in the order $ABCD$ and Job 2 in the order DBAC. The time taken for processing the jobs on machine is:

Machine	A	B	C	D
Job 1	4	6	7	3
Job 2	5	7	8	4

[5+9=14]

3. State the branch and bound algorithm for solving an integer programming problem. How to formulate an integer programming problem to a binary integer programming problem? Formulate a k out n system as an integer programming problem.

[6+4+4=14]

P.T.O

4. Define total float, free float and independent float for an activity in networks. Explai crashing mechanism in network optimization problem.

[3+3+3+5=14]

5. A small project is composed of seven activities whose time estimates are listed in the tabl as follow:

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

i) Draw the project network.
ii) Find the expected duration and variance of each activity.
iii) Calculate early and late occurrence times for each event. What is the expected project length.
iv) Calculate the variance and standard deviation of project length.

[3+4+3+4=14]

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

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

[14]

7. A project is conducting research on a certain problem that must be solved. Three research teams are currently trying three different approaches for solving the problem. The problem is to determine how to allocate the two additional scientists to minimize the probability of failure. Data for that project are as follows:

New scientist	Probability of Failure		
	Team		
	1	2	3
0	0.40	0.60	0.80
1	0.20	0.40	0.50
2	0.15	0.20	0.30

Use dynamic programming to solve this problem.

[14

8. Assignment

[30

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2015

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

Subject Name: Advanced Statistical Methods

Date of Examination: 16.09.2014 Maximum Marks: 75 Duration: 2½ hours

1. This paper carries 85 marks. Answer all questions but the maximum you can score is 75.
2. All notations have their usual meanings

Distinguish between *dependence* and *interdependence* techniques. Give examples. [4]

a) Explain the terms *Generalized Population Variance* and *Total Population Variance*

b) Consider two population dispersion matrices

$$\Sigma_1 = \begin{bmatrix} 14 & 8 & 3 \\ 8 & 5 & 2 \\ 7 & 2 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 6 & 6 & 1 \\ 6 & 8 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Given that the generalized variance of population-2 is greater than the generalized variance of population-1. Show that of the total variance population-2 is smaller than that of population-1. Comment on why this is true in terms of the variances and correlations.

[5 + 5 = 10]

Let Y = be $N_3(\mu, \Sigma)$ where

$$\underset{\sim}{\mu} = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \quad and \quad \Sigma = \begin{bmatrix} 9 & 0 & 7 \\ 0 & 12 & 0 \\ 7 & 0 & 10 \end{bmatrix}$$

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 [5]

Suppose Y is $N_3(\mu, \Sigma)$, where $\Sigma = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}$

Is there a value of ρ for which $Y_1 + Y_2 + Y_3$ and $Y_1 - Y_2 - Y_3$ are independent? [6]

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

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

i) Find $E(\mathbf{y}|\mathbf{x})$ and $Cov(\mathbf{y}|\mathbf{x})$
ii) How linear regression model is related to these results?

[6 + 4 = 10]

6. Perspiration from 20 females was observed. Three components were measured. They were X_1: Sweat rate, X_2: Sodium content and X_3: Potassium content. The Summary of the data are given below:

$$\overline{X} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix} \quad \text{and} \quad S^{-1} = \begin{bmatrix} 0.586 & -0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix}$$

The scientist wants to test the hypothesis if the mean vector $\mu = (4, 50, 10)'$

a. Write down the test statistic you propose to use. What distributions does it follow?
b. Write down the necessary assumptions
c. Test the hypothesis. Carryout univariate tests if necessary
d. Would you prefer univariate tests instead? Justify, your answer.

[3 + 2 + 6 + 4 = 15]

7. An engineer approaches you and says that

a) He has carried out a multiple linear regression analysis with 4 independent variables X_1, X_2, X_3 and X_4 and obtained multiple correlation coefficient $R_1^2 = 0.79$.

b) Later on he added two more variables X5 and X6 and obtained multiple correlation coefficient $R^2 = 0.84$.

c) The total number of observations were 50

He wanted to know if X_5 and X_6 add significantly to the information.

i) State the statistical hypothesis he wanted to test?

ii) Outline the method of testing the hypothesis stating clearly all the required assumptions.

iii) Show that the test statistic to test the significance of the engineer's hypothesis can be written as

$$F_0 = \frac{(R^2 - R_1^2)/r}{(1-R^2)/(n-p)}$$

Where,

r = number of variables whose effect has been hypothesized to be insignificant
n = total number of observations
p = number of variables the entire variable considered + 1.

iv) Test the hypothesis and interpret the results.

[2 + 7 + 6 + 3 = 18]

In a study involving 2 variables 10 sets of observations each were obtained from 3 different groups. The sample means for the three groups on each of the variables were 10 and 20; 15 and 20; 20 and 25. The sample covariance matrices for each group were

$$S_1 = \begin{bmatrix} 4 & 1 \\ 1 & 5 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 8 & 4 \\ 4 & 9 \end{bmatrix}$$

The experimenter wished to test the hypothesis H_0: $\mu_1 = \mu_2 = \mu_3$ about the means of these variables.

Answer the following questions:

i) Write down the underlying model and associated assumptions. Explain the notations.

ii) What is the value of the Wilk's lambda?

What is the critical value of the significance test at the α level of .05?

What is the decision regarding the null hypothesis at the α level of .05?

iii) What is the observed value of Roy's largest root test ?

What is the value of Hotelling - Lawley trace?

What is the observed value of the Pillai' test?

[5 + 9 + 3 = 17]

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination: (2015 - 2016)

Course Name: M. Tech. (QR&OR) **Year: 2nd year**

Subject Name: Database Management Systems

Date: September 17, 2015 **Maximum Marks: 50** **Duration: 2 hrs**

Answer as many questions as possible. But the maximum that you can score, is 50.

1. A car showroom, called SHOP, maintains a database for storing the following information.

   i. SHOP sells new and used cars, each of which being associated with car number, model, its manufacturer and year of manufacturing, to its clients.

   ii. A repairing unit is included in SHOP for repairing old cars bought by it from its clients.

   iii. The workers of SHOP can be salespersons, mechanics and other housekeeping staffs. The entry in the database for a worker comprises an ID, worker's name, address, date of birth, phone number(s) and his/her gender.

   iv. A client is associated with his/her ID, name, address and phone number(s). A client can sell his/her old car and/or buy a new one through a salesperson. Old cars are repaired by the mechanics and then they are sold as used cars.

   v. For an old car, a repair job is created, which is associated with description of repair, car number and cost. Cost has two parts - cost of parts and cost of work. An old car may need repairing once or more before and/or after its sale as a used car.

   vi. If a worker of SHOP buys a car from SHOP, he/she will get a discount on the price of the car.

   vii. A salesperson sells a car, and each sale is associated with date of sale, car license number, car price, discount, net price, license cost, tax and total price.

   viii. Similarly, a salesperson mediates purchasing an old car from a client, and each purchase is associated with date of purchase and price.

   ix. SHOP purchases, through salespersons on different dates, new cars directly from different manufacturing companies producing various models. Each manufacturing company is identified by its name and location.

   a) Draw an appropriate Entity-Relationship diagram to show the overall structure of the said database. State and justify clearly the assumptions you have made in this regard.

   b) Create appropriate tables, from the above Entity-Relationship diagram, to store the above information with minimum redundancy. [25 + 15 = 40]

2. State Armstrong's axioms, and prove their soundness, using truth tables or otherwise.

[(1+2+2) + (2+3+3) = 13]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2015-16

Course Name: M.Tech QR&OR *II year*

Subject Name: Reliability II

Date: 18/09/2015 **Maximum Marks:** 50 **Duration:** 2hrs

Note, if any: Answer all Questions. Marks alloted to each question is given in [ ]

(Q1) State and prove the one step posterior distribution of the prior $\Pi(h, t)$ which denotes the prior distribution of the unknown number of bugs in a software. Assume that there are only two possible outcomes when a test case is tested with a software. [7]

(Q2) Mathematically optimal software testing problem resembles the optimal drilling problem in oil and gas explorations – explain with the description of the similarities between two completely different fields. [8]

(Q3) (a) Describe and develop the Jelinski-Moranda Software reliability model, clearly stating the assumptions and expressions for the unknown parameters.

(b) Write your comments about the validity of the assumptions in real world scenario.

[10+5]

(Q4) (a) Prove that $F(x)$ is IFR if and only if $1 - F(x)$ is PF_2.

(b) Give an example which shows that $1 - F(x)$ may be PF_2 but the density $f(x)$ is not PF_2.

[4+4]

(Q5) Let $\Lambda(t) = \int_0^t h(u)du$, $h(t)$ has a positive derivative. Prove that $\Lambda(t)$ is strictly convex and that $\frac{d}{dt}\left(\frac{1}{t}\Lambda(t)\right) \geqslant 0$. [5]

(Q6) (a) Prove that IFRA $\Longrightarrow$ NBU.

(b) Is the converse true? If not give example. [3+4]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2015-16

Course name	:	M. Tech. (QR & OR)-II	
Subject Name	:	Industrial Experimentation	
Date: 21.09.2015		Maximum Marks: 75	Duration: 2 hours 30 minutes

NOTE: (i) This paper carries 85 marks. Answer as much as you can but the maximum you can score is 75. The marks are indicated in [ ] on the right margin.

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

1. Define/describe the followings with suitable example wherever feasible:
Experiment, Factor & levels, Random factor. [2 × 3 = 6]

2. The response time in milliseconds was determined for three different types of circuits that could be used in an automatic valve shutoff mechanism. A total of five measurements on response time for each circuit was obtained. Consider the usual effects model for the design.

i) Write the linear model for the design along with necessary assumptions.

ii) Is $\mu + \tau_2$ estimable? Is τ_3 estimable? Is $\tau_2 - \tau_3$ estimable? [If any of your answer is 'yes' then write an estimator and if it is 'no' then justify.]

iii) Obtain the $E(MS_{Treatments})$ under your model.

iv) What is a contrast? What are orthogonal contrasts? For the above problem, write the expression for Sum of Squares due to a contrast of treatment totals.

v) Discuss the Scheffe's Method for comparing all contrasts.

[$3 + 2\times3 + 5 + \overline{2+2+2} + 6 = 26$]

3. Consider a $p \times p$ Latin square with rows (α_i), columns (β_k), and treatments (τ_j) fixed. Write down the normal equations and obtain least squares estimators of the model parameters α_i, β_k, and τ_j. Discuss the process of randomization for a Latin square design.

[8 + 4 + 6= 18]

4. What is a Balanced Incomplete Block Design (BIBD) in a treatments, b blocks, where each block can hold k ($k < a$) treatments and each treatment is replicated r times? Prove that $b \geq a + r - k$. Verify that a BIBD with parameters $a = 22$, $b = 22$, $r = 7$, $k = 7$ and $\lambda = 2$ does not exist. Show that the variance of interblock estimators of treatment effect is $V(\tilde{\tau}_i) = \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2)$.

[3 + 6 + 5 + 8 = 22]

5. An aluminum master alloy manufacturer produces grain refiners in ingot form. The company produces the product in four furnaces. Each furnace is known to have its own unique operating characteristics, so any experiment run in the foundry that involves more than one furnace will consider furnaces as a nuisance variable. The process engineers suspect that stirring rate affects the grain size of the product. Each furnace can be run at four different stirring rates. A randomized block design is run for a particular refiner and the resulting grain size data is shown below:

	Furnace			
Stirring Rate (rpm)	1	2	3	4
5	8	4	5	6
10	14	5	6	9
15	14	6	9	2
20	17	9	3	6

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

(b) What should the process engineers recommend concerning the choice of stirring rate for this particular grain refiner if small grain size is desirable?

(c) Estimate the observation corresponding to a stirring rate of 5 rpm for furnace 2.

[7 + 4 + 2 = 13]

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

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	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	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	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

Semestral Examination: (2015 - 2016)

Course Name: M. Tech. (QR&OR) **Year: 2nd year**

Subject Name: Database Management Systems

Date: November 23, 2015 **Maximum Marks: 100** **Duration: 3 hrs**

Answer as many questions as possible. But the maximum that you can score, is 100.

1. Consider the following relational schema:

Articles(ID, title, journal, issue, year, startpage, endpage, TR-ID)

It contains information on articles published in scientific journals. Each article has a unique ID, a title, and information on where to find it (i.e., name of the journal, issue number, and on which pages). Also, if results of an article previously appeared in a "technical report" (TR), the ID of this technical report can be specified, otherwise its (TR-ID) value is zero. We have the following information on the attributes:

- For each journal, an issue with a given number is published in a single year.
- The value of the endpage of an article can never be smaller than the value of the startpage.
- There is never (part of) more than one article on a single page.

The following is an instance of the corresponding relation:

ID	Title	Journal	Issue	Year	Startpage	Endpage	TR-ID
42	Metabolic pathway modelling	Computer Science	51	2004	121	133	87
33	A new model of database	Computer Science	41	2001	69	85	62
33	A new model of database	Computer Science	41	2001	69	85	56
39	Software testing	Software Engineering	31	2001	111	133	47
57	A new graph partitioning algorithm	Graph Theory	51	2008	1	3	99
77	Big data analytics	Software Engineering	51	2008	1	5	98
78	Big data analytics	Big Data	2	2008	22	22	98

Answer the following questions related to the above relation.

a) Write down, with appropriate explanation, the possible key(s).
b) State the functional dependencies that hold on the above relation. Explain your answer.
c) Based on a) and b), check whether *Articles* is in BCNF. If not, perform normalization so that the resulting schemas are in BCNF. Explain your answer.

[5+6+(2+7) = 20]

2. Consider the following relational schemas along with the schema in Question 1:

Authors(ORCHID_ID, name)
Authoring(articleID, authorID)

These schemas contain information on names of authors, their unique ORCHID_IDs and their paper IDs. The foreign keys articleID and authorID of *Authoring* refer to ID and ORCHID_ID of *Articles* and *Authors* respectively.

Consider the following query and answer the questions:

"Find the list of co-authors of the journal papers of the author named *M. Banerjee*, along with the title of their article(s)."

a) Write expressions in relational algebra, tuple relational calculus and domain relational calculus for the above query.
b) Write the SQL statement of the corresponding relational algebra expression.
[(7+7+5)+6 = 25]

3. Describe various constraints that need to be satisfied while updating a relational database. [10]

4. Describe various mapping cardinalities, with suitable real life examples, between two entity sets. [15]

5. Define partial dependency. Show that every partial dependency is a transitive dependency. [2+4 = 6]

6. Discuss the followings using an example of a file of records:
a) Insertion and Deletion of records in fixed length record representation.
b) Insertion and Deletion of records in variable length record representation. [5+5 = 10]

7. Answer the followings.
a) State two advantages and two disadvantages of the following strategies for storing a relational database.
i) Storing each relation in one file.
ii) Storing multiple relations in one file.
b) In the case of variable length record representation, give an example of the use of null bitmap, offset and length fields for a null record. [(4+4)+2 = 10]

8. With an example of B+ tree file organization, show the steps for insertion and deletion of records. Explain, using an example of a transaction, the ACID properties. What is the role of checkpoints in database system recovery? [4+4+2 = 10]

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2015-2016 (First Semester)

M. Tech. (QR & OR), II Year

Applied Stochastic Processes

Date: November 26, 2015 | Maximum Marks: 100 | Duration: 3 hours.

Notes: *This paper carries 110 marks. Answer as many questions as you can. The maximum you can score is 100.*

1. Suppose a repairable system is modelled by an NHPP with intensity function $\lambda(t) = \beta\lambda(\lambda t)^{\beta-1}$, $\beta > 0$, $\lambda > 0$. The system is observed upto nth failure and $T_1 < \cdots < T_n$ are the failure times.

   (a) Derive the distribution of T_n.

   (b) Derive the maximum likelihood estimates of β and λ.

   (c) Device a test for $H_0 : \beta = 1$ vs. $H_1 : \beta \neq 1$.

   [Hint: The joint pdf of $T_1, \ldots, T_n$ is given by $f(t_1, \ldots, t_n) = \prod_{i=1}^{n} \lambda(t_i) \exp(-\int_0^{t_n} \lambda(u)du)$.]

   [$6 + 6 + 8 = 20$]

2. Suppose that customers arrive at a shopping mall in groups according to Poisson process with rate λ. Let X_i be the number of customers in the i-th the group. Suppose X_1, X_2, ... are independent and identical distributed with common pmf

   $$P[X_i = k] = pq^{k-1}, \quad k = 1, 2, \ldots$$

   Let $M(t)$ be the total number of customers arrive by time t. Find the p.g.f. of $M(t)$ and hence find the mean of $M(t)$. Assume that X_i's are indpendent of the arrival process of customers.

   [$7 + 3 = 10$]

3. Consider a linear birth and death process $\{X(t), t \geq 0\}$ with immigration, having rates $\lambda_n = n\lambda + \alpha$, $n \geq 0$ and $\mu_n = n\mu$, $n \geq 1$.

   (a) Show that $M(t) = E[X(t)]$ satisfies $\frac{d}{dt}M(t) = (\lambda - \mu)M(t) + \alpha$.

   (b) Find the expression of $M(t)$, when $X(0) = i$.

   (c) Derive the steady-state distribution of $X(t)$ and condition for existence of the same.

   [$7 + 3 + (5 + 3) = 18$]

4. Write down the axioms and differential equations of a pure death process $\{X(t), t \geq 0\}$ with rate μ_n. Derive the distribution of $X(t)$, when $\mu_n = n\mu$ and the process starts with N individuals.

   [$(2 + 2) + 8 = 12$]

5. Let $\{N(t), t \geq 0\}$ be a renewal process having interarrival times X_n, $n \geq 1$ with $\mu = E(X_n) < \infty$ and suppose that each time a renewal occurs we receive a reward. Let R_n be the reward earned at the completion of the nth renewal. Assume that the R_n, $n \geq 1$, are independent and identically distributed with $\nu = E(R_n) < \infty$. Suppose $R(t)$ represents the total reward earned by time t. Then show that

$$\lim_{t \to \infty} \frac{E[R(t)]}{t} = \frac{\nu}{\mu}.$$

[10]

6. Consider a replacement policy, where a system is replaced upon failure or at age T_0, whichever is earlier. Let C_1 be the cost of planned replacement at T_0 and C_2, the cost of replacement upon failure.

   (a) Formulate this replacement policy as a renewal process and derive the expected renewal length.

   (b) Suppose the lifetime of the system follows exponential distribution with parameter λ. Find the log-run average cost of replacement.

[(2+5)+5=12]

7. Consider a delayed renewal process $\{N_D(t), t \geq 0\}$ whose first inter-arrival time has distribution G and the others have distribution F. Show that $M_D(t) = E[N_D(t)]$ satisfies

$$M_D(t) = G(t) + \int_0^t M_D(t-x)dF(x).$$

[8]

8. Let $\{N(t), t \geq 0\}$ be a renewal process having interarrival times X_n, $n \geq 1$. Derive the distribution of residual life at time t.

[8]

9. Consider a stationary two-state continuous-time Markov chain having states 0 and 1. The waiting time in state 0 before moving to state 1 is exponentially distributed with mean $1/\lambda$, while it follows exponential distributon with mean $1/\mu$ in state 1 before returning to state 0.

   (a) Write down the Kolmogorov's forward differential equations.

   (b) Derive the transition matrix $P(t) = ((P_{ij}(t)))$, with the initial condition $P(0) = \mathbf{I}$.

   (c) Suppose state 0 and 1 represent functioning state and repair state, respectively, of a repairable system. Then find the limiting availability of the system.

[2 + 7 + 3 = 12]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2015-16

M. Tech. (QR & OR)-II

Industrial Experimentation

Date: 30/11/2015 Maximum Marks: 90 Duration 3 hours

NOTE: (i) This paper carries 105 marks. Answer as much as you can but the maximum you can score is 90. The marks are indicated in [ ] on the right margin.

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

1. How do you define an experiment? What is meant by the strategy of experimentation? What is a factorial experiment, and what is its advantage over one-factor-at-a-time experiment? What is a nested factor? When is a design said to be of resolution R? What is response surface methodology? [Give examples wherever feasible. **Your answers must be precise, to the point and appropriate for the marks indicated.**]

[3+1+(3+5)+2+3+3=20]

2. Write short notes on <u>any three</u> of the following:

a) Fisher's least significant difference method,
b) Graeco-Latin square design,
c) Intra-block analysis of a BIB design,
d) Central composite design. [$6 \times 3 = 18$]

3. A laboratory is given the task to analyse the dispersion stability of a certain type of paint in relation to the four factors - Type of solvent (A), Amount of solvent (B), Disperser type (C) and Mixing method (D), all at three levels. The measurement of the dispersion stability is done by producing a certain amount of paint which then is stored under controlled conditions for a fixed time period after which the degree of dispersion in the paint is assessed microscopically.

a) Initially it is suggested to carry out an investigation in which all factors combinations are taken into account. Unfortunately, however, the laboratory cannot store so many samples at the same time. Therefore the experiment has to be split up. Assuming that the laboratory can store about 30 samples at the same time, work out a reasonable design for the experiment. [You do not need to write out the detailed plan, but it suffices to show the principle used and how a few treatment combinations are handled.]

b) After seeing the design and calculating the total cost of the experiment it is agreed that it probably is not necessary to carry out all possible factor combinations. It is therefore decided to implement a fractional factorial design using 27 treatment combinations. Work out such a design and list some of the treatment combinations which are in the fraction and some which are not in it explaining the principle used. Write out the alias structure for ONE main effect and ONE two-factor interaction. What is the resolution of your design?

c) In order to improve the design worked out in question (b), the laboratory suggests to block the experiment such that it is split into 3 blocks each having 9 treatment combinations. The reason for this is that from one mix (batch) of raw material it is not possible to formulate all 27 samples required for the design, but 9 samples can be made in a practical and homogeneous way. You may assume that the interaction effect of CD need not be measured. [Answer this question by showing the principle used for constructing the design and by finding three treatment combinations from each of the block.]

[6+(6+2×2+1)+6 = 23]

4. An experiment was performed to investigate the capability of a measurement system. Ten parts were randomly selected, and two randomly selected operators measured each part three times. The tests were made in random order, and the data below resulted.

Part No.	Operator 1 Mesurements			Operator2 Measurements		
	1	2	3	1	2	3
1	50	49	50	50	48	51
2	52	52	51	51	51	51
3	53	50	50	54	52	51
4	49	51	50	48	50	51
5	48	49	48	48	49	48
6	52	50	50	52	50	50
7	51	51	51	51	50	50
8	52	50	49	53	48	50
9	50	51	50	51	48	49
10	47	46	49	46	47	48

(a) Write the linear model for the experiment with necessary assumptions and analyse the data from this experiment. Find point estimates of the variance components using the analysis of variance method.

(b) Reanalyse the experiment assuming that the tests are destructive and a total of sixty parts were randomly selected for the experiment. Estimate the appropriate variance components and restate the model assumptions.

[(3+7+4)+(4+3+1) = 22]

5. Briefly discuss the analysis of a second-order response surface with emphasis on the location of the stationary point and, on characterising the response surface in the immediate vicinity of this point.

[8]

6. In a captive foundry, an experiment is to be conducted to reduce the rejection due to blowholes on the spigot surface of a motor end cover casting. The factors and their levels considered are given in the next page (Table 1).

It is required to study all the main effects along with six two factor interactions $B\times C$, $B\times D$, $B\times F$, $C\times D$, $C\times F$ and $D\times F$.

Design the experiment in not more than 16 runs. Assuming two replications, outline the steps in analysis of the data.

TABLE 1: Factors and their Levels for the Foundry Experiment

Sl. No.	Factor Code	Factor	Levels	
			1	2
1.	*A*	Number of vents in the mould	4+1	8+1
2.	*B*	Gating area	Present	Increased by 25%
3.	*C*	Gating Design	Present	Modified
4.	*D*	Time lag between mould closing and metal pouring	Within 2 hours	Between 2-4 hours
5.	*E*	Venting design in the core	Present	Modified
6.	*F*	Core coating type	Alcohol based	Water based
7.	*G*	Core bench life in Alcohol base	Fresh core	One day old core
8.	*H*	Core finishing in Water base	Process-1	Process-2

[8+6 = 14]

----- x -----

TABLE 2: Orthogonal Array - OA(16,15,2,2) or $L_{16}(2^{15})$

No. \ Col.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1

TABLE 3: Interaction between columns

L_{16} (2^{15})

Col.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	(1)	3	2	5	4	7	6	9	8	11	10	13	12	15	14
		(2)	1	6	7	4	5	10	11	8	9	14	15	12	13
			(3)	7	6	5	4	11	10	9	8	15	14	13	12
				(4)	1	2	3	12	13	14	15	8	9	10	11
					(5)	3	2	13	12	15	14	9	8	11	10
						(6)	1	14	15	12	13	10	11	8	9
							(7)	15	14	13	12	11	10	9	8
								(8)	1	2	3	4	5	6	7
									(9)	3	2	5	4	7	6
										(10)	1	6	7	4	5
											(11)	7	6	5	4
												(12)	1	2	3
													(13)	3	2
														(14)	1

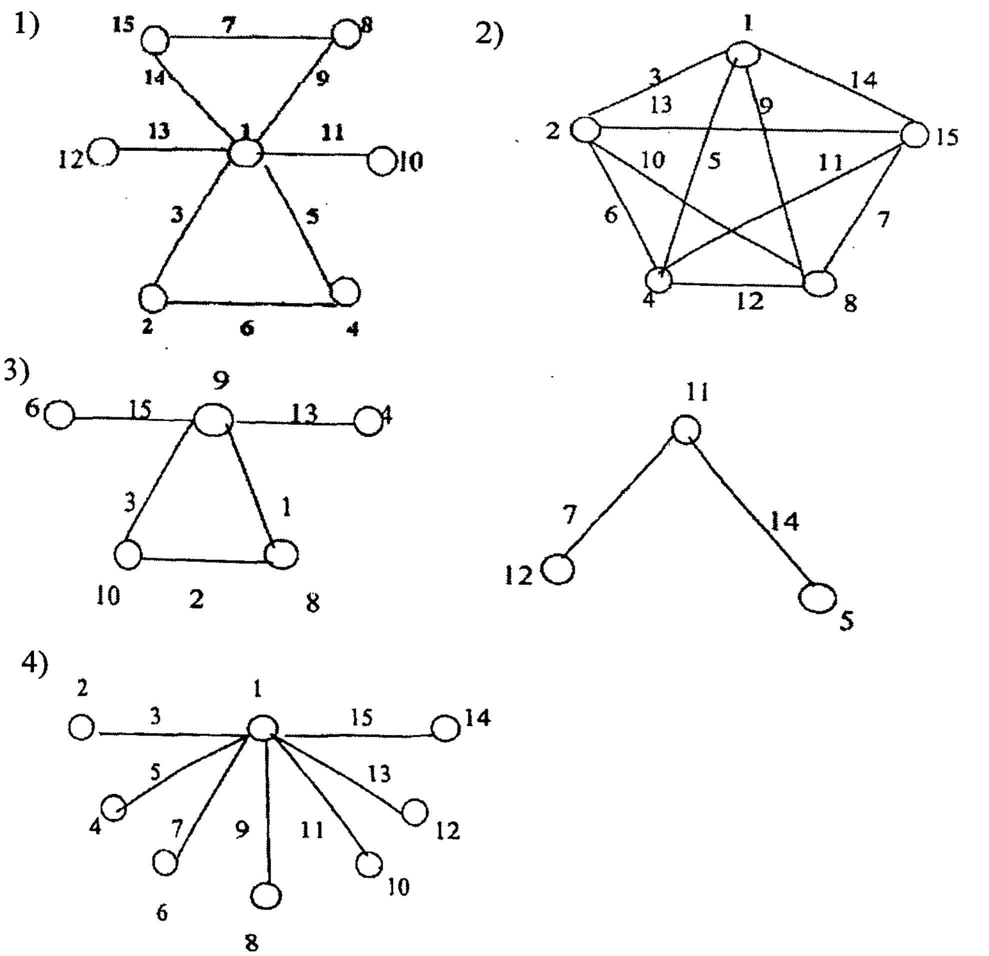

Fig. 1: Some Linear Graphs of OA(16,15,2,2)

Table 4: F distribution (5%) Table

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

Degree of freedom for the Denominator (v_2)	Degree of freedom for the Numerator (v_1)										
	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: 2015-16

Course Name: M. TECH (QROR) II

Subject Name: OR II

Date: 02.12.2015 **Maximum Marks: 100** **Duration: 3 hours**

Answer any four from 1. to 5.

1. a) Define convex, pseudo-convex and quasi-convex function.
   b) Let $f(x_1, x_2) = 2x_1 + 6x_2 + 2x_1^2 - 4x_1x_2 + 3x_2^2$. 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:

$$System\,1 \quad Ax \le 0 \text{ and } c'x > 0 \quad \text{for some } x \in R^n$$
$$System\,2 \quad A'y = c \text{ and } y \ge 0 \quad \text{for some } y \in R^m.$$

[6+6 + 8 = 20]

2. 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 $p'y > \alpha$ and $p'x \le \alpha$ for each $x \in S$.

[4+10 + 6 = 20]

3. a) State the duality theorem.
   b) 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

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

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

P.T.O

$$\nabla f(\bar{x}) + \sum_{i=1}^{n} \bar{u}_i \nabla g_i(\bar{x}) = 0$$

$$\bar{u}_i g_i(\bar{x}) = 0$$

$$\bar{u}_i \geq 0 \quad \text{with at least one } \bar{u}_i > 0.$$

[8+8+4=20]

4. a) Define linear complementarity problem LCP (q, M).
   b) Formulate linear programming problem and quadratic programming problem as linear complementarity problem.
   c) Consider an LCP (q, M) where

$$M = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 5 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 \end{bmatrix} \qquad q = \begin{bmatrix} -2 \\ -2 \\ -2 \\ -5 \end{bmatrix}$$

   Solve this LCP (q, M) by using Lemke's algorithm.

[4+8+8=20]

5. a) Define copositive, copositive-plus matrix and copositive star matrix.
   b) Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 & 8 \\ 2 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 1 & 8 & 0 & 1 \end{bmatrix}$$

   Is the matrix A copositive-star?

   c) State the formulation of linear fractional programming problem as linear programming problem.

   or

   State the stages of separable programming problem.

[6+6+8=20]

6. Assignment

[20]

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2015-16
(First Semester)

Course Name: M.Tech (QR&OR) II year

Subject Name : Reliability II

Date: 04/12/2015 Maximum Marks: 100 Duration: 3 hours

Notes: Answer all Questions. Marks allotted to each question are given in [ ]

(Q1) 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 censored. The table below shows 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:

$$h(x|Z=z)=h_0(x)e^{bz}$$

where x denotes age, $h_0(x)$ is the baseline hazard and z=0 for males and z=1 for females. Write down the partial likelihood for these observations using the model above. [10]

(Q2) Let X and Y be two random variables with hazard rate functions $h_X(x)$ and $h_Y(x)$ respectively. Suppose that X and Y are 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.

IFR : Increasing failure rate, IFRA : Increasing failure rate in average. [7×2=14]

(Q3) If a life distribution F is DFR (decreasing failure rate) then prove that F is DFRA (decreasing failure rate in average). [6]

(Q4) Explain in brief the software quality characteristics. Justify why reliability of the software is the most important among them. [8+2=10]

(Q5) (a) Consider the case where both stress and strength distributions of a particular device under operation follow Gamma distribution with parameters (μ,n) and (λ,m) respectively, where λ, μ, n, m > 0. Derive the expression for reliability of the device as a function of the above parameters.

(b) Prove using (a) that if both stress and strength distributions are exponential with μ and λ as the respective parameters, then the expression of reliability is $R=\mu/(\mu+\lambda)$.

(c) The strength of a component has a gamma distribution with parameters $\lambda=1$ and m=4. The failure inducing stress is also gamma distributed with $\mu=1$ and n=2. Compute the reliability of the component. [12+4+4=20]

(Q6) A system consists of five subsystems that must function if the system has to function properly. The system reliability goal is 0.950. Assume that all the five subsystems have identical reliability improvement effort functions. The estimated subsystem reliabilities at the present time are 0.65, 0.75, 0.85, 0.90 and 0.95. What reliability goals should be apportioned to the subsystems so as to minimize the total effect spent on the system improvement? [20]

(Q7) An extensive accelerated life testing experiment is conducted by subjecting the device to temperatures of 105°C, 210°C and 300°C. The average failure times at these temperatures are 9287 hrs., 5244 hrs. and 1295 hrs. respectively. Determine the mean time to failure of the device at 30°C. You may assume that the life distribution at each temperature and the normal operating condition follows exponential distribution. [20]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2015-16

Course Name: **M. Tech. (QR OR); II Year**

Subject Name: **Advanced Statistical Methods**

Date: 07 .12.2015 Maximum Marks: 100 Duration: 3½ hours

Note: This paper carries 120 marks. You can answer any part of any question, but maximum you can score is 100.

---

1) Comment on the following statements with justification:

   a) One may carry out separate ANOVAs for each variable instead of a MANOVA.

   b) A multiple linear regression model may contain terms like X_1^2, X_2X_3, $X_3^2X_4$ etc.

   c) A model developed by multiple linear regression method may or may not represent the underlying causal model.

   d) Single linkage method & complete linkage method would lead to the same cluster.

[3 x 4 =12]

2) A soft drink bottler is analyzing the vending machine service routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machines in an outlet. The service activity includes stocking the machine with beverage products and minor maintenance or housekeeping. The industrial engineer responsible for the study has suggested that the two most important variables affecting the delivery time (Y) are the number of cases of product (X_1) stocked and the distance travelled by the route driver (X_2). 25 observations were collected on the triplet (y, x_1 and x_2).

From the data $(x'x)^{-1}$ was obtained as follows:

$$\begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix}$$

The fitted model is

$y = 2.34123 + 1.615991x_1 + 0.01438x_2$

and following outputs were obtained

SSR = 5550.81092, SSE =223.73168

i) Does X_2 contribute significantly to the model given that X_1 is in the model?

ii) Is there any high leverage point in the data ? What is the effect of high leverage in the fitted model ?

iii) What do you mean by an influential observation? Is there any influential observation in the given data ?

You may refer to the table below:

	Observed Value (y)	**Predicted Value y-hat**	Residual (e_i)	Studentized Residual (r_i)	Cook's Distance (D_i)	Leverage (h_{ii})
1	16.68000	21.70808	-5.02808	-1.6277	0.100092	0.1018
2	11.50000	10.35361	1.14639	0.3648	0.003376	0.0707
3	12.03000	12.07979	-0.04979	-0.0161	0.000009	0.0987
4	14.88000	9.95565	4.92435	1.5797	0.077647	0.0854
5	13.75000	14.19440	-0.44440	-0.1418	0.000543	0.0750
6	18.11000	18.39957	-0.28957	-0.0908	0.000123	0.0429
7	8.00000	7.15538	0.84462	~ 0.2704	0.002172	0.0818
8	17.83000	16.67340	1.15661	0.3663	0.003051	0.0637
9	79.24000	71.82030	7.41970	3.2138	3.419313	0.4983
10	21.50000	19.12359	2.37641	0.8133	0.053845	0.1963
11	40.33000	38.09251	2.23750	0.7181	0.016200	0.0861
12	21.00000	21.59304	-0.59304	-0.1933	0.001596	0.1137
13	13.50000	12.47299	1.02701	0.3252	0.002295	0.0611
14	19.75000	18.68246	1.06753	0.3411	0.003293	0.0782
15	24.00000	23.32880	0.67120	0.2103	0.000632	0.041
16	29.00000	29.66293	-0.66293	-0.2227	0.003289	0.1659
17	15.35000	14.91364	0.43636	0.1380	0.000401	0.0594
18	19.00000	15.55138	3.44862	1.1130	0.043978	0.0963
19	9.50000	7.70681	1.79319	0.5788	0.011919	0.0964
20	35.10000	40.88797	-5.78797	-1.8735	0.132445	0.1017
21	17.90000	20.51418	-2.61418	-0.8778	0.050861	0. 1653
22	52.32000	56.00653	-3.68653	-1.4500	0.451045	0.3916
23	18.75000	23.35757	-4.60757	-1.4437	0.029899	0.0413
24	19.83000	24.40285	-4.57285	-1.4961	0.102322	0.1206
25	10.75000	10.96258	-0.21258	-0.0675	0.000108	0.0666

[6+5+5 = 16]

3) a) What is multicollinearity?

b) What is variance inflation factor? How does it help to detect the multicollinearity?

[2 + 4 = 6]

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

b) Six haematological variables x_1, x_2, x_3, x_4, x_5, 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, 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]

4) a) Suppose that Hotelling's T^2 test statistic was used to test the hypothesis $H_o : \mu_1 = \mu_2$ based on two samples of size n_1 and n_2 from two multivariate normal populations: $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$ respectively.

i) In the event of rejection of the hypothesis, explain how will you identify the significant variables responsible for rejection of hypothesis?

ii) Write down the sample linear discriminant function, which maximally separate the two populations.

iii) Is there any relationship between the discriminant criteria and Hotelling's T^2 test statistic ?

b) Observations are taken from two bivariate normal populations with equal variance-covariance matrix Σ. The data and some relevant computations are given below:

$$Y_1 : \begin{pmatrix} 7 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ 7 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 9 \end{pmatrix}, \begin{pmatrix} 7 \\ 7 \end{pmatrix}$$

$$Y_2 : \begin{pmatrix} 9 \\ 10 \end{pmatrix}, \begin{pmatrix} 6 \\ 8 \end{pmatrix}, \begin{pmatrix} 9 \\ 5 \end{pmatrix}, \begin{pmatrix} 8 \\ 7 \end{pmatrix}, \begin{pmatrix} 10 \\ 8 \end{pmatrix}$$

$$\bar{y}_1 = \begin{pmatrix} 6.0 \\ 6.6 \end{pmatrix}, \quad \bar{y}_2 = \begin{pmatrix} 8.0 \\ 7.60 \end{pmatrix} \quad S_1 = \begin{bmatrix} 0.80 & -0.40 \\ -0.40 & 2.60 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

i) Construct the discriminant function, which best separates the two populations. Which variable has highest contribution in group separation?

ii) Consider a new observation $y' = (4, 8)$. Classify this observation.

iii) Assuming prior probabilities $p_1 = .7$ and $p_2 = .3$ compute the error rate (APER) of the classification rule.

c) Show that the optimal classification rule to classify k multivariate normal population with same covariance matrix is:

Allocate **y** to G_i if $L_i(\mathbf{y})$ = largest of $\{L_1(\mathbf{y}), L_2(\mathbf{y}), \ldots L_k(\mathbf{y})\}$, $i = 1, \ldots, k$,

where G_i is the ith group, and

$$L_i(\mathbf{y}) = \bar{\mathbf{y}}_i' S_{pl}^{-1} \mathbf{y} - \tfrac{1}{2} \bar{\mathbf{y}}_i' S_{pl}^{-1} \bar{\mathbf{y}}_i + \ln(p_i), \quad i = 1, 2, \ldots k$$

$\bar{\mathbf{y}}_i$ is the sample mean corresponding to i^{th} population,

p_i is the prior probability, and

S_{pl} is the pooled sample variance-covariance matrix.

[(4+ 1 + 2) + (6+2+6)) + 9 = 30]

6) (a) What is the difference between principal component analysis and factor analysis?

(b) Explain the terms communality and specific variance in a factor model.

(c) Show that the assumptions of the factor models and communality remain unchanged under orthogonal transformation.

(d) In a consumer-preference study with five variables, the factor loadings were estimated by principal component method using correlation matrix. The estimated factor loadings are given below.

Variable	Estimated factor loadings	
	F_1	F_2
X_1	0.61	0.78
X_2	0.79	-0.59
X_3	0.64	0.75
X_4	0.93	-0.10
X_5	0.78	-0.53

Calculate communalities and specific variance for each variable. Calculate the proportion of total sample variance due to the first factor.

e) What is the purpose of rotation in factor analysis? How is the varimax rotation achieved?

[3 +4+ 6 +4 +5 = 22]

7) (a) What is the difference between classification analysis and cluster analysis?

(b) Differentiate between Partitional and Hierarchical clustering

(c) The distance matrix between pairs of five items is given below.

$$\begin{bmatrix} 0 & 2 & 6 & 1 & 10 \\ 2 & 0 & 9 & 7 & 3 \\ 6 & 9 & 0 & 4 & 5 \\ 1 & 7 & 4 & 0 & 12 \\ 10 & 3 & 5 & 12 & 0 \end{bmatrix}$$

Cluster the five items using any-linkage method. Draw the dendogram and interpret.

[4+4+(9 + 3) = 20]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : 2015-16

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

Subject Name : Industrial Experimentation

Date: 16/02/2016 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. a) What are the basic principles of experimentation?

   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 simple, hypothetical example.

   c) Draw the diagram for a general model of a process. Based on this diagram, list four objectives of an experiment.

   d) Why is nonstatistical knowledge considered invaluable in designing an experiment?

(3+4+$\overline{3+4}$+4) = [18]

2. The effect of insulin on the blood concentration of glucose was studied on rabbits. Three rabbits received insulin doses A, B and C (corresponding to 0, 1 and 2 units respectively) at different days. The experiment is given below with the glucose measurements (mg per 100 ml blood) taken 50 minutes after injection.

	Rabbit					
Day	1		2		3	
1	A	50	C	39	B	36
2	C	37	B	51	A	53
3	B	51	A	60	C	37

Write the associated statistical model describing the model parameters. State all the necessary assumptions, and analyse the data. Obtain estimates of the effects of interest. Why is it not possible to investigate if there is an interaction between rabbit and dose based on these data?

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

3. a) Consider a two factor factorial design in factors A and B, having a and b levels respectively. Both A and B are fixed factors and there are n replications. Write down the complete model describing the model parameters. Obtain the least square estimators of all the model parameters except error variance. What is the expression

for the fitted value of the kth observation based on your least square estimators, when A is at level i and B is at level j, $i = 1, \ldots a$, $j = 1, \ldots, b$, $k = 1, \ldots, n$?

b) Consider the following experiment to investigate the effect of the type of plate material (A) and Temperature (B) on the life of a battery. The response variable (y) is the life of battery (in hours) when tested at each combination of plate material and temperature. The coded data are as follows:

Material Type	Temperature		
	15	70	125
1	2	-8	-2
	4	-4	-1
	6	-4	-3
2	5	6	-5
	7	4	-6
	4	2	-7

Analyse the data. Is there any indication that either factor influences life? Do the two factors interact? (Use $\alpha = 0.05$ for both.)

$(\overline{3 + 12 + 3} + \overline{10 + 2 + 1}) = [31]$

4. Construct a 3^{4-2} design with $I = ABC$ and $I = AB^2D^2$. What is the resolution of this design and why? Write the alias set of A.

(8+3+5)=[16]

5. a) What is a response surface? Describe the method of steepest ascent.

b) The region of experimentation for three factors are time ($40 \leq T_1 \leq 80$ min), temperature ($200 \leq T_2 \leq 300$°C), and pressure ($20 \leq P \leq 50$ psig), A first order model in coded variables has been fit to yield data from a 2^3 design. The model is

$$\hat{y} = 30 + 5x_1 + 2.5x_2 + 3.5x_3$$

Are the points (i) $T_1 = 65$, $T_2 = 256.25$, $P = 37.625$ and (ii) $T_1 = 85$, $T_2 = 291.25$, $P = 55.125$ on the path of steepest ascent?

$(\overline{4 + 7} + 4) = [15]$

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
Back Paper Examination: 2015-2016 (First Semester)

M. Tech. (QR & OR), II Year

Applied Stochastic Processes

Date: 19/02/16 Full Marks: 100 Duration: 3 hours.

Note: *Answer all questions.*

1. Let $\{X(t), 0 \leq t \leq 1\}$ be the stochastic process defined from a Poisson process $\{N(t), t \geq 0\}$ with rate λ as follows:

$$X(t) = N(t) - tN(1), \quad \text{for } 0 \leq t \leq 1.$$

Find $\text{Cov}[X(t_1), X(t_2)]$, for $0 \leq t_i \leq 1$, $i = 1, 2$.

[10]

2. Suppose in a Poisson process with rate λ, n events have occurred during $(0, T]$. Let T_r $(r \leq n)$ be the time of occurrence of the rth event. Show that T_r/T follows a beta distribution with parameters $r - 1$ and $n - r$. Obtain the mean and variance of T_r.

[8]

3. Let Λ be a positive random variable with mean μ_λ and variance σ_λ and the counting process $\{N(t), t \geq 0\}$ given $\Lambda = \lambda$ is a Poisson process with rate λ.

   (a) Find the variance of $N(t)$.

   (b) Find the distribution of $N(t)$, if Λ has a gamma distribution with pdf

$$f_\Lambda(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \ \alpha > 0, \ \beta > 0.$$

[5+7=12]

4. Consider a birth and death process $\{X(t), t \geq 0\}$ with birth and death rates $\lambda_n = n\lambda$ and $\mu_n = n\mu$ (for $n \geq 1$), respectively and $\lambda_0 = \mu_0 = 0$. Derive the p.g.f. of $X(t)$. Compute the probability of extinction.

[10+2 = 12]

5. Describe the $M|M|s|k$ queueing system ($s=$ number of servers and $k=$system capacity) as birth and death process. Write down the differential equations and obtain the steady-state probabilities.

[3+2+5=10]

6. (a) Consider time-homogenous and stationary continuous-time Markov chain with R as the transition rate matrix. Show that

$$\frac{d}{dt}P(t) = P(t).R$$

(b) Write down the transition rate matrix of a linear birth and death process.

[8+4 =12]

7. A man works on a temporary basis. The mean length of each job he gets is three months. If the amount of time he spends between jobs is exponentially distributed with mean 2(months), then at what rates does the man get new jobs?

[5]

8. Consider a renewal process $\{N(t), t \geq 0\}$ having interarrival timse X_n, $n \geq 1$, and suppose $\mu = E[X_1] < \infty$. Define $S_{N(t)} = \sum_{i=1}^{N(t)} X_i$.

(a) Show that $E[S_{N(t)+1}] = \mu(M(t)+1)$, where $M(t) = E[N(t)]$.

(b) Using (a), prove that $\dfrac{M(t)}{t} > \dfrac{1}{\mu} - \dfrac{1}{t}$.

[8+2=10]

9. Consider a Renewal process $\{X_i\}$ with mean renewal time μ. With the ith renewal, there is an associated cost Y_i incurred at the beginning of the renewal interval. Assume Y_i's to be i.i.d. with mean ν. Let $W(t)$ be the accumulated cost up to time t and $A(t) = E[W(t)]$.

(a) Find $A(t)$.

(b) If F is not lattice, then show that $\lim_{t\to\infty} \dfrac{A(t)}{t} = \dfrac{\nu}{\mu}$.

[9+4=13]

10. Consider a discrete time branching process in which in each generation an individual either dies or is replaced by two offspring, the probability of two events being p_0 and p_2 respectively. Let X_n be the population size of the nth generation. Assuming $X_0 = 1$, derive the recursive relationship for the probability generating function of X_n.

[8]